On multivariate survival models with a skewed frailty and a correlated baseline hazard process

Sujit K. Sahu
Faculty of Mathematical Studies,
University of Southampton, UK

Dipak K. Dey
Department of Statistics,
University of Connecticut, USA

January 9, 2003

Abstract

Often the dependence in multivariate survival data is modeled through an individual level effect called the frailty. Due to its mathematical simplicity the gamma distribution is often used as the frailty distribution. However, it is well known that the gamma distribution for frailty has many drawbacks. For example, it weakens the effect of covariates. To overcome such drawbacks more heavy tailed distributions are needed to model the frailty distribution, e.g. the positive stable distribution. In this paper we develop a class of log-skew-t distributions for the frailty. This class includes the log-normal distribution along with many other heavy tailed distributions, e.g. log-Cauchy or log-t as special cases.

Conditional on the frailty, the survival times are assumed to be independent with proportional hazard structure. The modeling process is then completed by assuming an appropriate baseline hazard function. There are many prior processes for modeling the baseline hazard. An attractive choice here is a correlated prior process, which offers a great deal of flexibility. We consider such a process, which jumps according to a time-homogeneous Poisson process. We develop Bayesian methods to obtain posterior inference using a variable dimensional Markov chain Monte Carlo method. We illustrate and compare our methods using two practical examples.

KEYWORDS: Auto-correlated prior process, conditional predictive ordinate, frailty, Markov chain Monte Carlo methods, proportional hazard model, reversible jump.

1 Introduction

Multivariate survival data arise when each study subject may experience multiple events or when study subjects are clustered into groups. Examples of such data include the recurrence times of a certain disease and the survival times of members of a family or litter.

Suppose 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}. The conditional hazard function of T_{ij} is often modeled as the product of: (i) an individual level random effect called the frailty, (ii) a baseline hazard function, and (iii) a proportional hazard function, which takes into account the effect of the covariates. That is, given the covariates $\mathbf{z}_{ij}$ and the unobserved frailty parameters w_i, the hazard function is modeled as

$$h(t_{ij}|\mathbf{z}_{ij}, w_i) \;=\; w_i \; h_o(t_{ij}) \; \exp(\boldsymbol{\beta}^T \mathbf{z}_{ij}) \tag{1.1}$$

where $h_o(\cdot)$ is the baseline hazard function and $\boldsymbol{\beta}$ is the regression parameter. Observe that the frailty parameter w_i can be specified equivalently by the one-to-one transformation $b_i = \log(w_i)$ as well.

The main focus of the current paper is to develop flexible models for the frailty and the baseline hazard function. In particular, we adopt the following two new approaches to modeling.

- We develop a flexible class of log-skew-t frailty distribution for modeling the dependence. The class includes the log-normal distribution along with other heavy tailed distributions such as the log-t distribution as special cases. As is well known, inferences based on models that use heavy tailed frailty distributions are more robust to outliers.
- We propose a correlated prior process for the baseline hazard function, which jumps according to a time-homogeneous Poisson process. Thus the number and positions of the jump times are not fixed in advance, but are estimated using data and prior assumptions.

The *frailty* is a random effect common to the individuals of the same group or cluster. See for example the recent book by Ibrahim, Chen and Sinha (2001) for general introduction and

early references. A convenient choice of the frailty distribution is the gamma distribution since it provides conjugate sampling distributions for Gibbs sampling (Gelfand and Smith, 1990). However, it has many drawbacks, see e.g. Hougaard (1986). For example, the gamma distribution attenuates the covariate effect. To overcome such problems, Hougaard (1986) uses the positive stable frailty distributions, which have been followed up by Qiou, Ravishanker and Dey (1999) in the Bayesian context.

In this article we consider an alternative class of frailty distributions using the log-skew-t distributions. The class *before the transformation* reduces to the family of normal (t or Cauchy) distributions for particular values of the parameters. On the other extreme it behaves as a half-normal (half-t or half-Cauchy) distribution by placing all its mass on the positive side of the real line. Thus the family of distributions is quite flexible and general, and the existence of the mean and variance of the frailty distribution is not assumed since the degrees of freedom of the t-distribution can be less than two.

A suitable stochastic process needs to be considered for the baseline hazard function. Several parametric and non-parametric models are available, see e.g. Sinha and Dey (1997) for a review. In this article we adopt a piecewise constant baseline hazard function. This choice is popular when modeling univariate survival data, see for example Gamerman (1991), Arjas and Gasberra (1994) and McKeague and Tighiouart (2000).

The correlated prior process imposes smoothness on the baseline hazard function in adjacent intervals. In particular, we generalize a first order autoregressive process considered in Sahu *et al.* (1997). In addition, we assume that the endpoints of the interval themselves form a time-homogeneous Poisson process. This introduces further flexibility since the number of endpoints where jumps are allowed to occur is left unknown.

The full Bayesian model is rather complex and does not allow fitting and comparison using analytic methods. The straightforward Gibbs sampler is also not able to handle the computations since the parameter space is of varying dimension. Thus we develop Bayesian computation methods using the reversible jump MCMC method, see for example Green (1995).

A natural next step after Bayesian model fitting is to investigate the issues relating to model adequacy and model choice. Here we adopt familiar predictive Bayesian model choice criteria and adequacy checks for comparing different models. In particular, we use the pseudo-Bayes factor, see e.g. Geisser and Eddy (1979) for model comparison.

The remainder of the article is organized as follows. Section 2 introduces the frailty distributions. The baseline hazard function is discussed in Section 3. The likelihood and prior specifications are discussed in Section 4. Section 5 develops computing methods and may be omitted without loss on a first reading. In Section 6 we provide two numerical examples. We conclude with a few summary remarks in Section 7.

2 Frailty Models

2.1 Frailty Distributions

The popular gamma frailty distribution assumes that w_i, $i = 1, \ldots, n$ are i.i.d. each with the gamma density

$$f(w|\eta) = \frac{\eta^\eta}{\Gamma(\eta)} w^{\eta-1} \exp(-w\eta), \ w > 0. \tag{2.1}$$

Here η^{-1} is the unknown variance of the w_i, and the strength of the association between the survival times is a non-increasing function of η.

The positive stable distribution has been suggested as a suitable frailty distribution on the ground that it preserves the proportional hazard assumption in the marginal model. A good introduction to this distribution is given by Hougaard (2000). The distribution is specified by its Laplace transform $E[\exp(-sW)] = \exp(-s^\alpha)$ where $0 < \alpha \leq 1$. The components of the multivariate survival distribution induced by the stable frailty are locally independent when $\alpha = 1$. Other values of α introduce dependence among the components.

The density of the positive stable distribution is non-standard. We adopt the version obtained by Buckle (1995). See also Ravishanker and Dey (2000). The density is given by

$$f(w|\alpha) = \frac{\alpha\, w^{1/(\alpha-1)}}{|\alpha - 1|} \int_{-1/2}^{1/2} \exp\left\{ - \left| \frac{w}{d_\alpha(y)} \right|^{\alpha/(\alpha-1)} \right\} \left| \frac{1}{d_\alpha(y)} \right|^{\alpha/(\alpha-1)} dy \tag{2.2}$$

where

$$d_\alpha(y) = \frac{\sin(\pi\alpha y + s_\alpha)}{\cos \pi y} \left[\frac{\cos \pi y}{\cos\{\pi(\alpha-1)y + s_\alpha\}} \right]^{(\alpha-1)/\alpha},$$

and $s_\alpha = \min(\alpha, 2-\alpha)\pi/2$. It is well known that the mean and variance of the positive stable distribution do not always exist. However, the mean and and variance of $b = \log(w)$ are available

in closed form and are given by

$$E(b) = -\left(\frac{1}{\alpha} - 1\right)\psi(1), \quad \text{Var}(b) = \left(\frac{1}{\alpha^2} - 1\right)\frac{\pi^2}{6}$$

where $\psi(x)$ is the digamma function. These expressions will be useful when comparing the different frailty distributions.

We contrast the above two frailty distributions with a new class of frailty distributions called the log-skew-t distributions obtained by Sahu *et al.* (2001). We suppose that the frailty parameters $b_i, (= \log w_i)$ $i = 1, \ldots, n$ are independent and identically distributed for every group with the following density,

$$\begin{aligned} f_b(b|\delta,\nu) &= 2\,(1+\delta^2)^{-1/2}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)(\nu\pi)^{1/2}}\left[1+\frac{b^2}{\nu(1+\delta^2)}\right]^{-(\nu+1)/2} \\ &\quad T_{\nu+1}\left[\sqrt{q(b)}\,\delta\,\frac{b}{\sqrt{1+\delta^2}}\right] \end{aligned} \tag{2.3}$$

where

$$q(b) = \frac{\nu+1}{\nu + b^2/(1+\delta^2)}$$

and $T_m(\cdot)$ is the cumulative distribution function of the standard t distribution with m degrees of freedom (df). The parameters δ and ν influence the shape of the distribution as discussed below.

First, observe that with $\delta = 0$ the above density reduces to the standard t density with ν df. In addition if $\nu \to \infty$ then it approaches the normal distribution. Second, with $\delta = 0$ and $\nu = 1$, (2.3) is the pdf of the Cauchy distribution. Clearly, the mean and variance of b does not exist in this case. Third, skewed distributions emerge for non-zero values of δ. For positive values of δ the distribution is positively skewed and for negative values it is negatively skewed. The mean of the distribution exists if $\nu > 1$ and the variance exists if $\nu > 2$. We note the mean and variance of b for future reference. These are given by,

$$E(b) = \left(\frac{\nu}{\pi}\right)^{1/2}\frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)}\delta,$$

and

$$\text{Var}(b) = (1+\delta^2)\frac{\nu}{\nu-2} - \frac{\nu}{\pi}\left(\frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)}\right)^2\delta^2. \tag{2.4}$$

The density of $w = \exp(b)$ implied by (2.3) is given by,

$$f_w(w|\delta,\nu) = \frac{1}{w}\,f_b\left(\log(w)|\delta,\nu\right) \tag{2.5}$$

where $f_b(\cdot|\delta,\nu)$ is given in (2.3).

2.2 Comparison of Frailty Distributions

Frailty models are usually compared by studying the dependence structures induced by them. However, there are two types of dependence structures to consider: global and local. Measures for local dependence structures include the cross-ratio function which is often defined by taking an appropriate ratio of the derivatives of the joint and marginal survival functions. The quantities required to form the cross-ratio function are not available in analytic closed form under the proposed log-skew-t frailty distributions, thus limiting the scope of theoretical comparisons to be made.

To study the frailty models using a global measure of dependence we use the correlation between the log-survival times, though there are other measures available. It is not meaningful to study the correlation between the survival times themselves since the moments of the frailty distribution w do not always exist. To simplify the exposition suppose that the survival times are bivariate with two components denoted by T_1 and T_2. Assuming that the baseline hazard function is Weibull, Hougaard (2000, p227) shows that

$$\text{corr}(\log T_1, \log T_2) = \frac{\text{Var}(b)}{\text{Var}(b) + \pi^2/6}. \tag{2.6}$$

For the gamma frailty model $\text{Var}(\log w) = \psi'(\eta)$. For the stable frailty model $\text{corr}(\log T_1, \log T_2)$ is given by $1-\alpha^2$. The correlation of the log-survival times under the log-skew-t model is obtained using (2.4).

The above analytical results facilitate comparisons between the frailty distributions. To illustrate, we set $\nu = 8$ to have a moderate tail in the log-skew-t frailty distribution. Now we equate $\text{Var}(\log w)$ under the three different frailty models. In addition, we suppose that the $E(\log w)$ under the stable and the log-skew-t model are equal. Solving these equations we obtain $\eta = 1.14$, $\alpha = 0.74$, and $\delta = 0.23$. The resulting $\text{corr}(\log T_1, \log T_2)$ is $1 - 0.74^2 \approx 0.45$.

Since the tail of the frailty distribution plays an important role in dictating the dependence structure we investigate the shape and tail of the frailty densities for the above parameter values. The densities are plotted in Figure 1. The tail of the log-skew-t-density is heavier than the gamma but lighter than the stable density (2.2). Thus it is seen that for the same amount of correlations between the log survival times the log-skew-t distribution provides a flexible alternative to the heavy tailed stable distribution and light tailed gamma distributions.

Note that if w has the gamma distribution (2.1) then $E(\log(w)) = \psi(\eta) - \log(\eta)$, which is

always non-positive. However, under the stable frailty model $E(\log(w))$ is always non-negative. This is the reason for not equating $E(\log(w))$ for the gamma case with that of the other frailty distributions.

3 Baseline hazard function

The assumption of correlated prior processes for the baseline hazard function is very common in survival analysis. See for example, Gamerman (1991), Arjas and Gasberra (1994) and Sinha and Dey (1997) for a review. The setup is as follows.

Suppose that time is divided into g pre-specified intervals $I_k = (\tau_{k-1}, \tau_k]$ for $k = 1, 2, \ldots, g$ where $0 = \tau_0 < \tau_1 < \ldots < \tau_g < \infty$, τ_g being the last survival or censored time. Assume that the baseline hazard is constant within each interval. That is,

$$h_o(t_{ij}) = h_k, \quad \text{for} \quad t_{ij} \in I_k. \tag{3.1}$$

Following Sahu *et al.* (1997) we assume a martingale process prior for $\lambda_k = \log(h_k)$.

We assume that

$$\lambda_k | \lambda_1, \ldots, \lambda_{k-1} \sim N(\lambda_{k-1}, \sigma^2) \tag{3.2}$$

with $\lambda_0 = 0$. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_g)$. See e.g. Gamerman (1991) and Arjas and Gasbarra (1994) for more general models.

Several authors have suggested different choices for the number of grid points g. Some early references include Breslow (1974) and Kalbfleisch and Prentice (1973). More recent solutions to this problem suggest leaving g unspecified, see for example Arjas and Heikkinen (1997) and McKeague and Tighiouart (2000). Following this last article we assume that the jump times $\tau_1, \tau_2, \ldots$ form a time-homogeneous Poisson process with rate a. This has several advantages over a fixed value of g. One such advantage is that the number and positions of the grid points need not be fixed in advance.

4 Model Specification

4.1 Likelihood Specification

We only consider right-censored survival data and assume that the censoring is non-informative. 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. Thus the triplet $(t_{ij}, \delta_{ij}, \mathbf{z}_{ij})$ is observed for all i and j. Let $(\mathbf{x}, \mathbf{z})$ denote the collection of all such triplets $(t_{ij}, \delta_{ij}, \mathbf{z}_{ij})$. Let $\mathbf{w}$ denote the vector of unobserved w_i's.

The likelihood is derived as follows. The jth subject of the ith group has a constant hazard of $h_{ij} = w_i h_k \theta_{ij}$ 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., $t_{ij} > \tau_k$, the likelihood contribution is $\exp\{-h_k \Delta_k \theta_{ij} w_i\}$ where $\Delta_k = \tau_k - \tau_{k-1}$. If the subject has failed or was censored in the kth interval, i.e., $\tau_{k-1} < t_{ij} \leq \tau_k$ then the likelihood contribution is $(h_k \theta_{ij} w_i)^{\delta_{ij}} \exp\{-h_k (t_{ij} - \tau_{k-1}) \theta_{ij} w_i\}$. Hence, we arrive at the following likelihood, $L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{w}, g;\ \mathbf{x}, \mathbf{z})$ say,

$$\prod_{i=1}^{n} \prod_{j=1}^{m} \left\{ \prod_{k=1}^{g_{ij}} \exp(-h_k \Delta_k \theta_{ij} w_i) \right\} \left(h_{g_{ij}+1} \theta_{ij} w_i\right)^{\delta_{ij}} \exp\left\{-h_{g_{ij}+1} (t_{ij} - \tau_{g_{ij}}) \theta_{ij} w_i\right\}, \tag{4.1}$$

where g_{ij} is such that $t_{ij} \in (\tau_{g_{ij}}, \tau_{g_{ij}+1}] = I_{g_{ij}+1}$.

4.2 Prior Specification

The joint prior distribution of all the parameters is given by

$$\prod_{i=1}^{n} [\pi(w_i | \cdots)]\, \pi(\cdots)\ \pi(g) \pi(\boldsymbol{\lambda}|g)\ \pi(\boldsymbol{\beta}), \tag{4.2}$$

where $\cdots$ denote the hyper-parameters of the frailty distribution. We assume vague prior distributions for the components of $\boldsymbol{\beta}$. Thus each component of $\boldsymbol{\beta}$ is assumed to follow the normal distribution with mean zero and a large variance (10^4) independently. The frailty distribution can be any one of the gamma, stable and log-skew-t distributions given in (2.1), (2.2), and (2.5), respectively. In the gamma frailty case the hyper-parameter is η and we assume that η follows a Gamma distribution with mean 1 (for identifiability) and large variance to incorporate flexibility, $Gamma(\phi, \phi)$ say with a small choice of ϕ. For the parameter α in the stable frailty model we specify the beta prior distribution since the beta family of distributions is quite flexible.

Under the log-skew-t frailty distribution (2.5) we have two hyper-parameters, δ and ν. We assume that δ follows the uniform distribution in an interval $[0, q]$ with pre-specified q. We restrict δ to a bounded interval in order to avoid the conflict between fat tail and skewness in (2.5). Negative values of δ are not considered since we intend to have a positively skewed t distribution.

We now require a suitable prior distribution for ν, the degrees of freedom parameter. Observe that the log-skew-t frailty distribution (2.5) is well defined for any positive value of ν. Thus we treat ν as a continuous random variable taking positive values. We propose the exponential distribution with mean κ as the prior distribution for ν. Finally, we recommend a moderate value of κ (5–15) so that the heavy tailed log-skew-t-distributions are put forward as prior distributions.

The time-homogeneous Poisson process assumption on the jump times implies that $g \sim$ Poisson$(a\,\tau_{\max})$ where $\tau_{\max}$ is the maximum observed survival time. Given g, the martingale specification (3.2) implies that

$$\boldsymbol{\lambda} \sim N_g\left(\mathbf{0},\ \sigma^2 C^{-1}\right) \tag{4.3}$$

where the $g \times g$ matrix C has all elements zero except for $c_{kk} = 2$, $k = 1, \ldots, g$ and $c_{kk+1} = -1$, $k = 1, \ldots, g-1$ and $c_{k-1k} = -1$ for $k = 2, \ldots, g$.

Let $\boldsymbol{\zeta}$ denote the collection of parameters, for which the prior distribution is given by (4.2). The joint posterior density of $\boldsymbol{\zeta}$ is simply proportional to the likelihood (4.1) times the prior (4.2), i.e.

$$\pi(\boldsymbol{\zeta}|\mathbf{x}, \mathbf{z}) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{w}, g;\ \mathbf{x}, \mathbf{z}) \times \prod_{i=1}^{n} [\pi(w_i|\cdots)]\, \pi(\cdots)\ \pi(g)\pi(\boldsymbol{\lambda}|g)\ \pi(\boldsymbol{\beta}). \tag{4.4}$$

4.3 Hyper-parameter values and prior sensitivity

Several hyper-parameters and simulation constants need to be specified in order to successfully adopt the Bayesian approaches. The aim is to keep the prior distributions as vague as possible so that the data, rather than the prior, drives the inference. Also sensitivity of the assumed values to statistical inference need to be checked. The adopted hyper-parameter values for our two examples in Section 6 are discussed below.

For the gamma distribution prior on η of the gamma frailty model ϕ is taken to be 0.001.

This specifies a diffuse but proper prior distribution for the inverse-variance parameter. The skewness parameter, δ, was given a uniform prior in $[0, 5]$. The resulting interval was large enough to capture the skewness in the frailty distribution. The degrees of freedom parameter ν was given an exponential prior distribution with mean 10. This is to specify a moderate to heavy tail of the assumed t distribution.

Recall that for the stable frailty model we have assumed the beta distribution for the parameter α. We have experimented with many combination of values for the parameters of the beta distribution including the case for uniform distribution. In the latter case it was difficult to sample from the resulting full conditional distribution due to computer underflow problems similar to the ones reported in Buckle (1995). As a result we work with informative beta prior distribution which is further justified by the fact that in both of our examples there is only weak association present between the multivariate survival times. In particular, we assumed the parameters of the beta distribution to be proportional to a and b, say where $a/(a+b) = 3/4$. This gives a-priori mean of α to be $3/4$ which leads to moderate correlations between the log survival times as discussed earlier in Section 2.2.

The number of jump times, g was assumed to be a Poisson distribution with mean 10, and truncated in the interval 1 to 20. This automatically specifies a to be $10/\tau_{\max}$. The prior variance, σ^2, of the log baseline hazard levels was assumed to be 1. Bayesian inference was largely insensitive to changes in the values of these hyper-parameters. In particular we have always obtained a robust posterior distribution of g with about 6–8% acceptance in the reversible jump steps. The values $g = 7, \ldots, 15$ always accounted for more than 90% of the probability mass.

We have adopted the pseudo-Bayes factor (see e.g. Geisser and Eddy, 1979; Gelfand and Dey, 1994) for model comparison. The cross-validation predictive densities known as conditional predictive ordinates (CPO) have also been used, see e.g. Sahu *et al.*, (1997) for definition. The CPOs measure the influence of individual observations and are often used as predictive model checking tools.

5 Reversible Jump Steps

Observe that the dimension of $\boldsymbol{\zeta}$ changes as the number of jump times g associated with the baseline hazard function changes. Hence we fit the full Bayesian model using the reversible jump Markov chain Monte Carlo method, see Green (1995) and McKeague and Tighiouart (2000). The algorithm is described as follows.

Suppose that the Markov chain is at a current state x and it is intended to move to a new state y where the dimension of x and y can be different. The new point y and its acceptance probability is obtained as follows. The move will be implemented by drawing continuous random variables u and v of appropriate dimensions, such that the dimension of (x, u) is same as the dimension of (y, v). Further, the proposal point y is obtained using a one-to-one transformation $(y, v) = T(x, u)$. The proposed move is then accepted with probability

$$\alpha\{(x,u),(y,v)\} = \min\left\{1, \frac{\pi(y)q_2(v)}{\pi(x)q_1(u)}\left|\frac{\partial T(x,u)}{\partial(x,u)}\right|\right\} \tag{5.1}$$

where $\pi(\cdot)$ is the posterior distribution and $q_1(u)$ and $q_2(v)$ are the densities of u and v, respectively.

It is straightforward to check that the acceptance probability in (5.1) is the usual Metropolis-Hastings acceptance probability for accepting the proposal (y, v) when the current point is (x, u).

We consider the following three move types for our problem,

(a) updating all the parameters in $\boldsymbol{\zeta}$ except for g,

(b) birth of a new jump time,

(c) death of an existing jump time.

The move type (a) does not involve dimension change and is accomplished by the usual Gibbs sampling steps. Sahu *et al.* (1997) provide details for the steps involving $\boldsymbol{\beta}, \mathbf{w}$ and η for the gamma frailty model. The integral in the density of the stable frailty model (2.2) is calculated using the Newton-Cotes numerical quadrature formula, see e.g. Forsythe *et al* (1977). Under the log-skew-t frailty model the calculations make use of appropriate Metropolis-Hastings steps. The remaining two move types change the dimension of $\boldsymbol{\lambda}$ by 1. These are detailed as follows.

Consider the birth move (b) first. A new jump time τ^* is drawn uniformly in $(\tau_0, \tau_{\max})$.

Suppose that $\tau^* \in (\tau_{k-1}, \tau_k)$. Now we need to generate two new log baseline hazard rates λ'_k and λ'_{k+1} in the proposal when the current point is λ_k. Since there is one degree of freedom for proposing the move we simulate u uniformly in $(-\epsilon, \epsilon)$ for some $\epsilon > 0$. Now λ'_k is taken to be a convex combination of λ_{k-1} and $\lambda_k + u$, and λ'_{k+1} is taken to be a convex combination of $\lambda_k - u$ and λ_{k+1}. In particular we take

$$\lambda'_k = \frac{\tau^* - \tau_{k-1}}{\tau_k - \tau_{k-1}}\lambda_{k-1} + \frac{\tau_k - \tau^*}{\tau_k - \tau_{k-1}}(\lambda_k + u), \tag{5.2}$$

$$\lambda'_{k+1} = \frac{\tau^* - \tau_{k-1}}{\tau_k - \tau_{k-1}}(\lambda_k - u) + \frac{\tau_k - \tau^*}{\tau_k - \tau_{k-1}}\lambda_{k+1}. \tag{5.3}$$

Now we set $\lambda'_i = \lambda_i$ for $1 \le i \le k-1$ and $\lambda'_{i+2} = \lambda_{i+1}$ for $k \le i \le g-1$. Further, we let $\tau'_k = \tau^*$ and set $\tau'_i = \tau_i$ for $0 \le i \le k-1$ and $\tau'_{i+1} = \tau_i$ for $k \le i \le g$.

Let $\boldsymbol{\zeta}'$ be the proposed parameter vector where $\boldsymbol{\lambda}$ and g are replaced by $\boldsymbol{\lambda}'$ and $g' = g + 1$, respectively. The remaining parameters in $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}'$ are kept the same. Observe that the Jacobian for this type of move is

$$\frac{2\,\Delta'_k\,\Delta'_{k+1}}{\Delta_k^2}.$$

The density $q_1(u)$ in (5.1) is $\frac{1}{2\epsilon}I(-\epsilon < u < \epsilon)$ and generation of v is not required.

Now the ratio of the full posterior density (4.4) evaluated at $\boldsymbol{\zeta}'$ and $\boldsymbol{\zeta}$ is calculated. Observe that there are some obvious cancellations in the ratio since the parameters $\boldsymbol{\beta}, \mathbf{w}$ and the hyper-parameters of the frailty distribution $\cdots$ are un-changed in $\boldsymbol{\zeta}'$. The density ratio is multiplied by the Jacobian and divided by $q_1(u)$. Finally the acceptance probability (5.1) is calculated by taking the minimum.

The move type (c) is now straightforward. The proposal is generated as follows. An index k is uniformly selected from $\{1, \ldots, g-1\}$ and the corresponding jump time τ_k is removed. The remaining jump times are re-labeled as: $\tau'_i = \tau_i$ for $0 \le i \le k-1$ and $\tau'_i = \tau_{i+1}$ for $k \le i \le g-1$. Now λ_k and λ_{k+1} are to be combined to obtain a new log baseline hazard level λ'_k. Toward this end, we obtain solutions of (5.2) and (5.3) for λ'_k and u (which appear on the right hand sides of the equations). If u falls outside the interval $(-\epsilon, \epsilon)$ then the move is rejected forthwith since the corresponding birth move would not be reversible. Otherwise, we set $\lambda'_i = \lambda_i$ for $1 \le i \le k-1$ and $\lambda'_i = \lambda_{i+1}$ for $k+1 \le i \le g$. The acceptance probability is calculated using the inverse ratio in (5.1) and the fate of the move is decided accordingly.

6 Examples

6.1 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. The primary co-variate is sex of the patients. There were 28 female patients each with two recurrence times some of which were censored.

Table 1 shows the posterior mean, standard deviation and 95% credible intervals for all the parameters. Although the parameter β is significant under all three models, it is farthest from zero and has the smallest variance under the log-skew-t model. That is the covariate effect is most strongly pronounced under the log-skew-t model. The skewness parameter δ under this model is also significant and the estimate of the degrees of freedom parameter ν suggests that a log-t model is better than a log-normal model.

To further quantify the difference between the frailty models we also estimate the predictive survival curve for a typical female patient under the three models. Figure 2 plots the curves. The Kaplan-Meyer estimate of the survival function is also plotted in the same graph. The Kaplan-Meyer estimates ignore the dependence present in the data and are to be *used as rough guide* only. The observed survival times are shown as points on the time axis.

The majority of the observed survival times are below 200. Around time=200 the Kaplan-Meyer estimate shows a sharp decrease in the survival function as the data suggests. The predictive survival function under the log-skew-t model adapts to this most rapidly. Other models follow suit, but at a slower pace than the log-skew-t model.

The last issue we investigate is whether anything is gained by having an unknown number of jump times. We compare two versions of the log-skew-t model. The first is as described above and the second version keeps g fixed at 10, which was the choice adopted by other authors, see e.g. Sahu *et al.* (1997). Here the endpoints are placed at equal time intervals.

The posterior mean estimates of the log baseline hazard function under the fixed g and the random g model are plotted in Figure 3. Survival times are plotted as points on top of the graph. Some interesting conclusions can be made from the plot. The peaks in the hazard functions are caused by observed survival times. Similarly the troughs are seen in the intervals

where there are no observed survival times. The baseline hazard function for the model with random g quickly adapts to an occurrence of a failure while the function corresponding to the fixed case does not. The baseline hazard function in the fixed g case remains unchanged even after a failure has occurred in some intervals. As a result the function for the random g case is seen to be much smoother. That is, the random g model is more flexible.

To compare the fixed and random models using the pseudo-Bayes factor we have also calculated the pseudo-marginal log likelihood under the two models. The pseudo-Bayes factor for the random g model is 3.7 which suggests strong evidence in favor of the random g model. In conclusion, the random g model is much better than the fixed g model.

6.2 Litters Example

We consider the rat tumor data first studied by Mantel *et al.* (1977). There were 50 litters of rats with each litter consisting of three rats. A randomly selected rat from each litter was given a drug and the other two were selected as controls and were given a placebo. The survival time to develop tumor for each rat was recorded.

The parameter estimates are given in Table 2. Note that, as expected for the gamma frailty model the estimate of β is closest to zero. That is, the covariate effect is attenuated under the gamma frailty model. To see this more clearly, the kernel density estimates of the posterior for β under the three models are plotted in Figure 4. The right tail of this density is heavier under the stable frailty model. The figure also shows that the estimate from the proposed log-skew-t model lies in between the estimates from the gamma and the stable frailty model.

The dependence parameter α under the stable model is estimated to be 0.70 which shows moderate local dependence between the component survival times. The estimate of variance of the frailty distribution under the gamma model is also significant. Under the log-skew-t model the estimate of the skewness parameter δ is also significant and the estimate of the degrees of freedom shows that the frailty distribution has a moderate tail, which agrees with the stable frailty model.

It is of interest to check whether the form of the assumed frailty distribution affects the posterior distribution of the number jump times, g. To investigate this the posterior distribution of g for each of the three models is plotted in Figure 5. The estimated distributions seem not

to differ too much, although the distribution under the skew model tends to put less mass on the smaller values of g.

The log-pseudo-Bayes factor in favor of the log-skew-t model compared against the stable model is estimated to be 2.3. This suggests strong preference of the data for the log-skew-t model. In order to visualize this, Figure 6 plots the individual CPOs. The plot shows that 93 of the 150 observations support the log-skew-t model. The data do not show such strong evidence for the gamma frailty model. The pseudo-Bayes factor for the log-skew-t model versus the gamma model is estimated to be 2.45. This shows some positive evidence for the log-skew-t model as well. Hence the proposed frailty model is seen to be better than the stable or gamma frailty models.

7 Conclusion

In this paper we have extended the multivariate survival models in two directions. The log-skew-t frailty distribution adopted here is shown to be more flexible than the popular gamma and stable frailty distributions. The new development is also shown to provide better model fit according to some well known predictive Bayesian model checking and selection criteria in our example. Also we have shown in our examples that the gamma frailty distribution attenuates the covariate effect much more than the proposed log-skew-t frailty distribution. These reasons justify the consideration of the log-skew-t model.

The new frailty distribution cannot take multi-modal shape, however. If it is desired to obtain multi-modal frailty distribution then mixture distributions (Ravishanker and Dey, 2000) or a non-parametric specification in infinite dimensional parameter space (Walker and Mallick, 1997) should be considered.

The baseline hazard function conditional on the frailty distribution is modeled using a flexible martingale process. It imposes smoothing using its neighbors. The choice of jump times of the baseline hazard function is also made quite flexible using a time-homogeneous Poisson process. This removes the ad-hoc assumptions often made when specifying the number and position of the jump times.

We have developed the powerful reversible jump Markov chain Monte Carlo method for

multivariate survival analysis. Our methods can be extended to perform data analysis in many other scenarios including the models with time-dependent covariate effects.

We have illustrated our methods using two well known examples. Various aspects of the new models have been quantified using output of our MCMC implementation. The pseudo Bayes factors show an order of magnitude improvement provided by the new models. These improvements in model fit have been illustrated and explained using various diagnostic plots. The proposed models are shown to be viable alternatives to the gamma and stable frailty models. Thus the contribution of the current paper can be seen in the following comment made by Hougaard (2000). 'Finding the right tools for a given problem is more exciting than using a single tool for all problems.'

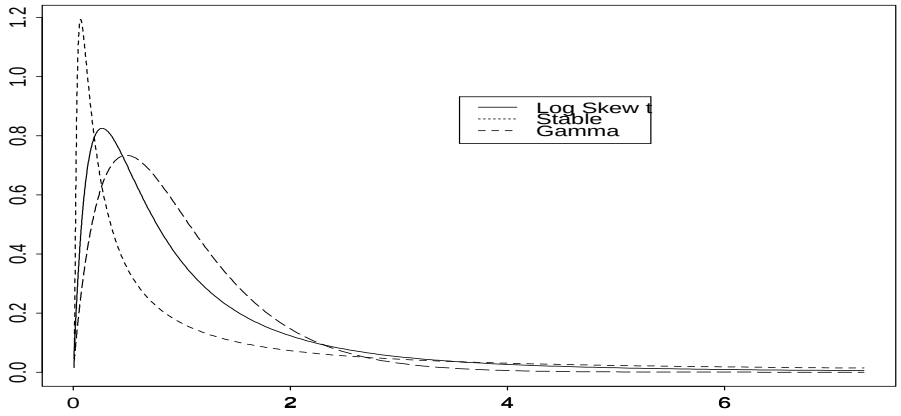

Figure 1: Densities of frailty distributions. The graphs are for (i) the log-skew-t distribution with $\delta = 0.23$ and $\nu = 8$, (ii) positive stable with $\alpha = 0.74$, and (iii) the gamma distribution with $\eta = 1.14$.

	β	φ	ν
Gamma	−1.95 (0.51)	0.58 (0.31)	–
	(−3.02, −1.02)	(0.13, 1.31)	–
Stable	−2.14 (0.55)	0.78 (0.0091)	–
	(−3.20, −0.48)	(0.76, 0.81)	–
Log-Skew-t	−2.34 (0.58)	0.34 (0.26)	13.4 (9.0)
	(−3.47, −1.25)	(0.01, 0.96)	(2.98, 37.10)

Table 1: Parameter estimates from the gamma, stable and log-skew-t model. Posterior means are followed by posterior standard deviations in the first row. 95% credible intervals are shown in the second row. φ is η^{-1} for the gamma model, α for the stable model and δ for the log-Skew-t model.

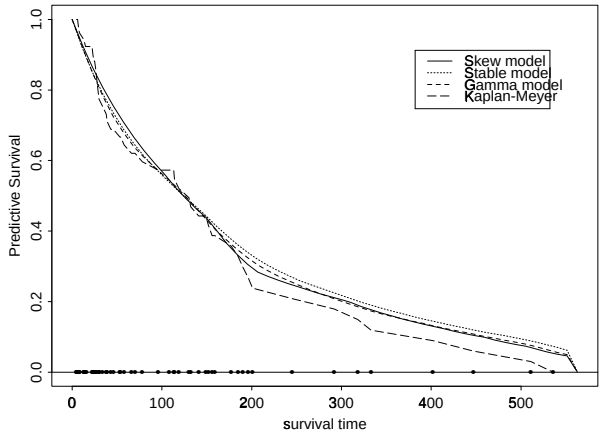

Figure 2: Predictive survival curves for a typical female patient for the kidney infection example.

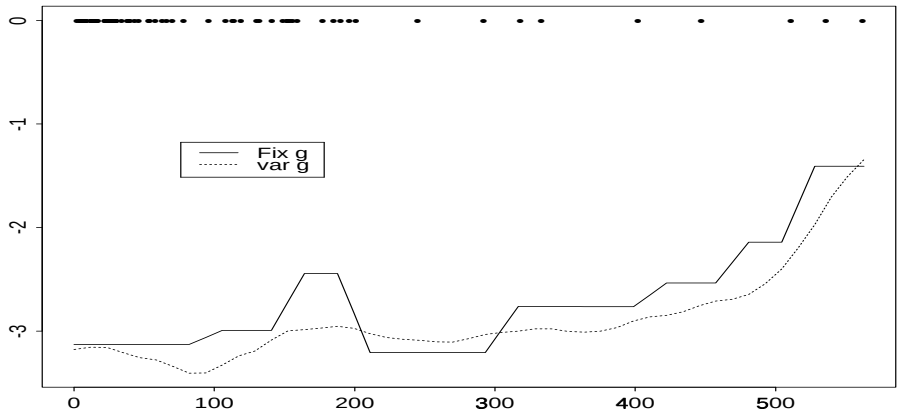

Figure 3: Posterior mean estimate of the log base line hazard function for fixed g and random g model for the kidney infection example.

	β	φ	ν
Gamma	0.69 (0.31)	0.51 (0.45)	–
	(0.09, 1.30)	(0.06, 1.69)	–
Stable	0.78 (0.32)	0.70 (0.0061)	–
	(0.13, 1.41)	(0.69, 0.71)	–
Log-Skew-t	0.71 (0.31)	0.33 (0.25)	15.6 (9.6)
	(0.09, 1.31)	(0.01, 0.95)	(3.32, 40.02)

Table 2: Parameter estimates from the gamma, stable and log-skew-t model for the litters example. Posterior means are followed by (standard deviations) in the first row. 95% credible intervals are shown in the second row. φ is η^{-1} for the gamma model, α for the stable model, and δ for the log-Skew-t model.

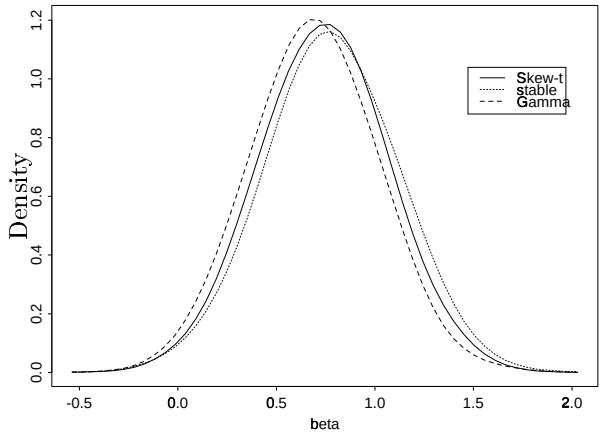

Figure 4: The kernel density estimates of posterior for β under the three different frailty distributions for the litters example.

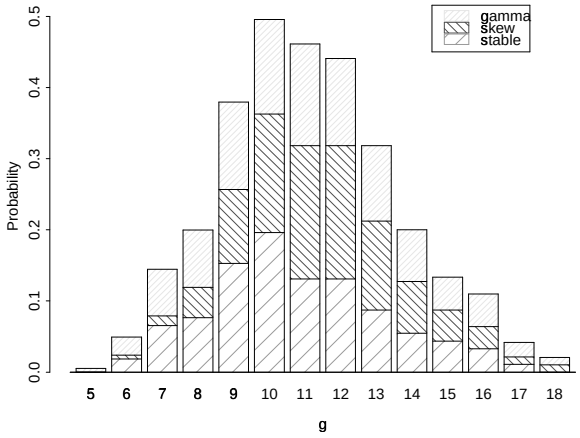

Figure 5: The posterior distribution of g under the three models for the litters example.

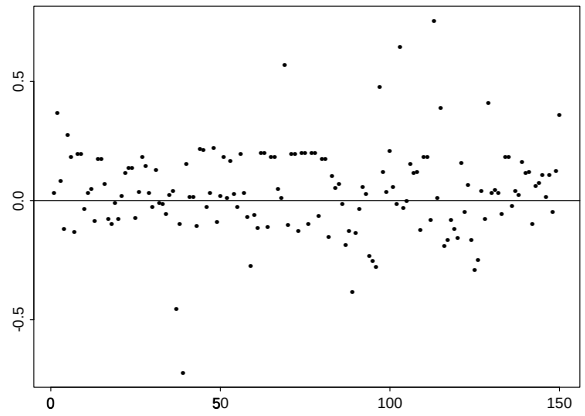

Figure 6: The CPO plot for comparing the log-skew-t model versus the stable model. 93 out of 150 observations support the skew model. The log pseudo-Bayes factor is 2.3.

References

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

Arjas, E. and Heikkinen, J. (1997) An Algorithm for Nonparametric Bayesian estimation of a Poisson intensity. *Journal of Computational Statistics*, **12**, 385–402.

Breslow, N. E. (1974) Covariance analysis of censored survival data, *Biometrics*, **30**, 89–99.

Buckle, D. J. (1995) Bayesian inference for stable distribution. *Journal of the American Statistical Association*, **90**, 605–613.

Forsythe, G.E., Malcolm, M.A. and Moler, C.B. (1977) *Computer Methods for Mathematical Computation.* Upper Saddle River, NJ: Prentice-Hall.

Gamerman, D. (1991) Dynamic Bayesian models for survival data, *Applied Statistics*, **40**, 63–79.

Geisser, S. and Eddy, W. (1979) A predictive approach to model selection. *Journal of the American Statistical Association* , **74**, 153–160.

Gelfand, A. E. and Dey, D. K. (1994) Bayesian model choice: asymptotics and exact calculations, *Journal of the Royal Statistical Society*, B **56**, 501–514.

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

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

Hougaard, P. (1986) A class of multivariate failure time distributions, *Biometrika*, **73**, 671–678.

Hougaard, P. (2000) *Analysis of Multivariate Survival Data.* New York: Springer.

Ibrahim, J. Chen, M.-H. and Sinha, D. (2001) *Bayesian Survival Analysis.* New York: Springer.

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

Mantel, N., Bohidar, N. R., and Ciminera, J. L. (1977) Mantel-Haenszel Analysis of Litter Matched Time-To-Response Data, with Modifications for Recovery of Interlitter Information. *Cancer Research*, **37**, 3863–3868.

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

McKeague, I. W. and Tighiouart (2000) Bayesian Estimators for Conditional Hazard Functions. *Biometrics*, **56**, 1007–1015.

Qiou, Z. Q., Ravishanker, N. and Dey, D. K. (1999) Multivariate survival analysis with positive stable frailties. *Biometrics*, **55**, 637–644.

Ravishanker, N. and Dey, D. K. (2000) Multivariate survival models with a mixture of positive stable frailties. *Methodology and Computing in Applied Probability*, **2**, 293–308.

Sahu, S. K., Dey, D. K., Aslanidou, H. and Sinha, D. (1997) A Weibull Regression Model with Gamma Frailties for Multivariate Survival Data. *Lifetime Data Analysis*, **3**, 123–137.

Sahu, S. K., Dey, D. K. and Branco, M. D. (2001) A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models.
Available from www.maths.soton.ac.uk/staff/Sahu.

Sinha, D. and Dey, D. K. (1997) Semiparametric Bayesian Analysis of Survival Data, *Journal of the American Statistical Association*, **92**, 1192–1212.

Walker, S. G. and Mallick, B. K. (1997) Hierarchical generalized linear models and frailty models with Bayesian non-parametric mixing. *Journal of the Royal Statistical Society*, B, **59**, 845–860.