Austrian Journal of Statistics
July 2023, Volume 52, 95–116.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i4.1481

Modeling Dependence in Survival Times Using Log-Skew Normal Shared Frailties

Sukhmani Sidhu
Department of Statistics,
Panjab University

Suresh K. Sharma
Department of Statistics,
Panjab University

Kanchan Jain
Department of Statistics,
Panjab University

Abstract

In survival studies, event times under a common influence are often grouped together in clusters. The association between and within these clusters can be studied using frailty models where randomness in the data or heterogeneity arising due to unknown covariates is described using a frailty variable. In shared frailty models, frailty value is common or shared for all observations within a cluster while it is conditionally independent for different clusters. In this article, we consider a model whose baseline distribution is Weibull and shared frailties follow log skew-normal distribution. This distribution increases flexibility of the model as it allows the frailty term to be positively or negatively skewed and estimation of skewness parameter enables us to comment on dependence structure of the random component. A simulation study is performed and Bayesian estimates of treatment effects, variance and skewness of frailty term are obtained using Metropolis-Hastings algorithm. It is shown that while bias and expected loss for estimates of all parameters reduce as dataset size increases, frailty parameters are more efficiently estimated when the random component is considered to be skewed. The model and a few popularly used benchmark models are applied to two real-life datasets where positive and negative skewness is observed in the frailty term. The Bayesian estimates obtained for log-skew normal shared frailty model have lower standard errors and provide additional information about the dependence structure in the frailty term.

Keywords: skewed frailty, Metropolis-Hastings, Jeffreys prior, cross ratio function.

1. Introduction

Clustered survival data arises when event times need to be grouped together due to some common factor shared by observations in the same cluster. Such data is observed in clinical, medical, demographical or epidemiological studies and many other related fields. For example, repeat measurements of the same individual or observations from family members, patients under a common treatment/doctor/hospital, students in the same community/school or individuals with exposure to the same environment/household, may need to be grouped together as they share some common risk factors. In all these cases, grouping of event times helps to analyze the data more effectively.

Cox proportional hazards model (Cox 1972) is the most widely used model to analyze sur-

vival data. However, it suffers from the drawback of assuming that given certain observed covariates, the study population is homogeneous. It is not reasonable to make this assumption in most of the studies as individuals differ in many aspects. For example, individuals with no significant difference in covariates, may have a vastly different response to the same treatment. A randomness term is included in the model to depict the heterogeneity due to common unknown risk factors. It acts multiplicatively on the hazard function modifying it for each individual. These common unknown risk factors (frailties) do not always impact the hazard function in a symmetric manner. For example, in a stock market, share price movements may be symmetric on a day to day basis but during certain time frames like union budget, elections, change in key policies, etc., the movements may be positively or negatively skewed. Similarly, susceptibility to infections increases during peak summer or winter seasons, increasing risk of sickness for the exposed population. When the random component is shared or common for all individuals in a cluster but different among clusters, the association between clusters is studied using the shared frailty models (SFM).

The SFM has variability between clusters due to different risks for different groups and dependence between cluster event times due to common frailty. The primary aim is to efficiently estimate parameters of the model and to make valid conclusions by evaluating the effect of a covariate or evaluating the degree of dependence in event times. Analysis of datasets from diverse fields such as economics, medicine, or public health suggest that data may exhibit early or late dependencies between event times. Frailty can be used to model this dependence in the data using a flexible distribution whose density function can take various shapes.

In this paper, we study clustered data under shared frailty effect where the frailty term is described using log skew-normal (LSN) distribution. This distribution allows frailty term to be positively or negatively skewed and hence this model can be used in situations where there are early or late dependencies in survival time. A Monte Carlo Markov Chain (MCMC) algorithm is used to find Bayesian estimates of parameters. Effect of degree of skewness in frailty term, cluster size and number of clusters is seen on the estimates through simulations. The effect of ignoring skewness is also observed on the estimates.

Callegaro and Iacobelli (2012) worked on the LSN shared frailty model and obtained maximum likelihood estimates using EM algorithm. Although they noted that the profile likelihood for the model is not log-concave in their case study, they attributed this to chance. However, as per literature, estimates obtained for parameters of the skew-normal (SN) distribution are not stable due to a monotonic likelihood function with respect to the skewness parameter λ and existence of a stationary point at $\lambda = 0$ for profile log likelihood. Hence, using a frequentist approach for estimating parameters may not be advisable as it may give significantly biased estimates. Various researchers (Azzalini 1985; Azzalini and Capitanio 1999; Liseo and Loperfido 2004, 2006; Sartori 2006; Bayes and Branco 2007; Canale and Scarpa 2013) have addressed this issue and given several numerical techniques to obtain estimates of the parameters. In this paper, we provide an alternative approach by using Jeffreys prior as a weight function in order to obtain finite Bayesian estimates for λ (detailed in Sect. 5.1). We also look into restrictive sampling to avoid issues arising due the stationary point at $\lambda = 0$ (discussed in Sect. 5.3).

In Sect. 2, the general shared frailty model is described followed by a brief discussion on LSN frailties in Sect. 3. In Sect. 4, LSN shared frailty model with a Weibull baseline is given and an estimation procedure is explained in Sect. 5. Simulation studies are carried out in Sect. 6. Sect. 7 consists of real life illustrations. Conclusions and discussions from the paper are presented in Sect. 8.

2. Shared frailty model with censoring

Consider a case where the data under study may be grouped into G clusters each with n_i (i

= 1, 2, ..., G) observations. The clusters can represent time to events like marriage or child birth for different demographies or time to death, recovery or relapse of patients under the care of different hospitals. Associated with each cluster is the unobserved frailty z_i that is assumed to be shared by all observations within a cluster.

For j^{th} individual in the i^{th} cluster, let T_{ij} denote the time to event that may or may not be observed and C_{ij} be the censoring time in cases where the event was not observed due to any reason such as inability to complete the treatment or loss of follow-up etc. The survival times are considered to be right censored where t_{ij}= min$\{T_{ij}, C_{ij}\}$ and

$$\delta_{ij} = \begin{cases} 1 \text{ if } t_{ij} = T_{ij}, \text{ for observed event} \\ 0 \text{ if } t_{ij} = C_{ij}, \text{ for censored event} \end{cases}$$

denotes the censoring indicator.

In shared frailty model, it is assumed that given the frailties, observations belonging to the same cluster are conditionally independent. For j^{th} individual in the i^{th} cluster and given frailty Z = z_i, the conditional hazard function at time $t_{ij} > 0$ is given as

$$h(t_{ij}|z_i, \mathbf{X}_{ij}) = z_i h_0(t_{ij}) e^{\mathbf{X}'_{ij}\boldsymbol{\beta}}, \qquad j = 1, 2, ...n_i, \quad i = 1, 2, ...G. \tag{1}$$

Here $h_0(t)$ denotes the baseline hazard function, $\mathbf{X} = (X_1, X_2, ..., X_r)'$ is a vector of known covariates and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_r)$ is the corresponding vector of treatment effects.

Hence, the conditional survival function, given frailty Z = z_i is

$$\begin{aligned} S(t_{ij}|z_i, \mathbf{X}_{ij}) &= e^{-\int_0^{t_{ij}} h(u|z_i, \mathbf{X}_{ij})du} \\ &= e^{-z_i H_0(t_{ij}) e^{\mathbf{X}'_{ij}\boldsymbol{\beta}}} \end{aligned} \tag{2}$$

where $H_0(t)$ is the cumulative hazard function.

3. Log skew-normal frailties

As studied by Azzalini (1985), skew-normal (SN(ξ,ν,λ)) distribution with location parameter ξ, scale parameter ν, and skewness parameter λ, has the PDF given by

$$f(b : \xi, \nu, \lambda) = \frac{2}{\nu}\varphi\left(\frac{b-\xi}{\nu}\right)\Phi\left(\lambda\frac{b-\xi}{\nu}\right), \qquad b, \xi, \lambda \in \Re; \nu > 0 \tag{3}$$

where φ and Φ denote the PDF and CDF of standard normal distribution respectively. Note that, $\lambda > 0$ (< 0) gives a positively (negatively) skewed distribution and for $\lambda = 0$, $f(b)$ is the PDF of standard normal distribution. Mean and variance of $SN(\xi, \nu, \lambda)$ distribution are

$$\mu_b = \xi + \nu\lambda'\sqrt{\frac{2}{\pi}}, \qquad \sigma_b^2 = \nu^2\left(1 - 2\frac{\lambda'^2}{\pi}\right)$$

where $\lambda' = \dfrac{\lambda}{\sqrt{1+\lambda^2}}$ can be regarded as an alternate parametrization of λ, with $|\lambda'| < 1$.

Skewness coefficient of the distribution is given by

$$\sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\frac{\lambda'^3}{\left(1 - 2\frac{\lambda'^2}{\pi}\right)^{3/2}}$$

and is limited to the interval $(-0.9953, 0.9953)$ when $\lambda' = \pm 1$. Hence this distribution can be used to describe moderate to low degree of skewness in the frailty distribution. In this piece

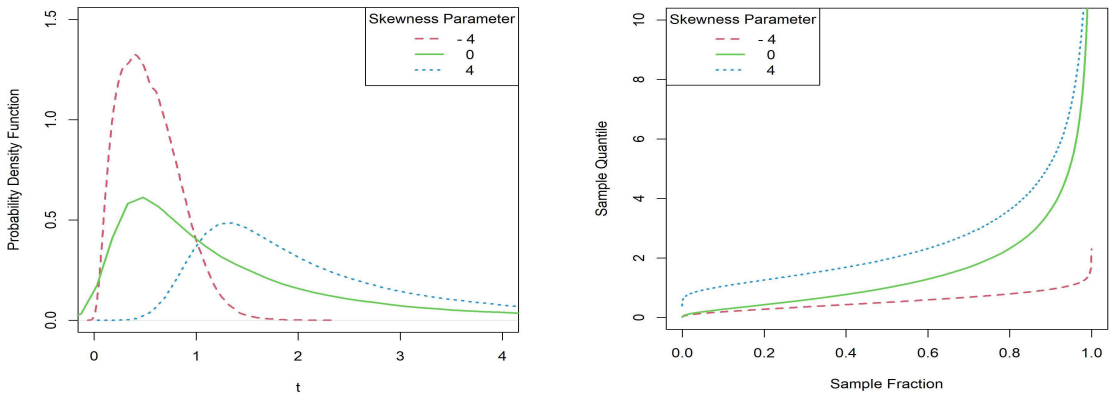

Figure 1: Probability density function (left) and quantile function (right) plots for LSN(0, 1, λ) with varying skewness parameter

of work, we consider $\lambda' \in [-0.9998, 0.9998]$ corresponding to a skewness of [-0.9936,0.9936] since beyond this, the shape of the distribution does not change significantly.

For the frailty variable in the hazard function described in (1), log-transformed distribution of $SN(\xi, \nu, \lambda)$ is used. That is, $log(z) \sim SN(\xi, \nu, \lambda)$, hence $z \sim LSN(\xi, \nu, \lambda)$ with PDF

$$f(z : \xi, \nu, \lambda) = \frac{2}{\nu z}\varphi\left(\frac{log(z) - \xi}{\nu}\right)\Phi\left(\lambda\frac{log(z) - \xi}{\nu}\right) \qquad \xi, \lambda \in \Re; z, \nu > 0. \tag{4}$$

The density function and quantile function in Fig. 1 highlight the heavy tails and skewed nature of this distribution. LSN is moment indeterminate (Lin and Stoyanov 2009) although finite moments exist for positive integer order. The distribution has been studied in detail by various authors (Azzalini, Cappello, and Kotz 2003; Liao, Peng, and Nadarajah 2013; Martinez-Florez, Vergara-Cardozo, and Gonzalez 2013) and has also been generalized further (Hutson, Mashtare Jr, and Mudholkar 2020). The mean and variance of LSN(ξ, ν, λ) are

$$\mu_z = 2e^{\xi + \frac{\nu^2}{2}}\Phi(\lambda'\nu), \qquad \sigma_z^2 = 2e^{2(\xi+\nu^2)}\Phi(2\lambda'\nu) - {\mu_z}^2.$$

For identifiability of parameters, we take

$$\mu_z = 1 \Rightarrow \xi = -log[2e^{\nu^2/2}\Phi(\lambda'\nu)].$$

This gives

$$\sigma_z^2 = \frac{1}{2}e^{\nu^2}\frac{\Phi(2\lambda'\nu)}{\Phi^2(\lambda'\nu)} - 1.$$

It is of interest to estimate the skewness parameter of frailty distribution as it is indicative of the dependence structure of survival times.

A strong early dependence is indicated by a long right tail (positively skewed) and strong late dependence is indicated by a long left tail (negatively skewed). In other words, if more failures occur in the beginning (end) of a study time period, it will be appropriate to assume a positively (negatively) skewed frailty distribution.

One of the measures to study early or late dependence structure exhibited by survival times is the Cross-Ratio Function (CRF) (Clayton 1978) defined as

$$CRF(t_1, t_2) = \frac{H(t_1|T_2 = t_2)}{H(t_1|T_2 > t_2)}.$$

CRF measures the degree of association between a pair of survival times (t_1, t_2) and a higher value indicates greater dependence at that time in the study. For shared frailty model, CRF

can also be written in terms of derivative of the joint survival function $S(t_1, t_2)$ (Oakes 1989) as

$$CRF(t_1, t_2) = \frac{S(t_1, t_2)S_{12}(t_1, t_2)}{S_1(t_1, t_2)S_2(t_1, t_2)} \tag{5}$$

where $S_{ij}(t_i, t_j)$ is the derivative of $S(t_1, t_2)$ with respect to t_i and t_j. Using (2), the bivariate survival function can be written as

$$\begin{aligned} S(t_1, t_2) &= \int_0^\infty S(t_1|z, X_1)S(t_2|z, X_2)f(z)dz \\ &= \int_0^\infty e^{-z[H_0(t_1)e^{X_1\beta}+H_0(t_2)e^{X_2\beta}]}f(z)dz. \\ &= \int_0^\infty e^{-zG(\mathbf{t})}f(z)dz \end{aligned}$$

where $[H_0(t_1)e^{X_1\beta} + H_0(t_2)e^{X_2\beta}] = G(\mathbf{t})$, with $\mathbf{t} = (t_1, t_2)$. The integral is approximated by using Gauss-Laguerre quadrature rule and we get

$$S(t_1, t_2) \approx \sum_n w_n e^{u_n} e^{-u_n G(\mathbf{t})} f(u_n) \tag{6}$$

where w_n and u_n are the weights and nodes of the Laguerre polynomial. Taking derivative of (6) and using (5), the CRF for the general shared frailty model can be approximated as

$$CRF(t_1, t_2) \approx \frac{\sum_n w_n e^{u_n} e^{-u_n G(\mathbf{t})} f(u_n) \sum_n w_n e^{u_n} e^{-u_n G(\mathbf{t})} u_n^2 f(u_n)}{[\sum_n w_n e^{u_n} e^{-u_n G(\mathbf{t})}(-u_n) f(u_n)]^2} \tag{7}$$

where $f(.)$ is PDF of the frailty distribution.

The varying degrees of dependencies between pair of event times (t_1, t_2) are shown in the level plots (Fig. 2) for $\lambda = \pm 4$, $\mu_z = 1$ and $\nu = 0.5$.

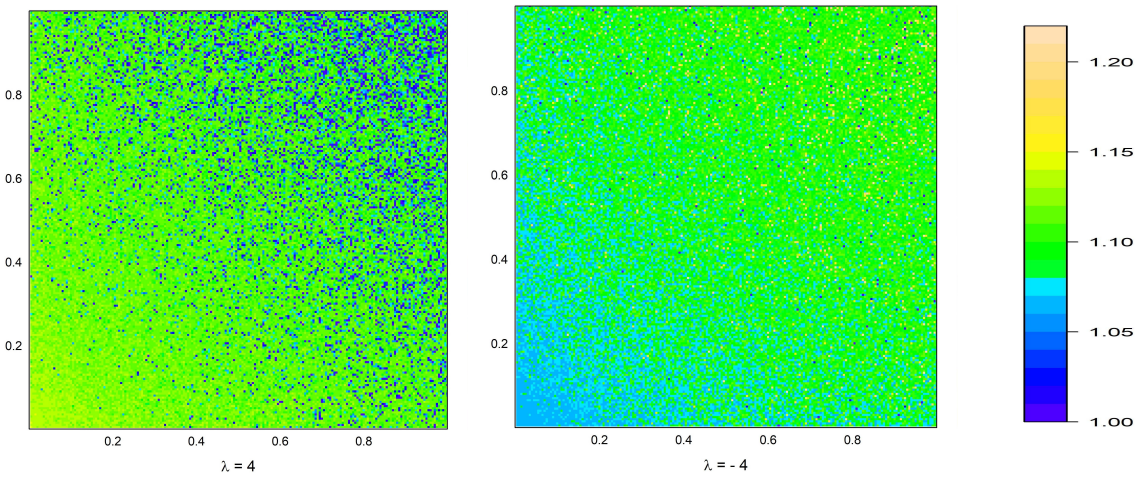

Figure 2: Level plots for CRF (t_1, t_2) for positive and negative skewness

It can be observed from the plots that positive skewness (Fig. 2, Left) gives a higher CRF$(0, 0)$ indicating a higher degree of dependence at earlier stages of study while negative skewness (Fig. 2, Right) gives a higher CRF$(1, 1)$ indicating a higher degree of dependence at later stages of the study. This can also be seen from the graph in Fig. 3 where the CRF(t, t) has been plotted for positive and negative skewness. For comparison, CRF(t, t) is plotted for Gamma, Inverse Gaussian and Compound Poisson distributions with mean 1 and variance 0.5.

It is of interest to note that $\lambda = 0$ corresponds to an early dependence in survival times. Due to log transformation, b although negatively skewed for $\lambda < 0$ does not give a negatively skewed

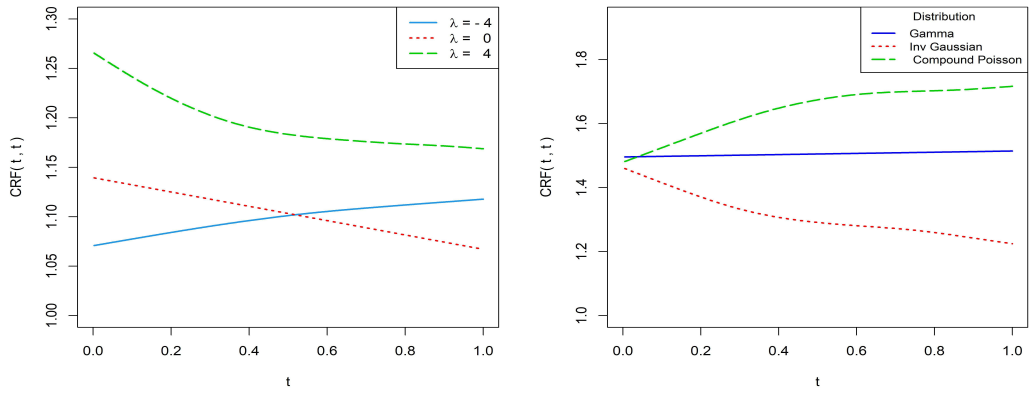

Figure 3: CRF(t,t) for LSN distribution with varying degrees of skewness (left) and CRF(t,t) for Gamma, inverse Gaussian and compound distribution (right)

random variable $z = e^b$. No expression for skewness of LSN distribution is currently available in literature as moments of the distribution require numerical approximations. Hence, using simulations to study the skewness pattern of LSN distribution, we observe that on assuming $\mu_z = 1$ and $\nu = 0.5$, LSN distribution becomes approximately negatively skewed for $\lambda < -4$. This cut-off increases with a decrease in value of ν. Hence, in case of late dependencies, we need to consider a wider range for λ which is not limited to [-10,10] (corresponding to $\lambda' \in [-0.9998, 0.9998]$), as considered in the estimation procedure adopted in this article.

4. Log skew-normal shared frailty model

4.1. Weibull baseline distribution

The two parameter Weibull distribution is frequently used for analysis of life time data as different values of shape parameter ψ lead to different shapes of the hazard function. The corresponding probability density function (PDF) with $\psi > 0$ as shape parameter and $\omega > 0$ as scale parameter is written as

$$f(t) = \psi\omega t^{\psi-1} e^{-\omega t^{\psi}}.$$

The baseline hazard and cumulative baseline hazard functions are given by

$$h_0(t) = \omega\psi t^{\psi-1},$$
$$H_0(t) = \omega t^{\psi}.$$

For $\psi < 1$, this distribution has a decreasing hazard rate with time, which is useful for depicting early life or infantile high mortality. For $\psi > 1$, it increases with time and can be used for wear out or old age data analysis. While for $\psi = 1$, it is fairly constant and can be used for depicting useful or mid-life phase. These three shapes are indicative of the "bath tub curve" of human mortality from infancy till death.

4.2. Likelihood function

The full conditional likelihood function for general shared frailty model with non-informative right censoring has the form

$$L_c = \prod_{i=1}^{G} \prod_{j=1}^{n_i} [h(t_{ij}|z_i, \mathbf{X}_{ij})]^{\delta_{ij}} \, S(t_{ij}|z_i, \mathbf{X}_{ij}), \tag{8}$$

where (1) and (2) give the conditional hazard and survival functions respectively.

Likelihood function L_c, is integrated to obtain the full unconditional likelihood function, that is

$$L = \int_0^\infty \prod_{i=1}^{G} \prod_{j=1}^{n_i} [h(t_{ij}|z_i, \mathbf{X}_{ij})]^{\delta_{ij}} S(t_{ij}|z_i, \mathbf{X}_{ij}) f(z_i) dz_i.$$

Using (1) and (2) and putting $\eta_{ij} = e^{\mathbf{X}'_{ij}\boldsymbol{\beta}}$, we have

$$\begin{aligned}
L &= \prod_{i=1}^{G} \left\{ \int_0^\infty \left(\prod_{j=1}^{n_i} \left[z_i h_0(t_{ij}) e^{\mathbf{X}'_{ij}\boldsymbol{\beta}} \right]^{\delta_{ij}} e^{-z_i H_0(t_{ij}) e^{\mathbf{X}'_{ij}\boldsymbol{\beta}}} \right) f(z_i) dz_i \right\} \\
&= \prod_{i=1}^{G} \left\{ \left(\prod_{j=1}^{n_i} [h_0(t_{ij})\eta_{ij}]^{\delta_{ij}} \right) \int_0^\infty z_i^{\sum_{j=1}^{n_i} \delta_{ij}} e^{-z_i \sum_{j=1}^{n_i} H_0(t_{ij})\eta_{ij}} f(z_i) dz_i \right\} \\
&= \prod_{i=1}^{G} \left\{ \left(\prod_{j=1}^{n_i} [h_0(t_{ij})\eta_{ij}]^{\delta_{ij}} \right) \int_0^\infty z_i^{D_i} e^{-z_i A_i} f(z_i) dz_i \right\}, \\
&\text{where } D_i = \sum_{j=1}^{n_i} \delta_{ij} \text{ and } A_i = \sum_{j=1}^{n_i} H_0(t_{ij})\eta_{ij}.
\end{aligned}$$

Let $I_i = \int_0^\infty z_i^{D_i} e^{-z_i A_i} f(z_i) dz_i$.
Approximating the integral using Gauss-Laguerre quadrature method, with $q_i = z_i A_i$, we have

$$\begin{aligned}
I_i &= \int_0^\infty \left(\frac{q_i}{A_i} \right)^{D_i} e^{-q_i} f\left(\frac{q_i}{A_i} \right) \frac{1}{A_i} dq_i \\
&\approx \frac{1}{A_i^{D_i+1}} \sum_n w_n {u_n}^{D_i} f\left(\frac{u_n}{A_i} \right).
\end{aligned}$$

Therefore, the unconditional likelihood function is

$$L = \prod_{i=1}^{G} \left\{ \left(\prod_{j=1}^{n_i} [h_0(t_{ij})\eta_{ij}]^{\delta_{ij}} \right) I_i \right\}. \tag{9}$$

5. Estimation procedure

In the Bayesian method of estimation, parameters of the model $\boldsymbol{\theta} = (\psi, \omega, \beta_1, ..., \beta_r, \nu, \lambda)$ are considered to be random variables rather than unknown constants. Available information about $\boldsymbol{\theta}$ is incorporated in the model using a prior distribution $p(\boldsymbol{\theta})$ which along with data collected from an independent study $L(x|\boldsymbol{\theta})$, gives the updated posterior distribution as

$$\pi(\boldsymbol{\theta}|x) \propto L(x|\boldsymbol{\theta}) p(\boldsymbol{\theta})$$

$\pi(\boldsymbol{\theta}|x)$ is then used to draw conclusions about $\boldsymbol{\theta}$.

Assuming the parameters to be independently distributed, joint posterior density can be written as

$$\pi(\psi, \omega, \boldsymbol{\beta}, \nu, \lambda) \propto L(\psi, \omega, \boldsymbol{\beta}, \nu, \lambda) g_1(\psi) g_2(\omega) g_3(\nu) g_4(\lambda) \prod_{i=1}^{r} p_i(\beta_i), \tag{10}$$

where $g(.)$ denotes the prior density function of ψ, ω, ν, and λ and $p_i(.)$ is the prior density function of treatment effect β_i. The hyperparameters of prior density functions are assumed to be known.

5.1. Prior densities

There is generally little prior information available about the parameters and the prior distribution may often be based on guesses made by experts. Hence it is not desirable for $p(\boldsymbol{\theta})$ to have a significant impact on the estimation procedure. Also, information collected in the current study would outweigh any prior information incorporated in the model as specified prior probabilities. Thus, "reference priors" (Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin 2013), that are vague, diffused, or flat and cover all reasonable values taken by $\boldsymbol{\theta}$, are used in the estimation procedure to enable data to play a more significant role in the inference made about the parameters.

Normal prior N(0,ϕ) is used for the treatment effects β_i and Gamma prior G(ϵ,ϵ) is used for positive parameters ψ, ω, and ν. Hyperparameters ϕ and ϵ are taken to be 1000 and 0.0001 respectively, such that the prior variance is large. High prior variances ensure that the prior is flat in order to reflect uncertainties about treatment effects. While such priors have been used efficiently in literature (Sahu, Dey, Aslanidou, and Sinha 1997; Ibrahim, Chen, and Sinha 2001; Santos and Achcar 2011; Hanagal and Dabade 2013; Sidhu, Jain, and Sharma 2018) for the shape parameter λ such diffused prior distributions do not yield a proper posterior density. Earlier researchers (Azzalini 1985; Azzalini and Capitanio 1999; Liseo and Loperfido 2004, 2006; Sartori 2006; Bayes and Branco 2007; Canale and Scarpa 2013) have also noted that estimation of λ poses some intrinsic problems. This can be explained by considering a sample from $LSN(0, 1, \lambda)$ whose likelihood function is of the form

$$L = \prod_{i=1}^{n} \frac{2}{y_i} \varphi(log(y_i)) \Phi(\lambda log(y_i)).$$

If $y_i > 1$ $(0 < y_i < 1)$ $\forall$ $i = 1, 2, ...n$, the likelihood function is a monotonically increasing (decreasing) function of λ. Hence any estimation procedure maximizing L will give $\pm\infty$ as an estimate for λ even with positive and negative observations.

In Metropolis-Hastings algorithm, the accept-reject rule makes use of the value of posterior density at two points. The algorithm jumps from a point $\theta^{(i)}$ to $\theta^{(*)}$, always accepting $\theta^{(*)}$ in the sample if posterior density increases, but only sometimes accepting $\theta^{(*)}$ when the posterior density decreases. Since posterior density depends on L, hence if L is monotonic and the algorithm accepts points that increase the posterior density, Bayesian estimate for $\lambda(\lambda')$ will tend to $\pm\infty(\pm 1)$ with a positive probability (Liseo and Loperfido 2006).

Due to this problem, we suggest using the Jeffreys prior for λ. This prior is symmetric about 0 and decreasing for $|\lambda|$ and hence acts as a weight function to obtain finite estimates. Jeffreys prior for λ is given by

$$\sqrt{E\left[\frac{\partial}{\partial\lambda} log f(z;\xi,\nu,\lambda)\right]^2}.$$

Using (4), we get

$$\frac{\partial}{\partial\lambda} log f(z;\xi,\nu,\lambda) = \frac{\varphi\left[\lambda\left(\frac{log(z)-\xi}{\nu}\right)\right]\left\{\frac{log(z)-\xi}{\nu}\right\}}{\Phi\left[\lambda\left(\frac{log(z)-\xi}{\nu}\right)\right]}.$$

This gives

$$E\left[\frac{\partial}{\partial\lambda}logf(z;\xi,\nu,\lambda)\right]^2 = \int_0^\infty \frac{\varphi^2\left[\lambda\left(\frac{log(z)-\xi}{\nu}\right)\right]\left\{\frac{log(z)-\xi}{\nu}\right\}^2}{\Phi^2\left[\lambda\left(\frac{log(z)-\xi}{\nu}\right)\right]} f(z;\xi,\nu,\lambda)dz$$

$$= \int_0^\infty \frac{\varphi^2\left[\lambda\left(\frac{log(z)-\xi}{\nu}\right)\right]\left\{\frac{log(z)-\xi}{\nu}\right\}^2}{\Phi^2\left[\lambda\left(\frac{log(z)-\xi}{\nu}\right)\right]} \frac{2}{\nu z}\varphi\left[\frac{log(z)-\xi}{\nu}\right]\Phi\left[\lambda\left(\frac{log(z)-\xi}{\nu}\right)\right]dz.$$

Writing $\frac{log(z)-\xi}{\nu} = s$, we have

$$E\left[\frac{\partial}{\partial\lambda}logf(z;\xi,\nu,\lambda)\right]^2 = \int_{-\infty}^{\infty} \frac{2s^2\varphi^2(\lambda s)\,\varphi(s)}{\Phi(\lambda s)}ds$$

$$= E_s\left[\frac{s^2\varphi^2(\lambda s)}{\Phi^2(\lambda s)}\right], \qquad \text{where } s \sim SN(0,1,\lambda). \tag{11}$$

Here the expectation in (11) is approximated using Monte Carlo method by drawing a sample from $SN(0,1,\lambda)$.

5.2. Conditional posterior densities

Full conditional distribution of each parameter is obtained using (9) and (10), by considering the terms involving only that parameter. For example, for ψ we have

$$\pi(\psi|\omega,\boldsymbol{\beta},\nu,\lambda) \propto \prod_{i=1}^{G}\left\{\left(\prod_{j=1}^{n_i}[h_0(t_{ij})\eta_{ij}]^{\delta_{ij}}\right)I_i\right\}g_1(\psi)$$

Hence, collating the terms that involve ψ, we have,

$$\left.\begin{aligned}
\pi(\psi|\omega,\boldsymbol{\beta},\nu,\lambda) &\propto \prod_{i=1}^{G}\left(\prod_{j=1}^{n_i}h_0(t_{ij})^{\delta_{ij}}\right)I_i\ g_1(\psi).\\
\text{Similarily}\qquad\qquad&\\
\pi(\omega|\psi,\boldsymbol{\beta},\nu,\lambda) &\propto \prod_{i=1}^{G}\left(\prod_{j=1}^{n_i}h_0(t_{ij})^{\delta_{ij}}\right)I_i\ g_2(\omega),\\
\pi(\nu|\psi,\omega,\boldsymbol{\beta},\lambda) &\propto \prod_{i=1}^{G}I_i\ g_3(\nu),\\
\pi(\lambda|\psi,\omega,\boldsymbol{\beta},\nu) &\propto \prod_{i=1}^{G}I_i\ g_4(\lambda),\\
\pi(\beta_i|\psi,\omega,\boldsymbol{\beta_i},\nu,\lambda) &\propto \prod_{i=1}^{G}\left(\prod_{j=1}^{n_i}\eta_{ij}^{\delta_{ij}}\right)I_i\ p_i(\beta_i) \text{ for i = 1, 2, ..., r}
\end{aligned}\right\} \tag{12}$$

where $\boldsymbol{\beta_i} = (\beta_1,\beta_2,...,\beta_{i-1},\beta_{i+1},...,\beta_r)$ is a vector of all treatment effects excluding β_i.

5.3. Algorithm

Bayesian estimates of LSN shared frailty model are found using Metropolis-Hastings algorithm (Metropolis and Ulam 1949; Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953) by

generating a sample from the conditional posterior densities of each parameter as given in (12)). Proposal distributions q(.) for ψ, ω, $\boldsymbol{\beta}$, and ν are taken as normal centered around the previous point accepted in the chain and a variance 's' such that approximately 25% of the proposals are accepted (Gelman, Roberts, and Gilks 1996).

For λ, the proposal distribution requires some restrictions due to two reasons. Firstly, the profile log likelihood has a stationary point at 0 and there is always a point of maxima on either side with $\lambda = 0$ being a point of low density. As a result, to ensure log-concavity of the likelihood function, the range for λ proposal is restricted to be either positive or negative. Since the type of dependence exhibited by the survival times is generally known in advance, either a positively skewed ($\lambda > 0$) or a negatively skewed ($\lambda < 0$) distribution may be considered during estimation , as appropriate for the data. Secondly, there are still chances of the Markov chain getting stuck at extreme values ($\pm\infty$) even though such cases are reduced by using Jeffreys prior. Hence to improve efficiency of the algorithm, a Uniform proposal, U(0,M) (U(-M,0)) is used for $\lambda > 0$ ($\lambda < 0$) where M is chosen such that the algorithm is able to explore the parametric space corresponding to skewness coefficient in the range [-0.9936, 0.9936].

Algorithm for sampling from the conditional posterior densities is summarized as follows:

1. Parameters are assigned initial values $\theta^{(0)} = (\psi^{(0)}, \omega^{(0)}, \boldsymbol{\beta^{(0)}}, \nu^{(0)}, \lambda^{(0)})$,
2. Iteration counter 'i' is set to 1,
3. A new value is generated for a parameter using proposal distribution q(.). For example, ψ^* is proposed by generating a random number from $N(\psi^{(i-1)}, s)$,
4. The acceptance ratio (AR), is calculated for proposed value. For ψ^*,
$$AR = min\left\{1, \frac{\pi(\psi^*|\omega, \boldsymbol{\beta}, \nu, \lambda)q(\psi^{(i-1)}|\psi^*)}{\pi(\psi^{(i-1)}|\omega, \boldsymbol{\beta}, \nu, \lambda)q(\psi^*|\psi^{(i-1)})}\right\},$$
5. Random number 'r' is generated using U(0,1),
6. ψ^* is accepted as $\psi^{(i)}$ if $r \leq AR$ else $\psi^{(i)} = \psi^{(i-1)}$,
7. Steps (3) - (6) are repeated for remaining parameters using $\theta^{(*)} = (\psi^{(1)}, \omega^{(0)}, \boldsymbol{\beta^{(0)}}, \nu^{(0)}, \lambda^{(0)})$, each time revising θ using the value accepted in step (vi). N samples are selected from $\pi(\boldsymbol{\theta})$ by repeating these steps.

5.4. Posterior summaries

Graphical methods are used to study behaviour of Markov chains in order to decide burn-in and lag for obtaining a sample of independent observations. As the first few iterations are expected to be dependent on initial values, Cumulative Mean (CM) plots help to ascertain time of convergence of the chains and number of iterations (from the beginning) that need to be removed as burn-in. Trace plots help to ascertain whether the chain is mixing well and explores the parametric space, while Autocorrelation Function (ACF) plots are used to examine autocorrelation in the data. Markov chain has autocorrelation as each point in the chain is generated using the previous accepted point. Autocorrelation is expected to decrease as the lag increases, hence after removing the first few iterations as burn-in, the chain is thinned by selecting every l^{th} value to get a sample of independent observations. To test for convergence and stationarity of Markov chain, the Geweke (Geweke 1992) and Heidelberger - Welch tests (Heidelberger and Welch 1983) are used.

Posterior summaries are obtained using the thinned sample. Squared error loss function (SELF) is used for ψ, ω, $\boldsymbol{\beta}$ and ν, with corresponding Bayes' estimator as the posterior mean. In case of a skewed frailty distribution, there is risk involved in obtaining estimate of λ close

to 0 as that would indicate absence of skewness. Hence, in order to penalize the risk of underestimation (overestimation) when $\lambda > 0$ ($\lambda < 0$), we use the LINEX loss function which is an asymmetric loss function (Varian 1975; Zellner 1986) given by

$$L(\triangle) = e^{c\triangle} - c\,\triangle - 1, c \neq 0$$

where $\triangle = \hat{\lambda} - \lambda$ and 'c' is a scalar that controls the degree of penalization. For this study, it is taken to be -0.5 (0.5) for $\lambda > 0$ ($\lambda < 0$) when underestimation (overestimation) has to be penalized.

The Bayes' estimator of λ under LINEX loss function is given by

$$-\frac{1}{c} log(E[e^{-c\lambda}]). \tag{13}$$

Here, the expectation is estimated using the posterior sample obtained for λ.

All computations were performed using a code written in R statistical software with the help of functions available as a part of 'coda' (Plummer, Best, Cowles, and Vines 2006) and 'statmod' (Smyth 2005) R packages.

6. Simulation studies

The effect of number and size of clusters and skewness of frailty term on the estimation procedure was observed by using different setups consisting of 50, 100, and 200 clusters, with equal cluster size of 4 and 8, and $\lambda = \pm 1, \pm 2$, and $\pm\,3$. For computational convenience, equal clusters have been considered, although the model is applicable for different cluster sizes as well.

We simulate the survival times by generating random numbers from uniform distribution U(0,1) and equating them to the conditional survival function given by (2). Hence, the survival times are obtained as

$$t_{ij} = \left[\frac{-log(rand)}{\omega z_i e^{\mathbf{X'_{ij}}\boldsymbol{\beta}}}\right]^{\frac{1}{\psi}}. \tag{14}$$

In the simulation study, we considered a model with one treatment effect X_i generated from N(0,1) with corresponding $\beta = -1$. The baseline distribution was taken to be Weibull with ψ and ω both equal to 2 and the frailty effect was simulated from the log-skew normal distribution with mean 1 and $\nu = 0.5$. The survival times simulated using (14) were right censored with censoring variable following U(0,C) where C is selected to ensure a censoring of approximately 10% for each setup.

500 datasets were generated under each setting. For each dataset, a chain of 45,000 was run for each parameter using the algorithm given in Sect. 5.3. After studying the CM plots, trace plots and ACF plots for approximately 50 datasets under each setting, initial 10,000 iterations were discarded as burn-in and the sample was thinned by choosing every 150^{th} iteration.

The convergence and stationarity of Markov chains for each parameter for all datasets was checked using the Geweke and Heidelberger - Welch tests. Tables 1 and 2 report the percentage of convergent and stationary chains (p-value ≥ 0.05).
After removing the burn-in and thinning the chain, the remaining posterior sample was used to obtain an estimate of parameter for the corresponding dataset. It was observed in the simulations that the posterior samples obtained for ψ, β, and ν were symmetric. Hence the simple mean was used to summarize the thinned sample obtained for each of the 500 datasets and the corresponding Bayes' estimate that is, posterior mean was calculated. Frailty variance (σ^2_{FR}) was estimated at each iteration in the chain using estimates of ν and λ. The bias in the estimates obtained for ψ, β, ν, and σ^2_{FR} are reported in Table 3 followed by estimates of

Table 1: Percentage of convergent chains as per Geweke test

	G = 200					G = 100					G = 50				
λ	ψ	β	ν	λ	σ^2_{FR}	ψ	β	ν	λ	σ^2_{FR}	ψ	β	ν	λ	σ^2_{FR}
						$\boldsymbol{n_i = 4}$									
-3	93	93	92	91	93	92	91	89	91	92	92	88	92	93	93
-2	92	89	91	91	93	92	92	90	91	89	90	92	92	94	91
-1	95	92	94	91	90	94	92	93	93	94	92	91	91	91	89
1	91	93	93	91	91	91	90	92	92	89	91	90	90	92	89
2	91	91	92	91	93	91	92	92	93	90	92	91	91	88	90
3	88	92	93	91	91	93	93	94	92	93	92	92	91	88	91
						$\boldsymbol{n_i = 8}$									
-3	92	92	92	91	93	91	92	91	93	92	93	91	91	92	89
-2	93	91	91	93	90	92	93	94	90	91	92	91	93	92	91
-1	91	94	91	90	92	91	91	91	91	92	92	93	92	91	91
1	92	91	91	92	93	93	92	92	92	92	91	91	91	91	89
2	92	91	91	91	92	91	92	91	92	90	93	92	92	92	92
3	91	90	91	92	92	93	92	92	92	91	93	92	90	92	91

Table 2: Percentage of stationary chains according to Heidelberg and Welch test

	G = 200					G = 100					G = 50				
λ	ψ	β	ν	λ	σ^2_{FR}	ψ	β	ν	λ	σ^2_{FR}	ψ	β	ν	λ	σ^2_{FR}
						$\boldsymbol{n_i = 4}$									
-3	99	98	99	99	99	99	99	99	99	99	99	98	100	99	99
-2	100	99	99	100	99	99	99	99	100	99	99	100	99	99	99
-1	99	99	99	99	99	100	99	99	99	99	100	100	99	99	99
1	99	98	100	99	99	99	99	100	99	99	99	99	99	100	98
2	99	99	99	100	98	100	99	100	99	100	99	99	99	100	99
3	99	100	99	99	99	100	99	99	99	99	99	100	99	99	98
						$\boldsymbol{n_i = 8}$									
-3	99	100	99	99	100	99	99	99	99	100	99	100	99	100	99
-2	99	99	99	99	100	99	100	99	99	99	99	99	100	99	100
-1	99	99	98	98	99	99	99	99	99	99	99	99	98	99	98
1	99	99	99	99	99	99	99	98	99	98	99	100	99	99	99
2	100	99	99	100	99	100	99	99	100	98	99	99	99	100	99
3	99	98	99	99	99	99	99	99	99	99	99	99	99	99	98

the expected loss using SELF in Table 4.

For λ, the thinned sample was positively (negatively) skewed for $\lambda > 0$ (< 0), hence median was used to summarize the thinned sample for each of the 500 datasets and corresponding Bayes' estimate given by (13) was calculated. The corresponding bias in the estimates and estimate of expected loss using LINEX are reported in Table 5.
Further, to observe the consequence of ignoring skewness, same datasets were analyzed using log-normal shared frailty model. Table 6 gives the bias of estimates obtained using log normal shared frailty model (LN) and log skew-normal shared frailty model (LSN) under some of the setups considered in the simulation study.
On the basis of the simulation study (Tables 1-6), it was observed that

1. On an average, 91% of the Markov chains (among 500 datasets) converge for each parameter,

2. Not more than 2% Markov chains fail to achieve stationarity for each parameter,

3. In general, with an increase in size or number of clusters, the biases are reduced and the estimated loss is lower,

Table 3: Bias in the Bayesian estimates for parameters in simulation study

	G = 200				G = 100				G = 50			
λ	ψ	β	ν	σ^2_{FR}	ψ	β	ν	σ^2_{FR}	ψ	β	ν	σ^2_{FR}
						$\boldsymbol{n_i = 4}$						
-3	0.0056	-0.0026	-0.1268	0.0232	0.0083	-0.0003	-0.1254	0.0352	0.0337	-0.0128	-0.1247	0.0467
-2	0.0014	-0.0001	-0.0976	0.0139	0.0031	-0.0007	-0.1073	0.0232	0.0111	-0.0081	-0.1180	0.0313
-1	0.0033	-0.0021	0.0138	-0.0107	0.0100	-0.0048	-0.0398	-0.0063	0.0210	-0.0069	-0.0450	-0.0084
1	0.0016	-0.0032	-0.0258	0.0042	0.0032	-0.0023	0.0185	0.0596	0.0057	0.0017	0.0060	0.1120
2	-0.0047	0.0027	-0.0626	-0.0005	0.0003	0.0018	-0.0623	0.0173	0.0057	-0.0012	-0.0767	0.0498
3	-0.0022	0.0013	-0.0789	-0.0031	-0.0006	-0.0004	-0.0931	0.0036	0.0155	-0.0063	-0.0964	0.0371
						$\boldsymbol{n_i = 8}$						
-3	-0.0004	-0.0024	-0.0562	0.0105	0.0010	-0.0018	-0.0778	0.0101	0.0059	-0.0053	-0.0881	0.0157
-2	-0.0012	0.0013	-0.0378	0.0028	0.0034	0.0023	-0.0464	0.0054	0.0024	0.0043	-0.0594	0.0102
-1	0.0000	0.0007	0.0166	-0.0072	0.0004	0.0000	0.0197	-0.0086	0.0051	-0.0044	0.0145	-0.0064
1	0.0020	0.0004	0.0187	0.0207	0.0056	-0.0026	0.0250	0.0330	0.0071	-0.0041	0.0131	0.0439
2	-0.0010	-0.0009	-0.0486	-0.0046	-0.0046	0.0014	-0.0639	-0.0045	-0.0020	-0.0004	-0.0862	0.0015
3	0.0017	-0.0015	-0.0712	-0.0120	0.0033	-0.0035	-0.0917	-0.0132	0.0008	-0.0010	-0.1142	-0.0076

Table 4: Estimated loss using SELF for parameters in the simulation study

	G = 200				G = 100				G = 50			
λ	ψ	β	ν	σ^2_{FR}	ψ	β	ν	σ^2_{FR}	ψ	β	ν	σ^2_{FR}
						$\boldsymbol{n_i = 4}$						
-3	0.0045	0.0030	0.0219	0.0011	0.0093	0.0062	0.0259	0.0021	0.0180	0.0128	0.0322	0.0030
-2	0.0045	0.0028	0.0155	0.0012	0.0084	0.0052	0.0216	0.0018	0.0175	0.0116	0.0302	0.0020
-1	0.0048	0.0028	0.0077	0.0030	0.0104	0.0058	0.0143	0.0018	0.0202	0.0110	0.0260	0.0016
1	0.0043	0.0025	0.0073	0.0009	0.0094	0.0055	0.0185	0.0264	0.0168	0.0092	0.0289	0.1109
2	0.0044	0.0028	0.0127	0.0056	0.0087	0.0053	0.0176	0.0116	0.0182	0.0113	0.0284	0.0524
3	0.0046	0.0028	0.0144	0.0043	0.0087	0.0054	0.0204	0.0087	0.0179	0.0113	0.0296	0.0322
						$\boldsymbol{n_i = 8}$						
-3	0.0021	0.0013	0.0072	0.0007	0.0043	0.0025	0.0151	0.0016	0.0082	0.0051	0.0241	0.0036
-2	0.0018	0.0012	0.0049	0.0006	0.0039	0.0025	0.0096	0.0016	0.0088	0.0055	0.0202	0.0044
-1	0.0022	0.0014	0.0036	0.0012	0.0045	0.0028	0.0068	0.0025	0.0085	0.0052	0.0154	0.0059
1	0.0019	0.0012	0.0047	0.0031	0.0043	0.0027	0.0082	0.0079	0.0083	0.0048	0.0140	0.0171
2	0.0019	0.0013	0.0071	0.0019	0.0037	0.0023	0.0135	0.0047	0.0085	0.0052	0.0251	0.0108
3	0.0021	0.0014	0.0096	0.0017	0.0042	0.0026	0.0168	0.0035	0.0084	0.0054	0.0288	0.0080

Table 5: Bias and estimated loss (EL) using LINEX for skewness parameter

	G = 200		G = 100		G = 50	
λ	Bias	EL	Bias	EL	Bias	EL
			$\boldsymbol{n_i = 4}$			
-3	-0.0255	(0.1768)	0.1598	(0.2326)	0.5651	(0.2774)
-2	-0.7255	(0.1324)	-0.5912	(0.1340)	-0.2765	(0.1109)
-1	-0.5740	(0.0597)	-0.6737	(0.0903)	-0.7085	(0.0975)
1	0.3165	(0.0506)	0.5587	(0.0558)	0.4914	(0.0412)
2	-0.2696	(0.0914)	-0.3846	(0.0728)	-0.4822	(0.0660)
3	-1.1058	(0.3763)	-1.3504	(0.3993)	-1.4367	(0.4037)
			$\boldsymbol{n_i = 8}$			
-3	0.6500	(0.3121)	1.1049	(0.3693)	1.3326	(0.3928)
-2	-0.1867	(0.1245)	0.1335	(0.0929)	0.2823	(0.0769)
-1	-0.4206	(0.0569)	-0.5108	(0.0599)	-0.5197	(0.0506)
1	0.5378	(0.0729)	0.5758	(0.0702)	0.4846	(0.0468)
2	-0.0201	(0.1270)	-0.1499	(0.1010)	-0.4024	(0.0715)
3	-0.7329	(0.3334)	-1.1524	(0.3579)	-1.3728	(0.3960)

Table 6: Bias in estimates obtained using Log Skew-Normal (LSN) vs. Log normal (LN) frailty model

		G = 200				G =100			
λ		ψ	β	ν	σ^2_{FR}	ψ	β	ν	σ^2_{FR}
					$\boldsymbol{n_i = 4}$				
-3	LN	0.0057	-0.0025	-0.1713	0.0348	0.0088	-0.0010	-0.1806	0.0388
	LSN	0.0056	-0.0026	-0.1268	0.0232	0.0083	-0.0003	-0.1254	0.0352
-1	LN	0.0065	-0.0038	-0.0927	0.0192	0.0018	-0.0001	-0.0862	0.0161
	LSN	0.0033	-0.0021	0.0138	-0.0107	0.0100	-0.0048	-0.0398	-0.0063
					$\boldsymbol{n_i = 8}$				
-3	LN	0.0005	-0.0027	-0.1694	0.0294	0.0006	-0.0018	-0.1813	0.0274
	LSN	-0.0004	-0.0024	-0.0562	0.0105	0.0010	-0.0018	-0.0778	0.0101
-1	LN	0.0018	-0.0001	-0.0862	0.0161	0.0025	-0.0007	-0.0912	0.0177
	LSN	0.0000	0.0007	0.0166	-0.0072	0.0004	0.0000	0.0197	-0.0086

4. Estimates of ψ and β are not affected significantly by degree of skewness. However, estimates of ν improve when λ is closer to 0,

5. If skewness is ignored during estimation, there is no significant effect on the estimates of ψ or β. However, estimates of ν and σ^2_{FR} improve when skewness is considered.

From the simulation study, it can be concluded that a larger dataset improves estimates of parameters as the estimated loss reduces with an increase in the number of clusters or cluster sizes. Also, skewness in the data affects only estimates of the frailty variables. Hence, the size of study should be determined keeping in mind the required accuracy. Also, the log skew-normal model should be considered only when the frailty distribution is of interest and there may be numerous unknown covariates affecting survival times that could not have been considered in the model for any reason such as budget or time constraints, destructive nature of recording etc.

7. Examples

Log skew-normal (LSN) shared frailty model is illustrated using two datasets where early and late dependence is observed in the survival times. Sect. 7.1 discusses the application of the model to exercise dataset by Danahy, Burwell, Aronow, and Prakash (1977) where the random effect is considered to be positively skewed. In Sect. 7.2, the model is applied to catheter infection data of kidney dialysis patients by McGilchrist and Aisbett (1991) where negative skewness is considered in the frailty effect. The Bayesian estimates obtained for LSN shared frailty model are compared to the those obtained with a parametric approach using the Marquardt algorithm for log-norm (LN) and Gamma shared frailty models along with a simple Cox proportional hazard (Cox PH) model. The parametric estimates are obtained using 'frailtypack' (Rondeau, Marzroui, and Gonzalez 2012) R package. In order to make a valid comparison between Bayesian and parametric approaches, the Monte Carlo standard error (MCSE) for Bayesian estimates is obtained using a thinned sample from each Markov chain to account for the dependence in posterior samples for each parameter. MCSEs are then compared to the standard errors of the parametric estimates.

7.1. Exercise data

The effect of high dose oral isosorbide dinitrate (ISDN) on 21 patients with coronary atherosclerotic heart disease was studied by measuring the patients' heart rate and blood pressure during exercise tests till the onset of angina pectoris. For each patient, there are 10 exercise times

recorded after sublingual placebo (SLP), sublingual nitroglycerin (SLTNG), at 1, 3, & 5 hours after oral placebo (OP1, OP3, OP5) and its control period (OPC) and at 1, 3 & 5 hours after oral ISDN (OI1, OI3, OI5) and its control period (OIC).

If the test was incomplete due to exhaustion, the corresponding observations were considered to be censored. The likelihood as given in (8) assumes that the distribution of censoring times is non-informative, that is, it carries no information about the parameters of interest which in this case is the effect of ISDN on time to angina pectoris. On the other hand, exhaustion that led to censoring, can be thought of as being related to an individual's frailty, with more frail individuals being more likely to experience failure earlier and/or leave the test incomplete. Thus, it becomes important to not only look at treatment effects but also the dependence structure in the event times in the model.

Shared frailty model can be used to analyze this dataset since observations in the same cluster are for the same patient and are hence correlated. Here, since the patients are diseased, the survival times are expected to have early dependencies. The same has been verified by using the built-in function 'frailtyPenal' in the 'frailtypack' (Rondeau *et al.* 2012) R package where the skewness of the predicted frailty term under Gamma frailty and Weibull baseline distribution is calculated as 1.8192. The log skew-normal frailty model is fitted to the exercise dataset with a Weibull baseline distribution and positive skewness ($\lambda > 0$).

The algorithm given in Sect. 5.3 is used to estimate nine treatment effects and parameters of baseline and frailty distribution. Initially, it is run for fewer set of iterations in order to determine standard deviations of the proposal distributions 's' so that approximately 25% proposed values may be accepted for each parameter. Subsequently, using two different sets of starting values, two chains of 1,20,000 iterations are obtained for each parameter.

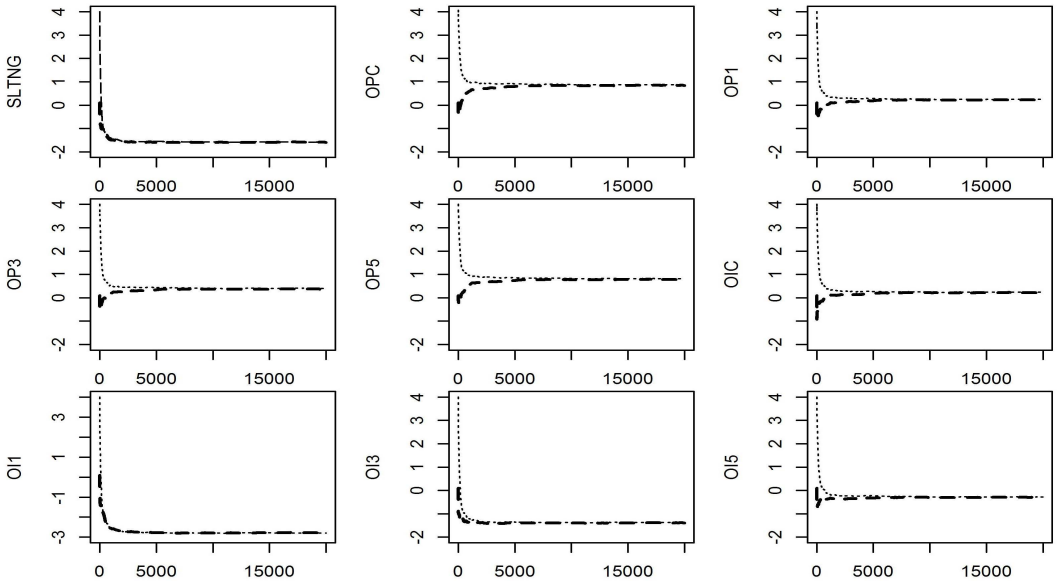

Figure 4: Cumulative mean plots for Exercise data: The initial 20,000 iterations for two chains with different set of initial values

Cumulative Mean (CM) plots (Fig. 4) are plotted using the two chains obtained as an outcome of using two different sets of starting values. The chains converge after approximately the first 5000 iterations, indicating that final estimates do not depend upon the starting values used in the algorithm. Different starting values help to check whether the likelihood function has multiple modes. Observing the initial iterations from the CM plots for all parameters, the first 20,000 iterations from each chain are removed as burn-in.
The trace plots (Fig. 5) indicate a well mixing chain while the ACF plots (Fig. 6) check for auto-correlation in the chain. At a lag of 150, auto-correlation reduces significantly (< 0.1).

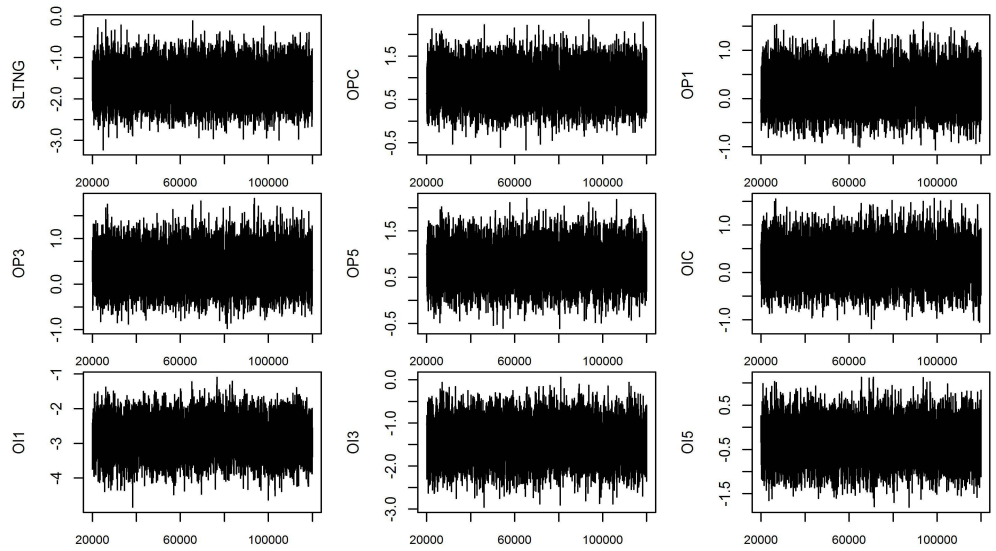

Figure 5: Trace plots for Exercise data: Posterior sample for treatment effects

Hence the chain is thinned by sampling every 150^{th} value.

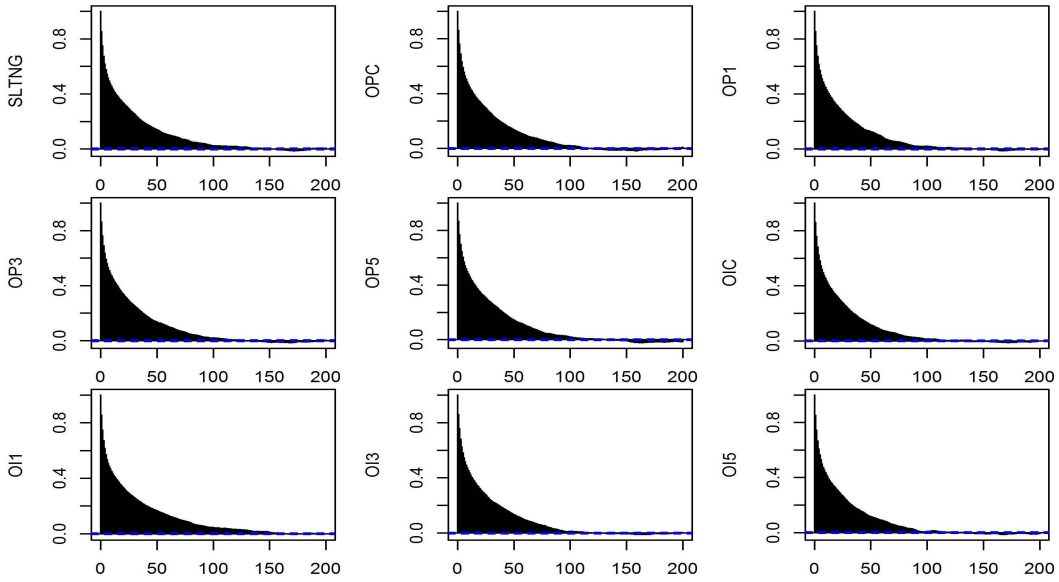

Figure 6: ACF plots for Exercise data: Auto-correlation in posterior sample for treatment effects

The density plots (Fig. 6) obtained for the posterior sample of treatment effects suggest a symmetric sample. Hence, they are summarized using mean of the thinned sample to find $\widehat{\beta}$.

Table 7 gives the estimates, credible intervals, and diagnostic tests results obtained for LSN shared frailty model along with the selected models for comparison. From the results it can be concluded that

1. while treatments SLTNG and ISDN (OI1, OI3, OI5) reduce the risk of angina pectoris ($\widehat{\beta} < 0$), the treatment effects SLTNG, OI1, and OI3 are significant. Hence, patients under SLTNG and ISDN have significantly less chance of angina pectoris specially after one and three hours of ISDN treatment,
2. $\widehat{\beta}$ corresponding to treatment effects OPC and OP5 are positive indicating that oral

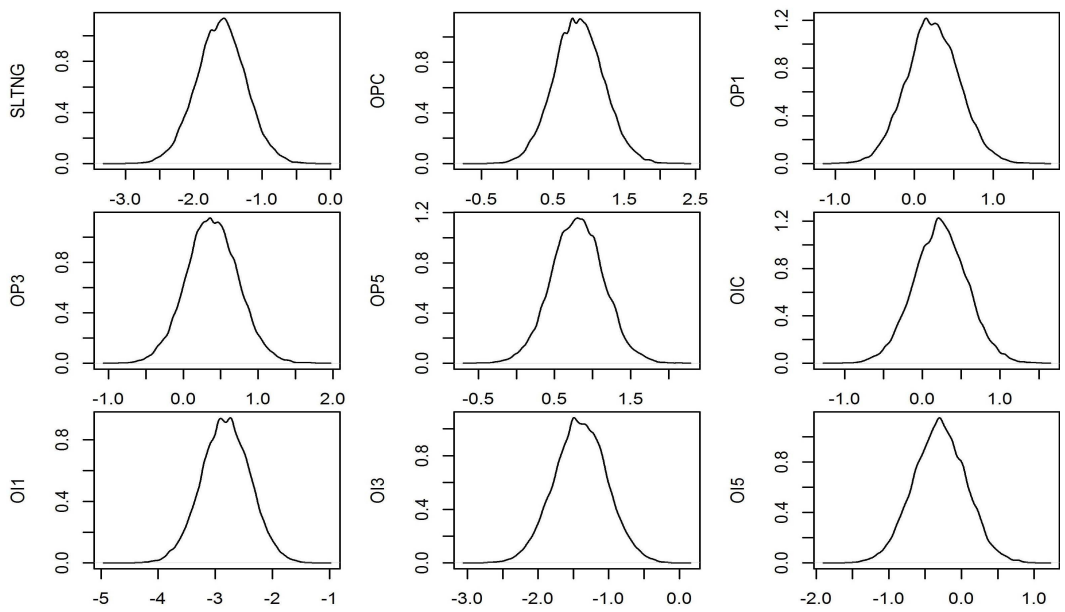

Figure 7: The density plots for treatment effects of the Exercise data

Table 7: Estimates and standard errors for parameters of LSN, LN, Gamma shared frailty models and simple Cox PH model for Exercise data

	LSN					LN	Gamma	Cox PH
Parameter	Estimate	Diagnostic Tests*		Credible Limits		Estimate	Estimate	Estimate
	(SE)	Geweke	H-W	Lower	Upper	(SE)	(SE)	(SE)
SLTNG	-1.6070	0.41	0.32	-2.3251	-0.9346	-1.5364	-1.5085	-0.8146
	(0.0139)					(0.3417)	(0.3401)	(0.3217)
OPC	0.8316	0.18	0.69	0.1107	1.5170	0.7142	0.6992	0.2334
	(0.0140)					(0.3290)	(0.3292)	(0.3102)
OP1	0.2303	0.33	0.05	-0.4500	0.8657	0.1365	0.1388	-0.0359
	(0.0131)					(0.3201)	(0.3199)	(0.3093)
OP3	0.3843	0.11	0.43	-0.3550	1.0734	0.2822	0.2841	-0.0065
	(0.0136)					(0.3284)	(0.3282)	(0.3096)
OP5	0.8052	0.55	0.54	0.1280	1.4907	0.6531	0.6423	0.1251
	(0.0136)					(0.3256)	(0.3252)	(0.3096)
OIC	0.2238	0.88	0.37	-0.4237	0.9243	0.1424	0.1490	-0.0195
	(0.0133)					(0.3254)	(0.3254)	(0.3088)
OI1	-2.8110	0.41	0.31	-3.6890	-2.0055	-2.6559	-2.6492	-1.4337
	(0.0170)					(0.4067)	(0.4063)	(0.3622)
OI3	-1.3980	0.24	0.26	-2.1519	-0.6516	-1.3777	-1.3718	-0.9478
	(0.0144)					(0.3553)	(0.3553)	(0.3342)
OI5	-0.3057	0.24	0.26	-1.0492	0.4208	-0.3564	-0.3568	-0.4385
	(0.0144)					(0.3450)	(0.3456)	(0.3189)
ψ	5.388	0.62	0.44	4.7845	6.0447	4.91	1.1500	
ν	2.4500	0.06	0.29	1.7181	3.5370			
λ	0.9151	0.05	0.26	0.0264	6.6069			
σ^2_{FR}	12.596						2.4871	3.90798
	11.27						(0.6888)	(1.3290)

*p-values

placebo increases the risk of heart disease significantly at time zero and five hours after using a placebo,

3. estimate of the baseline shape parameter indicates that the hazard function increases with time,

4. $\hat{\lambda}$ is estimated to be 0.9151 indicating low degree of positive skewness in the frailty parameter.
5. Comparatively, besides the parametric estimates having higher standard errors, the chosen models also fail to incorporate an evident dependence structure in the frailty parameter. The models also seem to be underestimating frailty variance.

7.2. Catheter infection data

The repeated occurrence of infection in kidney patients using a portable dialysis machine was studied where the observed event times are times from insertion of the catheter to time of occurrence of infection. The catheter is subsequently removed and reinserted after infection has cleared up and a second time to infection is noted. The dataset has 38 patients with two recordings of times to infection for each. The data is right censored for cases where the catheter is removed for other reasons or the final recurrence time is not noted. The covariates noted are age, sex, presence or absence of three disease types, Glomerulo Nephritis (GN), Acute Nephritis (AN), and Polycystic Kidney Disease (PKD).

The observations in the same cluster are times to infection for the same individual and hence are considered to share a common risk or frailty. Since infection is expected to build up with time and there is no infection at time of insertion of the catheter, we expect the survival times to have late dependencies. To verify this, the skewness of the predicted frailty term under Gamma frailty and Weibull baseline distribution is calculated to be -0.3354 using built-in functions (Rondeau *et al.* 2012) in R. Hence, the LSN shared frailty model with negative skewness ($\lambda < 0$) provides a fit to the dataset.

There are five treatment effects to be estimated (Age, Sex, GN, AN, and PKD) along with baseline and frailty parameters. The behaviour of Markov chains is studied using a similar procedure as given in Sect. 7.1. The posterior summaries are presented in Table 8.

Table 8: Estimates and standard errors for parameters of LSN, LN, Gamma shared frailty models and simple Cox PH model for the catheter infection data

	LSN					LN	Gamma	Cox PH
Parameter	Estimate	Diagnostic Tests*		Credible Limits		Estimate	Estimate	Estimate
	(SE)	Geweke	H-W	Lower	Upper	(SE)	(SE)	(SE)
Age	0.0027 (0.0007)	0.21	0.10	-0.0309	0.0290	0.0029 (0.0134)	0.0025 (0.0134)	0.0036 (0.0111)
Sex	-2.0730 (0.0276)	0.50	0.14	-3.1341	-0.9802	-1.8565 (0.4748)	-1.9081 (0.5299)	-1.4283 (0.3316)
GN	0.1969 (0.0278)	0.48	0.48	-0.9571	1.4668	0.1147 (0.4912)	0.1455 (0.5048)	0.2615 (0.4156)
AN	0.7312 (0.0277)	0.58	0.10	-0.3388	1.9348	0.5980 (0.4913)	0.6194 (0.5079)	0.4624 (0.4030)
PKD	-0.9063 (0.0431)	0.90	0.57	-2.6750	0.7582	-1.1645 (0.7669)	-0.9746 (0.9632)	-1.2423 (0.6304)
ψ	1.2460	0.12	0.62	0.9542	1.6225	1.1494	1.1624	
ν	1.0740	0.11	0.49	0.1931	2.1184			
λ	-4.3613	0.88	0.85	-8.3873	-0.0714			
σ^2_{FR}	0.5469 (0.0829)						0.2773 (0.3466)	0.2820 (0.3362)

From Table 8, the following conclusions can be made:

1. Sex of the patient appears to be a significant factor with negative $\hat{\beta}$, implying that female patients have a lower hazard rate, that is, lower risk of infections,

2. The hazard function is increasing in time since baseline shape parameter is estimated to be more than one,
3. $\hat{\lambda}$ is estimated to be -4.3613 indicating a moderate degree of negative skewness in the frailties.
4. Bayesian estimates for the LSN shared frailty model have lower standard errors and also highlights a moderate late dependence in survival times, a fact that the other models in comparison fail to report.

8. Conclusions

Bayesian estimation procedure is suggested for log skew-normal shared frailty model using the Metropolis-Hastings algorithm. In the simulation study, estimation procedure is applied to data simulated from a Weibull baseline and a skewed random effect following log skew-normal distribution. Besides the procedure performing better for larger datasets, it is also observed that random effect is estimated more efficiently when the skewness factor is considered in the model. However, there is no significant improvement in the estimation of treatment effects. Therefore, due to the complicated nature of estimation procedure and ready availability of software packages for other popular models, this method is advisable only when the random effect is of interest to the researcher.

The model has been fitted to the exercise dataset (Danahy *et al.* 1977) and catheter infection dataset (McGilchrist and Aisbett 1991) where the skewness parameter is found to be low to moderate. When there is early/late dependence in the random effect, common unknown risk factors need to be carefully studied further. In this case, changes need to be made to the model like inclusion or exclusion of covariates, choice of a more suitable baseline, removal of outliers or considering multiple populations with different hazard functions.

Bayesian estimation of the LSN model is also compared to the parametric log-normal and Gamma shared frailty models along with the simple Cox PH model. Other than having lower standard errors the LSN model is able to highlight the dependence in survival times, something that the other models fail to incorporate.

Acknowledgments

The authors are grateful to the Department of Science and Technology (DST), Government of India for grant provided under PURSE. The first author also acknowledges the financial support provided by the UGC, India.

References

Azzalini A (1985). "A Class of Distributions which Includes the Normal Ones." *Scandinavian Journal of Statistics*, pp. 171–178.

Azzalini A, Capitanio A (1999). "Statistical Applications of the Multivariate Skew Normal Distribution." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(3), 579–602. doi:10.1111/1467-9868.00194.

Azzalini A, Cappello TD, Kotz S (2003). "Log-skew-normal and Log-skew-t Distributions as Models for Family Income Data." *Journal of Income Distribution*, **11**(3), 12. doi:10.25071/1874-6322.1249.

Bayes CL, Branco MD (2007). "Bayesian Inference for the Skewness Parameter of the Scalar Skew-normal Distribution." *Brazilian Journal of Probability and Statistics*, **21**(2), 141–163.

Callegaro A, Iacobelli I (2012). "The Cox Shared Frailty Model with Log-skew-normal Frailties." *Statistical Modelling*, **12**(5), 399–418. doi:10.1177/1471082X12460146.

Canale A, Scarpa B (2013). "Informative Bayesian Inference for the Skew-normal Distribution." *arXiv preprint arXiv:1305.3080*. doi:10.48550/arXiv.1305.3080.

Clayton DG (1978). "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence." *Biometrika*, **65**(1), 141–151. doi:10.1093/biomet/65.1.141.

Cox DR (1972). "Regression Models and Life-tables." *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**(2), 187–220. doi:10.1111/j.2517-6161.1972.tb00899.x.

Danahy DT, Burwell DT, Aronow WS, Prakash R (1977). "Sustained Hemodynamic and Antianginal Effect of High Dose Oral Isosorbide Dinitrate." *Circulation*, **55**(2), 381–387. doi:10.1161/01.cir.55.2.381.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis. ISBN 9781439840955.

Gelman A, Roberts GO, Gilks WR (1996). "Efficient Metropolis Jumping Rules." In *Bayesian Statistics, 5 (Alicante, 1994)*, volume 5, pp. 599–607. Oxford University Press.

Geweke J (1992). "Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments." *Bayesian Statisitcs*, **4**, 169–188. doi:10.21034/sr.148.

Hanagal DD, Dabade AD (2013). "Bayesian Estimation of Parameters and Comparison of Shared Gamma Frailty Models." *Communications in Statistics - Simulation and Computation*, **42**(4), 910–931. doi:10.1080/03610918.2012.661909.

Heidelberger P, Welch PD (1983). "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research*, **31**(6), 1109–1144. doi:10.1287/opre.31.6.1109.

Hutson AD, Mashtare Jr TL, Mudholkar GS (2020). "Log-epsilon-skew Normal: A Generalization of the Log-normal Distribution." *Communications in Statistics-Theory and Methods*, **49**(17), 4197–4215. doi:10.1080/03610926.2019.1595655.

Ibrahim JG, Chen MH, Sinha D (2001). *Bayesian Survival Analysis*. Springer-Verlag., Berlin.

Liao X, Peng Z, Nadarajah S (2013). "Tail Properties and Asymptotic Expansions for the Maximum of the Logarithmic Skew-normal Distribution." *Journal of Applied Probability*, **50**(3), 900–907. doi:10.1239/jap/1378401246.

Lin GD, Stoyanov J (2009). "The Logarithmic Skew-normal Distributions Are Moment-indeterminate." *Journal of Applied Probability*, **46**(3), 909–916. doi:10.1239/jap/1253279858.

Liseo B, Loperfido N (2004). "Default Bayesian Analysis of the Skew-normal Distribution." *Journal of Statistical Planning and Inference*, **136**(2), 373–389.

Liseo B, Loperfido N (2006). "A Note on Reference Priors for the Scalar Skew-normal Distribution." *Journal of Statistical Planning and Inference*, **136**(2), 373–389. doi:10.1016/j.jspi.2004.06.062.

Martinez-Florez G, Vergara-Cardozo S, Gonzalez L (2013). "The Family of Log-Skew-Normal Alpha-Power Distributions Using Precipitation Data." *Revista Colombiana de Estadistica*, **36**, 43–57.

McGilchrist CA, Aisbett CW (1991). "Regression with Frailty in Survival Analysis." *Biometrics*, **47**(2), 461–466. doi:10.2307/2532138.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953). "Equation of State Calculations by Fast Computing Machines." *Journal of Chemical Physics*, **21**, 1087–1092. doi:10.1063/1.1699114.

Metropolis N, Ulam S (1949). "The Monte Carlo Method." *Journal of the American Statistical Association*, **44**(247), 335–341. doi:10.2307/2280232.

Oakes D (1989). "Bivariate Survival Models Induced by Frailties." *Journal of the American Statistical Association*, **84**(406), 487–493. ISSN 01621459. doi:10.2307/2289934. URL http://www.jstor.org/stable/2289934.

Plummer M, Best N, Cowles K, Vines K (2006). "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News*, **6**(1), 7–11. URL https://journal.r-project.org/archive/.

Rondeau V, Marzroui Y, Gonzalez J (2012). "frailtypack: An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation." *Journal of Statistical Software, Articles*, **47**(4), 1–28. ISSN 1548-7660. doi:10.18637/jss.v047.i04. URL https://www.jstatsoft.org/v047/i04.

Sahu SK, Dey DK, Aslanidou H, Sinha D (1997). "A Weibull Regression Model with Gamma Frailties for Multivariate Survival Data." *Life Time Data Analysis*, **3**, 123–137. doi:10.1023/a:1009605117713.

Santos C, Achcar J (2011). "A Bayesian Analysis in the Presence of Covariates for Multivariate Survival Data: An Example of Application." *Revista Colombiana de Estadistica*, **34**, 111–131.

Sartori N (2006). "Bias Prevention of Maximum Likelihood Estimates for Scalar Skew Normal and Skew t Distributions." *Journal of Statistical Planning and Inference*, **136**(12), 4259–4275. doi:10.1016/j.jspi.2005.08.043.

Sidhu S, Jain K, Sharma SK (2018). "Bayesian Estimation of Generalized Gamma Shared Frailty Model." *Computational Statistics*, **33**(1), 277–297. doi:10.1007/s00180-017-0728-0.

Smyth GK (2005). "Numerical Integration." *Encyclopedia of Biostatistics*, pp. 3088–3095. doi:10.1002/0470011815.b2a14026.

Varian HR (1975). "A Bayesian Approach to Real Estate Assessment." *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, pp. 195–208.

Zellner A (1986). "Bayesian Estimation and Prediction Using Asymmetric Loss Functions." *Journal of the American Statistical Association*, **81**(394), 446–451. doi:10.2307/2289234.

Affiliation:

Sukhmani Sidhu
Department of Statistics
Panjab University
Chandigarh, India - 160014
E-mail: sukhmani.15@gmail.com

Austrian Journal of Statistics
published by the Austrian Society of Statistics
Volume 52
July 2023

http://www.ajs.or.at/
http://www.osg.or.at/

Submitted: 2022-01-27
Accepted: 2022-07-26