Statistics and Applications {ISSN 2454-7395 (online)}
Volume 19, No. 2, 2021 (New Series), 41-62

Generalized Lindley Shared Frailty Models

Arvind Pandey[1], **David D. Hanagal**[2] **and Shikhar Tyagi**[1]

[1]*Department of Statistics, Central University of Rajasthan, Rajasthan, India*
[2]*Department of Statistics, Savitribai Phule Pune University, Pune-411007, India.*

Received: 03 June 2020; Revised: 22 July 2020; Accepted: 24 July 2020

Abstract

We propose a new frailty distribution named as the generalized Lindley frailty distribution with generalized Weibull and exponential power as baseline distributions. To estimate the parameters in the models, the Bayesian paradigm of the Markov Chain Monte Carlo technique was designed. Bayesian comparison techniques have been performed for the comparison of models. We analyze kidney infection data and suggest a better model.

Key words: Bayesian estimation; Exponential power distribution; Generalized Lindley frailty; Generalized Weibull distribution; MCMC; Random censoring.

1. Introduction

In survival data, a common approach is that each individual under study experiencing the same risk factors which act as multiplicatively. Sometimes, in real-life situations risk (hazard rate) changes from one family to another family, one group to another group, one cluster to another cluster. Heterogeneity in the population exists, because of the mixture of groups of individuals with different risk factors. This heterogeneity is called as a frailty. Ignoring frailty may have adverse consequences. A random impact that is unobservable risk shared by the subject characterized as frailty which was introduced by Vaupel *et al.* (1979). To handle such kind of problems, many models have been derived in survival analysis. Since the establishment of the proportional hazard model given by Cox (1972), survival function has been dominated by hazard rate models. The reason behind the popularity of this model is, the significance of known covariates can be tested, also a relationship between lifetimes and covariates can be incorporated. Cox (1972) gave the following proportional hazard model or multiplicative hazard model as

$$\phi(t|\underline{K}) \;=\; \phi_0(t)e^{\underline{K}^{'}\underline{\beta_0}} \tag{1}$$

where, $\phi(t|\underline{K})$ stands for conditional hazard rate given the covariates, $\phi_0(t)$ stands for baseline hazard rate. $\underline{K}^{'} = (K_{1j}, K_{2j}, ..., K_{mj})$ are vector of known covariate and $\underline{\beta_0}$ is the vector of

Corresponding author: David D. Hanagal
E-mail: david.hanagal@gmail.com

regression parameters of order m corresponding to $\underline{K}$. Augmentation of Cox's proportional hazard model provided away to introduce the unknown covariates,

$$\phi(t|\underline{V}) = \phi_0(t)e^{\underline{K}^{'}\underline{\beta_0}+\underline{V}^{'}\underline{\beta_1}}$$

or

$$\phi(t|w) = w\phi_0(t)e^{\underline{K}^{'}\underline{\beta_0}} \tag{2}$$

where, $\underline{V}^{'} = (V_{1j}, V_{2j}, ..., V_{mj})$ are considered as the vector of unknown covariates respectively, $\underline{\beta_1}$ are indicated as the vector of regression coefficients of order m corresponding to $\underline{V}$. $w = e^{\underline{V}^{'}\underline{\beta_1}}$ called as frailty random effect. The conditional cumulative hazard function is given by

$$\Phi(t|w) = w\Phi_0(t)e^{\underline{K}^{'}\underline{\beta_0}} \tag{3}$$

where $\Phi_0(t) = \int_0^t \phi_o(t)dt$. The conditional survival function is given by

$$S(t|w) = \exp\left(-w\Phi_0(t)e^{\underline{K}^{'}\underline{\beta_0}}\right) \tag{4}$$

Frailty models firstly introduced by Vaupel *et al.* (1979) in univariate survival models that can be separated into multiplicative components. It has been assumed that the baseline hazard function has a multiplicative effect of frailties. Several frailty models had been proposed by Oakes (1989). As a frailty distribution, gamma, inverse Gaussian, positive stable distributions had been claimed by Hougaard (1986). Hougaard (1985, 1991, 2000) had discussed the different aspects of frailty on a broad scale. Log-normal distribution was proposed as frailty distribution by Flinn and Hackman (1982). In the last decade, frailty regression models in mixture distribution have been discussed by Hanagal (2008). Hanagal and Dabade (2013, 2015) proposed modeling of the inverse Gaussian frailty model and comparison of different frailty models for analyzing kidney infection data. Modeling kidney infection data for inverse Gaussian shared frailty was done by Hanagal and Pandey (2014a). Gamma frailty models for bivariate survival data were given by Hanagal and Pandey (2015a). Hanagal and Pandey (2017a) were used the shared inverse Gaussian frailty models based on additive hazard. For reversed hazard rate setup, Hanagal and Pandey (2014b, 2015b, 2016a, 2016b, 2017b) have contemplated gamma and inverse Gaussian shared frailty models with different baseline distribution functions. Hanagal and Sharma (2013, 2015a, 2015b, 2015c) analyzed acute leukemia data, kidney infection data and diabetic retinopathy data using shared gamma and inverse Gaussian frailty models for the multiplicative model. Compound Poisson frailty was used by Hanagal and Kamble (2015) for Bayesian estimation. Analysis of kidney infection data and Australian twin data were done by Hanagal and Bhambure (2014, 2015, 2016) with different frailty distributions. Hanagal (2011, 2017, 2019) gave extensive literature review on different shared frailty models.

The main aim of this article has three objectives. First, generalized Lindley (GL) shared frailty models for hazard rate with generalized Weibull and exponential power as baseline distributions have been introduced. Second, Bayesian approach of estimation has been employed to estimate the unknown parameters under random censoring. Third, simulation study and data analysis have been done for the kidney infection data set.

2. Generalized Lindley Frailty Model

Lindley (1958) proposed a distribution with one parameter. Because of having only one parameter, the Lindley distribution does not provide enough flexibility for modeling purposes. It will be useful to consider further alternatives of this distribution. Zakerzadeh and Dolati (2009) proposed generalized Lindley distribution which generalizes Lindley distribution and includes exponential and gamma distributions as special cases. For a frailty distribution, generalized Lindley (GL) distribution has been considered in this paper. This distribution is the mixture of two gamma distributions G(θ,μ) and G(θ,η) with mixing coefficient $\theta/(\theta+1)$. That is the reason why GL frailty model is more adaptable in comparison with gamma frailty model. Probability density function of GL distribution has been specified below:

$$f_W(w) = \begin{cases} \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}w^{\mu-1}}{\Gamma\mu} + \frac{\theta^{\eta}w^{\eta-1}}{\Gamma\eta}\right]e^{-\theta w} & ; w \in \mathbb{R}^+, \mu, \eta, \theta \in \mathbb{R}^+ \\ 0 & ; \textit{otherwise} \end{cases}$$

with mean $E[W] = \frac{1}{1+\theta}\left[\mu + \frac{\eta}{\theta}\right]$. And corresponding variance is,

$$V(W) = \frac{1}{(1+\theta)}\left[\left(\mu^2 + \frac{\eta^2}{\theta}\right)\left(\frac{1}{\theta(1+\theta)}\right) + \left(\frac{\mu+\eta}{\theta}\right) - \left(\frac{2\mu\eta}{\theta(1+\theta)}\right)\right]$$

after applying identifiability property, *i.e.*, $E[W] = 1$ we get a relation between parameters $\eta = \theta\,(1+\theta-\mu) > 0$. Consequently, the density function, Laplace transformation and variance for GL reduced to,

$$f_W(w) = \begin{cases} \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}w^{\mu-1}}{\Gamma\mu} + \frac{\theta^{\theta(1+\theta-\mu)}w^{\theta(1+\theta-\mu)-1}}{\Gamma\theta(1+\theta-\mu)}\right]e^{-\theta w} & ;\ w, \theta \in \mathbb{R}^+, \mu \in (0, 1+\theta) \\ 0 & ;\ otherwise. \end{cases}$$

$$L_W(s) = \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}}{(s+\theta)^{\mu}} + \frac{\theta^{\theta(1+\theta-\mu)}}{(s+\theta)^{\theta(1+\theta-\mu)}}\right] \tag{5}$$

$$V(W) = \frac{\theta^4 - \theta^3\mu + 3\theta^2(1+\theta) - 4\theta^2\mu + 3\theta\mu(\mu-1) + \mu^2}{\theta(1+\theta)^2} \tag{6}$$

n objects are postulated to be under study. (T_{1j}, T_{2j}) are contemplated as first and second survival time of $i^{th}(i = 1, 2)$ component of $j^{th}(j = 1, 2, ..., n)$ objects. The unconditional bivariate survival function at time $t_{1j} \in \mathbb{R}^+$ and $t_{2j} \in \mathbb{R}^+$ can be written as,

$$\begin{aligned} S(t_{1j}, t_{2j}) &= \int_{w_j \in \mathbb{R}^+} S(t_{1j}, t_{2j}|w_j) f_W(w_j) dw_j \\ &= \int_{w_j \in \mathbb{R}^+} e^{-W_j(\Phi_{01}(t_{1j}) + \Phi_{02}(t_{2j}))\rho_j} f_W(w_j) dw_j \\ &= L_{W_j}\left[(\Phi_{01}(t_{1j}) + \Phi_{02}(t_{2j}))\,\rho_j\right] \end{aligned} \tag{7}$$

where, $L_{W_j}(.)$ is Laplace transformation of frailty variable W_j. $\Phi_0(.)$ stands for cumulative baseline hazard rate and $\rho_j = e^{\underline{K}_j^{'}\underline{\beta}_j}$ is the therm containing the regression coefficients corresponding to known covariates. To get unconditional survival function, using equations (5) and (7),

$$S(t_{1j}, t_{2j}) = \frac{1}{(1+\theta)}\left[\frac{\theta^{\mu+1}}{(\theta+\rho(\Phi_{01}(t_{1j})+\Phi_{02}(t_{2j})))^{\mu}} + \frac{\theta^{\theta(1+\theta-\mu)}}{(\theta+\rho(\Phi_{01}(t_{1j})+\Phi_{02}(t_{2j})))^{\theta(1+\theta-\mu)}}\right] \tag{8}$$

corresponding cross-ratio function given by Clayton (1978) and Oakes (1989) is given by,

$$\theta^{*}(t_{1j}, t_{2j}) = \frac{A * B}{\left(\theta^{\theta(\theta+1)}(-\mu+\theta+1)(C*\rho+\theta)^{\mu} + \mu\theta^{\mu(\theta+1)}(C*\rho+\theta)^{\theta(-\mu+\theta+1)}\right)^{2}}$$

where,

$$A = \theta^{\mu\theta-1}\left(\theta^{\theta(-\mu+\theta+1)}\left(\left(\Phi_{01}(t_{1j})+\Phi_{02}(t_{2j})\right)\rho+\theta\right)^{\mu} + \theta^{\mu+1}\left(\left(e^{\lambda_1 t_{1j}{}^{\alpha_1}} + e^{\lambda_2 t_{2j}{}^{\alpha_2}} - 2\right)\rho+\theta\right)^{\theta(-\mu+\theta+1)}\right)$$

$$B = \left(\theta^{\theta^2+\theta+1}(\mu^2 - 2\mu(\theta+1) + \theta(\theta+2)+2) - (\mu-1)\theta^{\theta(\theta+1)}\right)\left(\left(\Phi_1(t_1)+\Phi_2(t_2)\right)\rho+\theta\right)^{\mu} + \mu(\mu+1)\theta^{\mu(\theta+1)}\left(\left(\Phi_1(t_1)+\Phi_2(t_2)\right)\rho+\theta\right)^{\theta(-\mu+\theta+1)}$$

$$C = \left(\left(\Phi_{01}(t_{1j})+\Phi_{02}(t_{2j})\right)\right)$$

in the absence of frailty effect, model in the hazard rate setup will be,

$$S(t_{1j}, t_{2j}) = \exp\left(-\rho_j(\Phi_{01}(t_{1j})+\Phi_{02}(t_{2j}))\right) \tag{9}$$

One can have different baseline distributions for T_1 and T_2. After substituting different cumulative hazard functions in (8), we get different generalized Lindley frailty distributions.

3. Baseline Distributions

3.1 Generalized Weibull distribution

Here, the generalized Weibull distribution has been postulated as a baseline distribution. If a continuous random variable T follows the generalized Weibull distribution then the survival, hazard, and cumulative hazard function, are respectively,

$$S(t) = \begin{cases} 1-\left(1-e^{-\delta t^{\xi}}\right)^{\zeta} & ; t \in \mathbb{R}^{+}, \delta, \zeta, \xi \in \mathbb{R}^{+} \\ 1 & ; \textit{otherwise} \end{cases} \tag{10}$$

$$\phi_0(t) = \begin{cases} \frac{\xi\zeta\delta t^{\xi-1}e^{-\delta t^{\xi}}(1-e^{-\delta t^{\xi}})^{\zeta-1}}{1-\left(1-e^{-\delta t^{\xi}}\right)^{\zeta}} & ; t \in \mathbb{R}^{+}, \delta, \zeta, \xi \in \mathbb{R}^{+} \\ 1 & ; \textit{otherwise} \end{cases} \tag{11}$$

$$\Phi_0(t) = \begin{cases} -\log\left(1-\left(1-e^{-\delta t^{\xi}}\right)^{\zeta}\right) & ; t \in \mathbb{R}^{+}, \delta, \zeta, \xi \in \mathbb{R}^{+} \\ 0 & ; \textit{otherwise} \end{cases} \tag{12}$$

3.2 Exponential power distribution

Another baseline distribution we considered is exponential power distribution. A continuous random variable T is said to follow exponential power distribution if survival, hazard, and cumulative hazard function is,

$$S(t) = \begin{cases} e^{(1-e^{\delta t^{\zeta}})} & ; t \in \mathbb{R}^{+}, \delta, \zeta \in \mathbb{R}^{+} \\ 1 & ; \textit{otherwise} \end{cases} \tag{13}$$

$$\phi_0(t) = \begin{cases} \zeta \delta t^{\zeta-1} e^{\delta t^{\zeta}} & ; t \in \mathbb{R}^{+}, \delta, \zeta \in \mathbb{R}^{+} \\ 0 & ; \textit{otherwise} \end{cases} \tag{14}$$

$$\Phi_0(t) = \begin{cases} e^{\delta t^{\zeta}} - 1 & ; t \in \mathbb{R}^{+}, \delta, \zeta \in \mathbb{R}^{+} \\ 0 & ; \textit{otherwise} \end{cases} \tag{15}$$

Kolmogorov–Smirnov (K–S) statistic for goodness of fit shows that both baseline distributions are fitting well to kidney infection data set(see section 7, Figure 1-4).

4. Proposed Model

Due to group variation or frailty and individual variation described by the hazard function, a shared frailty model can be considered as a mixture model in survival analysis. After substituting cumulative hazard function for generalized Weibull and exponential power baseline distributions in equations (8) and (9), we get the following four survival functions.

$$S(t_{1j}, t_{2j}) = \frac{1}{(1+\theta)} \left[\frac{\theta^{\mu+1}}{\left[\theta + \left\{ \log\left(1 - \left(1 - e^{\delta_1 t_{1j}^{\xi_1}}\right)^{\zeta_1}\right) + \log\left(1 - \left(1 - e^{\delta_2 t_{2j}^{\xi_2}}\right)^{\zeta_2}\right)\right\}\rho\right]^{\mu}} + \frac{\theta^{\theta(1+\theta-\mu)}}{\left[\theta + \left\{ \log\left(1 - \left(1 - e^{\delta_1 t_{1j}^{\xi_1}}\right)^{\zeta_1}\right) + \log\left(1 - \left(1 - e^{\delta_2 t_{2j}^{\xi_2}}\right)^{\zeta_2}\right)\right\}\rho\right]^{\theta(1+\theta-\mu)}} \right] \tag{16}$$

$$S(t_{1j}, t_{2j}) = e^{\rho_j \left(\log\left(1 - \left(1 - e^{\delta_1 t_{1j}^{\xi_1}}\right)^{\zeta_1}\right) + \log\left(1 - \left(1 - e^{\delta_2 t_{2j}^{\xi_2}}\right)^{\zeta_2}\right)\right)} \tag{17}$$

$$S(t_{1j}, t_{2j}) = \frac{1}{(1+\theta)} \left[\frac{\theta^{\mu+1}}{\left(\theta + \rho\left\{e^{\delta_1 t_{1j}^{\zeta_1}} + e^{\delta_2 t_{2j}^{\zeta_2}} - 2\right\}\right)^{\mu}} + \frac{\theta^{\theta(1+\theta-\mu)}}{\left(\theta + \rho\left\{e^{\delta_1 t_{1j}^{\zeta_1}} + e^{\delta_2 t_{2j}^{\zeta_2}} - 2\right\}\right)^{\theta(1+\theta-\mu)}} \right] \tag{18}$$

$$S(t_{1j}, t_{2j}) = e^{-\rho\left\{e^{\delta_1 t_{1j}^{\zeta_1}} + e^{\delta_2 t_{2j}^{\zeta_2}} - 2\right\}} \tag{19}$$

Here, equations (16), (17) can be called as Model-I, Model-III respectively that have been established for generalized Weibull baseline distribution with and without frailty and equations (18), (19) can be called as Model-II and Model-IV respectively that have been established for exponential power baseline distribution with and without frailty.

5. Likelihood Design and Bayesian Paradigm

For the study, n individuals have been considered. Observed failure times have been indicated by (t_{1j}, t_{2j}). We are using the random censoring scheme. Censoring time, supposed to be indicated by c_{1j} and c_{2j} for j^{th} individual $(j = 1, 2, 3, ..., n)$. Independence between censoring schemes and lifetimes of individuals has been presumed. Likelihood function can be described for bivariate lifetime random variable of the j^{th} individual as,

$$L_j(t_{1j}, t_{2j}) = \begin{cases} f_1(t_{1j}, t_{2j}), & ; t_{1j} < c_{1j}, t_{2j} < c_{2j}, \\ f_2(t_{1j}, c_{2j}), & ; t_{1j} < c_{1j}, t_{2j} > c_{2j}, \\ f_3(c_{1j}, t_{2j}), & ; t_{1j} > c_{1j}, t_{2j} < c_{2j}, \\ f_4(c_{1j}, c_{2j}), & ; t_{1j} > c_{1j}, t_{2j} > c_{2j}. \end{cases}$$

and likelihood function will be,

$$L(\underline{\Theta}, \underline{\beta}, \theta, \mu) = \prod_{j=1}^{n_1} f_1(t_{1j}, t_{2j}) \prod_{j=1}^{n_2} f_2(t_{1j}, c_{2j}) \prod_{j=1}^{n_3} f_3(c_{1j}, t_{2j}) \prod_{j=1}^{n_4} f_4(c_{1j}, c_{2j}) \tag{20}$$

where,$\underline{\Theta}$, $\underline{\beta}$, θ and μ are vector of baseline parameters and the vector of regression coefficients and frailty parameters respectively. Likelihood function for without frailty model is,

$$L(\underline{\Theta}, \underline{\beta}) = \prod_{j=1}^{n_1} f_1(t_{1j}, t_{2j}) \prod_{j=1}^{n_2} f_2(t_{1j}, c_{2j}) \prod_{j=1}^{n_3} f_3(c_{1j}, t_{2j}) \prod_{j=1}^{n_4} f_4(c_{1j}, c_{2j}) \tag{21}$$

let n_1, n_2, n_3 and n_4 be the number of pairs for which first and second failure times (t_{1j}, t_{2j}) lie in the ranges $t_{1j} < c_{1j}, t_{2j} < c_{2j}$; $t_{1j} < c_{1j}, t_{2j} > c_{2j}$; $t_{1j} > c_{1j}, t_{2j} < c_{2j}$ and $t_{1j} > c_{1j}, t_{2j} > c_{2j}$ respectively and let

$$\begin{aligned} f_1(t_{1j}, t_{2j}) &= \frac{\partial^2 S(t_{1j}, t_{2j})}{\partial t_{1j} \partial t_{2j}} \\ f_2(t_{1j}, c_{2j}) &= -\frac{\partial S(t_{1j}, c_{2j})}{\partial t_{1j}} \\ f_3(c_{1j}, t_{2j}) &= -\frac{\partial S(c_{1j}, t_{2j})}{\partial t_{2j}} \\ f_4(c_{1j}, c_{2j}) &= S(c_{1j}, c_{2j}). \end{aligned} \tag{22}$$

substituting cumulative hazard rates $\Phi_{01}(t_{1j})$ and $\Phi_{02}(t_{2j})$ and survival function $S(t_{1j}, t_{2j})$ in equation (22) for Model-I and Model-II and by differentiating we get the likelihood function. The maximum likelihood method has a crucial importance in computing efficient estimators. Inappropriately, due to a convergence problem, maximum likelihood failed to estimate the parameters, because of Model-I has thirteen-dimensional, Model-II, Model-III have eleven-dimensional and Model-IV has nine-dimensional optimization problem. The Bayesian scenario has been discussed by several researchers for estimating parameters of the frailty models. For gamma and log-normal frailty models, the Bayesian paradigm has been contemplated by Santos and Achcar (2010). Weibull and piecewise exponential model have been discussed by Ibrahim *et al.* (2001) with gamma frailty. The joint posterior density function of parameters for given failure times is obtained as,

$$\pi(\Theta, \theta, \mu, \underline{\beta_0}) \propto L(\Theta, \mu, \underline{\beta}_0) g_1(\zeta) g_2(\xi) g_3(\delta) g_4(\theta) g_5(\mu) \prod_{i=1}^{5} p_i(\beta_{0i\times 1})$$

where $g_i(.)$ indicates the prior density function with known hyperparameters of corresponding argument for baseline parameters and frailty variance; $p_i(.)$ is prior density function for regression coefficient β_{0i} and likelihood function is $L(.)$. An important assumption here is, all the parameters are independently distributed. In a similar way, joint posterior density function can be written for without frailty models. To estimate the parameters of the models, Metropolis-Hastings algorithms and Gibbs samplers have been used. Geweke test (see Geweke, 1992) and Gelman-Rubin (see Gelman and Rubin, 1992) statistics have been used to monitor the convergence of a Markov chain to a stationary distribution.

Due to the high-dimensions of conditional distributions, it is not unproblematic to integrate out. Thus, it has been considered that full conditional distributions can be obtained as they are proportional to the joint distribution of the parameter of the model. The conditional distribution for single parameter δ with frailty as,

$$\psi_1(\delta \mid \xi, \zeta, \theta, \mu, \underline{\beta_0}) \quad \propto \quad L(\delta, \xi, \zeta, \theta, \mu, \underline{\beta_0}) \cdot g_1(\delta) \tag{23}$$

and the conditional distribution for single parameter δ without frailty as,

$$\psi_1(\delta \mid \xi, \zeta, \underline{\beta_0}) \quad \propto \quad L(\delta, \xi, \zeta, \underline{\beta_0}) \cdot g_1(\delta)$$

similarly full conditional distributions can be obtained.

6. Simulation Study

A simulation study has been executed to appraise the Bayesian estimation paradigm for Model-I and Model-II. Single covariate K_1 has been considered as follows normal distribution. The frailty variable W is assumed to follow generalized Lindley distribution. Independence between lifetimes of individuals has been considered. Samples are generated using the subsequent mechanism,

1. From the binomial distribution with probability 0.6, 25 values for K_1 has been generated.

2. For known covariates, compute $\rho = e^{K_1\beta_1}$.

3. Lifetimes reckoned to follows generalized Weibull and exponential power baseline distributions for given frailty W_j. 25 values of lifetimes have been spawned after using ensuing manners.
Conditional survival function for lifetime t_j $(j = 1, 2, ..., n)$ for given frailty $W_j = w_j$ and covariate K_1 is,

$$S(t_j \mid w_j, K_1) \quad = \quad e^{-w_j H_0(t_j)\rho}$$

Equating $S(t_j \mid w_j, K_1)$ to random number, say $v_j (0 < v_j < 1)$ spawned from $U(0, 1)$ over $t_j > 0$ we get,
for Model-I,

$$t_j \quad = \quad \left(-\frac{1}{\delta}\log(1-(1-v_j^{\frac{1}{w_j\rho_j}})^{\frac{1}{\zeta}})\right)^{\frac{1}{\xi}}$$

for Model-II,

$$t_j \quad = \quad \left(\frac{1}{\delta}\log(1-\frac{1}{w_j\rho_j}\log(v_j))\right)^{\frac{1}{\zeta}}$$

4. Censoring time c_j has been spawned from $G(0.9, 0.01)$ for Model-I.

5. Observe the j^{th} survival time $t_j^* = min(t_j, c_j)$ and the censoring indicator δ_j for the j^{th} individual $(j = 1, 2, ..., 25)$ where,

$$\delta_j = \begin{cases} 1, & ; t_j < c_j \\ 0, & ; t_j > c_j \end{cases}$$

thus we have data consisting of 25 pairs of survival times t_j^* and the censoring indicator δ_j.

Concurrently, with different priors and starting points, two chains based on two priors (one is based on gamma prior and another is based on uniform prior) have been operated. Both chains recapitulated 100,000 times. Gelman-Rubin test (see Gelman and Rubin, 1992) values are very close to one. Due to small values of Geweke test statistic (see Geweke, 1992) and corresponding p-values, the chains reach stationary distribution for both prior sets. In view of, estimates of parameters were about the same, no impact of prior distributions has been founded on posterior summaries. Here, the analysis for one chain has been exhibited because both the chains have shown similar results. Tables 1 and 2 present the estimates and the credible intervals of the parameters for the Models I and II based on the simulation study. The Gelman-Rubin convergence statistic values are nearly equal to one and also the Geweke test values are quite small and the corresponding p-values are large enough to say that the chain attains stationary distribution.

7. Analysis of Kidney Infection Data

To elucidate the Bayesian estimation paradigm, kidney infection data of McGilchrist and Aisbett (1991) has been considered. This data consists of 38 patients, recurrence times (in days) of infection are given which can be outlined as these are recorded from the insertion of the catheter until it has to be removed due to infection. Data having five known covariates age, sex (Female=1, Male=0), and disease type Glomerulo Neptiritis (GN), Acute Nephritis (AN) and Polycystic Kidney Disease (PKD). Opine first and second time to infection is symbolized by T_{1j} and T_{2j}. Five covariates age, sex, GN, AN and PKD are symbolized by K_1, K_2, K_3, K_4 and K_5. To check goodness of fit of kidney data set, we consider Kolmogrove-Smirnov (K-S) test for two baseline distributions. Table 3 gives the p-values of goodness of fit test for Model I and Model II. Thus from p-values of K-S test we can say that there is no statistical evidence to reject the hypothesis that data are from the Model I and Model II in the marginal case and we assume that they also fit for bivariate case. Figures 1-4 show the parametric plot with semi-parametric plot for models (Model I and Model II) with frailty for T_1 and T_2 separately and both lines are close to each other.

For frailty parameters, gamma prior distribution with very small shape and scale parameters (say, 0.0001) has been used. Additionally, it can be considered, regression coefficients are normally distributed with mean zero and high variance (say 1000). A similar type of prior was used in Ibrahim et al (2001) and Santos and Achcar (2010). That's why for frailty parameters θ, μ and regression coefficients $\beta_{0i}, i = 1, ..., 5$, vague priors have been used. Because of no information about baseline parameter having, therefore, prior distribution corresponding to baseline parameters are also considered flat. We considered two different vague prior distributions for baseline parameters, one is gamma distribution with shape and scale hyperparameters ϵ_1, ϵ_2 respectively and another is uniform distribution with interval (ν_1, ν_2). All the hyperparameters are known. Under the Bayesian paradigm, for both models, two parallel chains have been run. Also, two sets of prior distributions have been used with different starting points using the Metropolis-Hastings algorithm and Gibbs sampler based on normal transition kernels. It can be said that estimates are independent of the different prior distributions because, for both sets of priors, estimates of parameters are approximately similar. We got almost similar convergence rate of Gibbs sampler for both sets of priors. Here, the analysis for one chain has been exhibited because both the chains have shown similar results.

Markov chain has seemed to reach the stationary state because of the zigzag pattern of the trace plots for all the parameters that gesture parameters move and mix more freely (See Figure 5). Coupling from the past plot has been applied to fix up the burn-in period (See Figure 6). A sequence of draws may have serial correlation after the burn-in period. Randomness may not be shown in successive draws. But almost independence can be seen in values at the extensive split. After using the values from the single run of the Markov chain, a vague sample can be obtained from the posterior distribution. Because of the burn-in period, it has been founded at extensive spaced time points. Autocorrelation function (ACF) plots can be utilized to examine the appropriate blend of our chains (See Figure 7). ACF plot for each parameter is converging to the posterior mean of the parameter, thus, represents a good mixing of the chain. Thus, our diagnostic plots suggest that the MCMC chains are mixing very well. After a certain lag, the serial correlation of the parameters turns out to almost negligible for all the parameters. Observations are shown independently after thinning the serial correlation function plot (See Figure 8). For visual approximate estimates as

confirmative measures such as posterior density plots also drawn for Model-I. It has been observed in some of posterior densities of the parameters depict multi-modal shapes which are quite possible in frailty models. The Gelman-Rubin convergence statistic values are closely equal to one. The Geweke test statistic values are somewhat small, and the corresponding p-values are large enough to say that the chains reach stationary distribution. Tables 4-7 give the values of posterior mean and the standard error with 95% credible intervals, the Gelman-Rubin statistics values and the Geweke test with p-values for Model I, II, III and IV. Table 8 present the values of AIC, BIC and DIC values for both models. Values of AIC, BIC, and DIC, given in Table 8, have been used to the comparison of all models. Model-I holds the lowest possible values of AIC, BIC, and DIC. For all models, regression coefficients contained different values. For Model-I and Model-II, the credible interval of $\beta_{02}, \beta_{03}, \beta_{04}, \beta_{05}$ are not contained zero. It indicates that covariates sex, diseases GN, AN and PKD have a significant effects on all four models. It is being indicated that sex (β_2), disease PKD (β_5) are significant factors for kidney infection, having negative effects for all the four models. Negative value of β_2 indicates that the female patients have a slightly lower risk for infection. Negative value of β_5 indicates that the patients with the disease PKD has a slightly lower risk for infection.

8. Conclusions

Generalized Lindley frailty model under generalized Weibull and exponential power baseline distributions have been proposed. To fit the proposed models M-H algorithm and Gibbs samplers have been applied. Analysis has been done in R statistical software with self-written programs. The value of both frailty parameters for Model-I ($\theta = 3.08680, \mu = 2.89438$) and Model-II ($\theta = 2.89271, \mu = 2.49934$) are very high (See Tables 4 and 5) and corresponding variances are 1.38334 and 1.42248 by using equation (2.2). This exhibits that there is a strong indication of heterogeneity among the patient in the population for the data set. To take the decision about all models, different tools have been utilized. With the lowest value of AIC, BIC and DIC, given by Table 8, it can be said that Model-I and Model-II are better than Model-III and Model-IV for analyzing kidney infection data. The generalized Lindley frailty with generalized Weibull baseline (Model-I) is the best among all four models. For kidney infection data, sex, diseases AN, GN, and PKD have been found statistically significant factors for both with frailty and without frailty models (See Tables 4-7). Our proposed frailty model (Model-I) has been founded better in compare to Hanagal and Pandey's (2015a) frailty model with baseline generalized Weibull distribution. In a similar way, with a minimum value of AIC, our proposed frailty model (Model-II) has been founded better in compare of Hanagal and Dabade's (2015) frailty model.

Acknowledgments

The authors are thankful to the reviewer for the valuabe suggestions and comments.

References

Block, H. W., Savits, T. H. and Singh, H. (1998). On the reversed hazard rate function. *Probability in the Engineering and Informational Sciences*, **12**, 69- 90.

Clayton, D. G. (1978). A model for association in bivariate life tables and its applications to epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141-151.

Cox, D. R. (1972). Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society*, **B34**, 187-220.

Duffy, D. L., Martin, N. G. and Mathews, J. D. (1990). Appendectomy in Australian twins. *Australian Journal of Human Genetics*, **47(3)**, 590-92.

Flinn, C. J. and Heckman, J. J. (1982). New methods for analyzing individual event histories. *Sociological Methodology*, **13**, 99-140.

Gelman, A. and Rubin, D. B. (1992). A single series from the Gibbs sampler provides a false sense of security. *In Bayesian Statistics* **4** (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith). Oxford: Oxford University Press. pp. 625-632.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *In Bayesian Statistics* **4** (eds. J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith), Oxford: Oxford University Press, pp. 169-193.

Hanagal, D. D. (2008). Frailty regression models in mixture distributions. *Journal of Statistical Planning and Inference*, **138(8)**, 2462-68.

Hanagal, D. D. (2011). *Modeling Survival Data Using Frailty Models*. Chapman & Hall/CRC. New York.

Hanagal, D. D. (2017). Frailty Models in Public Health. *Handbook of Statistics*, **37(B)**, 209-247. Elsevier Publishers; Amsterdam.

Hanagal, D. D. (2019). *Modeling Survival Data Using Frailty Models*. 2nd Edition. Springer; Singapore.

Hanagal, D. D., and Bhambure, S. M. (2014). Analysis of kidney infection data using shared positive stable frailty models. *Advances in Reliability*, **1**, 21-39.

Hanagal, D. D. and Bhambure, S. M. (2015). Comparison of shared gamma frailty models using Bayesian approach. *Model Assisted Statistics & Applications*, **10**, 25-41.

Hanagal, D. D. and Bhambure, S. M. (2016). Modeling bivariate survival data using shared inverse Gaussian frailty model. *Communications in Statistics, Theory & Methods*, **45(17)**, 4969-4987.

Hanagal, D. D., and Dabade, A. D. (2013). Modeling of inverse Gaussian frailty model for bivariate survival data. *Communications in Statistics-Theory and Methods*, **42(20)**, 3744-3769.

Hanagal, D. D., and Dabade, A. D. (2015). Comparison of shared frailty models for kidney infection data under exponential power baseline distribution. *Communications in Statistics-Theory and Methods*, **44(23)**, 5091-5108.

Hanagal, D. D., and Kamble, A. T. (2015). Bayesian estimation in shared compound poisson frailty models. *Journal of Reliability and Statistical Studies*, **8(1)**, 159-180.

Hanagal, D. D., and Pandey, A. (2014a). Inverse Gaussian shared frailty for modeling kidney infection data. *Advances in Reliability*, **1**, 1-14.

Hanagal, D. D. and Pandey, A. (2014b). Gamma shared frailty model based on reversed hazard rate for bivariate survival data. *Statistics & Probability Letters*, **88**, 190-196.

Hanagal, D. D., and Pandey, A. (2015a). Gamma frailty models for bivariate survival data. *Journal of Statistical Computation and Simulation*, **85(15)**, 3172-3189.

Hanagal, D. D., and Pandey, A. (2015b). Inverse Gaussian shared frailty models with generalized exponential and generalized inverted exponential as baseline distributions. *Journal of Data Science*, **13(2)**, 569-602.

Hanagal, D. D., and Pandey, A. (2016a). Gamma shared frailty model based on reversed hazard rate. *Communications in Statistics-Theory and Methods*, **45(7)**, 2071-2088.

Hanagal, D. D. and Pandey, A. (2016b). Inverse Gaussian shared frailty models based on reversed hazard rate. *Model Assisted Statistics and Applications*, **11**, 137-151.

Hanagal, D. D., and Pandey, A. (2017a). Shared inverse Gaussian frailty models based on additive hazards. *Communications in Statistics-Theory and Methods*, **46(22)**, 11143-11162.

Hanagal, D. D., and Pandey, A. (2017b). Shared frailty models based on reversed hazard rate for modified inverse Weibull distribution as baseline distribution. *Communications in Statistics-Theory and Methods*, **46(1)**, 234-246.

Hanagal, D. D. and Sharma, R. (2013). Modeling heterogeneity for bivariate survival data by shared gamma frailty regression model. *Model Assisted Statistics and Applications*, **8**, 85-102.

Hanagal, D. D. and Sharma, R. (2015a). Bayesian inference in Marshall-Olkin bivariate exponential shared gamma frailty regression model under random censoring. *Communications in Statistics, Theory and Methods*, **44(1)**, 24-47.

Hanagal, D. D. and Sharma, R. (2015b). Comparison of frailty models for acute leukaemia data under Gompertz baseline distribution. *Communications in Statistics, Theory & Methods*, **44(7)**, 1338-1350.

Hanagal, D. D. and Sharma, R. (2015c). Analysis of bivariate survival data using shared inverse Gaussian frailty model. *Communications in Statistics, Theory & Methods*, **44(7)**, 1351-1380.

Hougaard, P. (1985). Discussion of the paper by D.G. Clayton and J. Cuzick. *Journal of the Royal Statistical Society*, A, **148**, 113-14.

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

Hougaard, P. (1991). Modeling heterogeneity in survival data. *Journal of Applied Probability*, **28**, 695-701.

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

Ibrahim, J. G., Ming-Hui C. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer Verlag, New York.

Lindley, D. V. (1958). Fiducial distributions and Bayes's theorem. *Journal of the Royal Statistical Society*, B, **20**, 102-107.

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

Oakes, D. (1989). Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association*, **84(406)**, 487-493.

Santos, C. A. and Achcar, J. A.(2010). A Bayesian analysis for multivariate survival data in the presence of covariates. *Journal of Statistical Theory and Applications*, **9**, 233-253.

Shaked, M. and Shantikumar, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, New York.

Vaupel, J. W., Manton, K. G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**, 439-454.

Zakerzadeh, H. and Dolati, A. (2009). Generalized Lindley distribution. *Journal of Mathematical Extension*, **3(2)**, 1-17.

Appendix
Summary of Tables and Figures

Table 1: Posterior Summary of Generalized Lindley Frailty with Baseline Generalized Weibull (Simulation Study: Model I)

Parameter	Estimate	S.E.	L.C.L.	U.C.L.	Geweke test	p-value	Gelman-Rubin test
$\zeta_1(3.1)$	3.08997	0.17672	2.68349	3.44157	0.00101	0.50040	0.99999
$\delta_1(0.15)$	0.15417	0.01551	0.12216	0.17852	-0.01145	0.49543	1.00378
$\xi_1(0.85)$	0.85822	0.03033	0.79395	0.91321	0.01542	0.50615	1.00239
$\zeta_2(5.0)$	4.99795	0.55329	4.07576	5.95502	-0.00825	0.49671	0.99998
$\delta_2(0.26)$	0.27186	0.02817	0.20969	0.31781	0.00043	0.50017	1.00494
$\xi_2(0.74)$	0.74822	0.02897	0.69046	0.81146	-0.00174	0.49930	1.00096
$\theta(3.0)$	2.99983	0.15562	2.71870	3.28869	0.01826	0.50729	1.00002
$\mu(2.5)$	2.49520	0.08979	2.33103	2.65688	0.00415	0.50166	1.00201
$\beta_1(0.005)$	0.00458	0.00349	-0.00237	0.01135	-0.00530	0.49788	1.00021

Table 2: Posterior Summary of Generalized Lindley Frailty with Baseline Exponential Power (Simulation Study: Model II)

Parameter	Estimate	S.E.	L.C.L.	U.C.L.	Geweke test	p-value	Gelman-Rubin test
$\zeta_1(0.75)$	0.71106	0.03077	0.65061	0.77292	0.00752	0.50300	1.00123
$\delta_1(0.09)$	0.09774	0.00960	0.07452	0.10972	-0.00332	0.49868	1.01043
$\zeta_2(0.7)$	0.75231	0.05995	0.62533	0.86400	0.00035	0.50014	1.00299
$\delta_2(0.06)$	0.06937	0.00764	0.05270	0.07915	0.00347	0.50138	1.00051
$\theta(1.2)$	1.19076	0.09615	1.01631	1.37833	-0.00444	0.49823	1.00352
$\mu(0.7)$	0.70361	0.03565	0.63321	0.76945	-0.00327	0.49870	1.00050
$\beta_1(0.003)$	0.00303	0.00173	-0.00049	0.00664	-0.00322	0.49871	0.99997

Table 3: p-values of K-S Statistics for goodness of fit test for Kidney Infection data set

	Recurrence Time	
Distribution	**First**	**Second**
Model I	0.5174	0.6060
Model II	0.1184	0.4185

Table 4: Posterior Summary of Generalized Lindley Frailty with Baseline Generalized Weibull for Kidney Infection Data (Model I)

Parameter	Estimate	S.E.	L.C.L.	U.C.L.	Geweke test	p-value	Gelman-Rubin test
ζ_1	2.99002	0.15637	2.70025	3.29677	-0.00910	0.49637	1.00016
δ_1	0.18320	0.01455	0.15510	0.21414	-0.00871	0.49653	1.00094
ξ_1	0.78464	0.02354	0.73846	0.83274	0.00234	0.50094	1.00005
ζ_2	8.99452	0.97203	7.14211	10.92913	-0.00559	0.49777	0.99998
δ_2	0.30022	0.01887	0.26317	0.34148	-0.00381	0.49848	0.99996
ξ_2	0.67188	0.02468	0.62380	0.71751	-0.00787	0.49686	1.00005
θ	3.08680	0.12306	2.84842	3.35547	0.00019	0.50008	1.00048
μ	2.89438	0.12997	2.63819	3.16898	0.00218	0.50087	1.00010
β_1	0.00091	0.00054	-0.00004	0.00185	0.00781	0.50312	1.00117
β_2	-2.02839	0.22274	-2.46437	-1.58541	0.00661	0.50312	1.00306
β_3	-0.00446	0.00259	-0.00928	-0.00005	-0.00167	0.49933	1.00001
β_4	0.44121	0.20265	0.07885	0.80228	-0.00178	0.49929	1.00117
β_5	-1.06697	0.25150	-1.49606	-0.59086	0.00912	0.50364	1.00081

Table 5: Posterior Summary of Generalized Lindley Frailty with Baseline Exponential Power for Kidney Infection Data (Model II)

Parameter	Estimate	S.E.	L.C.L.	U.C.L.	Geweke test	p-value	Gelman-Rubin test
ζ_1	0.59126	0.01185	0.56663	0.61349	-0.00265	0.49894	1.00119
δ_1	0.08507	0.00315	0.07859	0.09102	0.00065	0.50026	1.00011
ζ_2	0.67026	0.01459	0.64434	0.69888	-0.00376	0.49850	0.99998
δ_2	0.05114	0.00301	0.04569	0.05758	-0.00204	0.49919	1.00012
θ	2.89271	0.09667	2.70694	3.09136	0.00038	0.49965	1.00047
μ	2.49934	0.03404	2.43316	2.56418	-0.00088	0.49831	1.00025
β_1	-0.00044	0.00205	-0.00431	0.00306	0.00464	0.50185	1.00006
β_2	-1.67836	0.10168	-1.86985	-1.49478	-0.00517	0.50185	1.00406
β_3	0.21870	0.02916	0.17235	0.26824	0.00703	0.50280	1.00004
β_4	0.75038	0.11075	0.55099	0.93289	-0.01367	0.49455	1.00029
β_5	-0.67039	0.05381	-0.76419	-0.57894	-0.00423	0.49831	0.99998

Table 6: Posterior Summary of Generalized Weibull Distribution for Kidney infection Data (Model III)

Parameter	Estimate	S.E.	L.C.L.	U.C.L.	Geweke test	p-value	Gelman-Rubin test
ζ_1	2.48490	0.30943	1.89920	3.08310	-0.00829	0.49670	1.00010
δ_1	0.20310	0.06863	0.09050	0.35890	-0.00614	0.49750	0.99990
ξ_1	0.60490	0.07812	0.46010	0.76200	0.01195	0.50470	0.99990
ζ_2	5.04010	0.50556	4.09990	5.94920	0.00073	0.50030	1.00030
δ_2	0.32220	0.08144	0.17580	0.49420	-0.00887	0.49650	1.00000
ξ_2	0.51290	0.06162	0.38820	0.63330	0.01124	0.50450	1.00000
β_1	0.00070	0.00279	-0.00440	0.00630	-0.00968	0.49610	1.00040
β_2	-1.07160	0.31695	-1.67560	-0.46080	-0.01568	0.49370	0.99990
β_3	-0.01590	0.02781	-0.06770	0.03750	0.00845	0.50340	1.00040
β_4	-0.00410	0.00660	-0.01670	0.00780	-0.00533	0.49780	0.99990
β_5	0.00120	0.00185	-0.00210	0.00460	0.00589	0.50240	1.00000

Table 7: Posterior Summary of Exponential Power Distribution for Kidney infection Data (Model IV)

Parameter	Estimate	S.E.	L.C.L.	U.C.L.	Geweke test	p-value	Gelman-Rubin test
ζ_1	0.61387	0.01685	0.57520	0.64394	0.00142	0.50057	1.00138
δ_1	0.06108	0.00327	0.05428	0.06775	0.00493	0.50197	1.00050
ζ_2	0.63406	0.01645	0.60266	0.66541	0.00632	0.50252	1.00090
δ_2	0.05032	0.00305	0.04411	0.05608	-0.00380	0.49849	1.00051
β_1	-0.00193	0.00178	-0.00579	0.00084	-0.00567	0.49774	1.00285
β_2	-1.61959	0.09413	-1.81647	-1.48638	0.00300	0.49774	0.99998
β_3	0.21887	0.02533	0.17269	0.26366	-0.01487	0.49407	0.99999
β_4	0.76834	0.09830	0.57241	0.93181	0.00916	0.50365	1.00803
β_5	-0.67269	0.03430	-0.74420	-0.60866	-0.00258	0.49897	1.00013

Table 8: AIC, BIC and DIC Comparison

Model	AIC	BIC	DIC
Model-I	682.2537	703.5423	661.4339
Model-II	689.5993	707.6128	670.3702
Model-III	690.2814	708.2949	678.103
Model-IV	690.3103	705.0486	676.7688

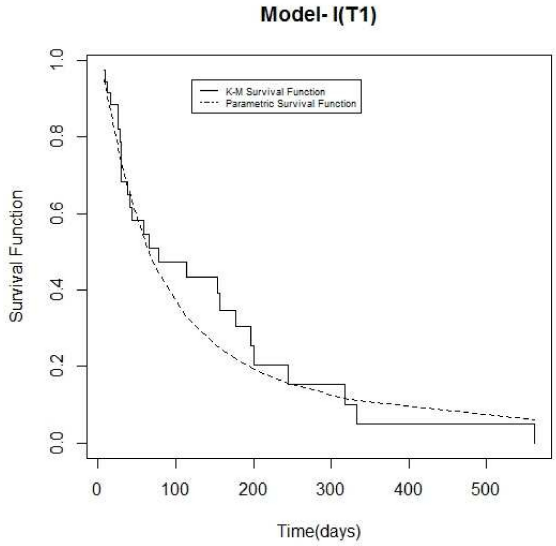

Figure 1

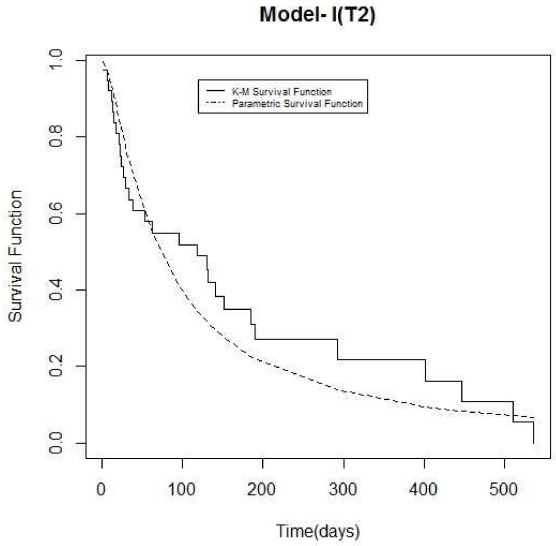

Figure 2

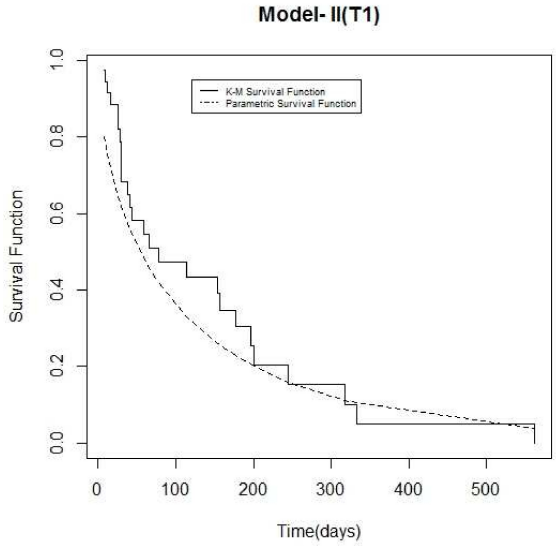

Figure 3

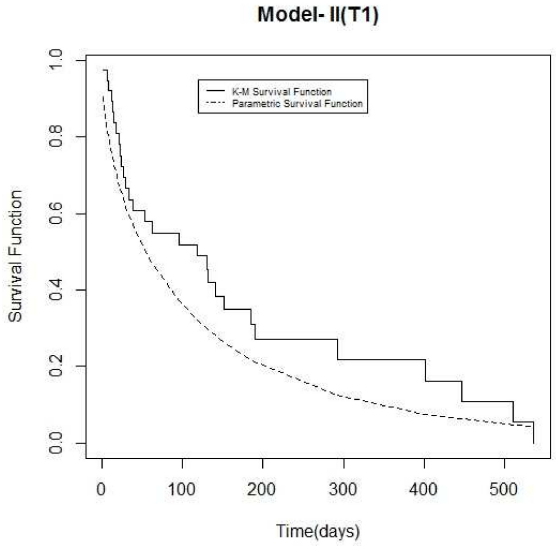

Figure 4

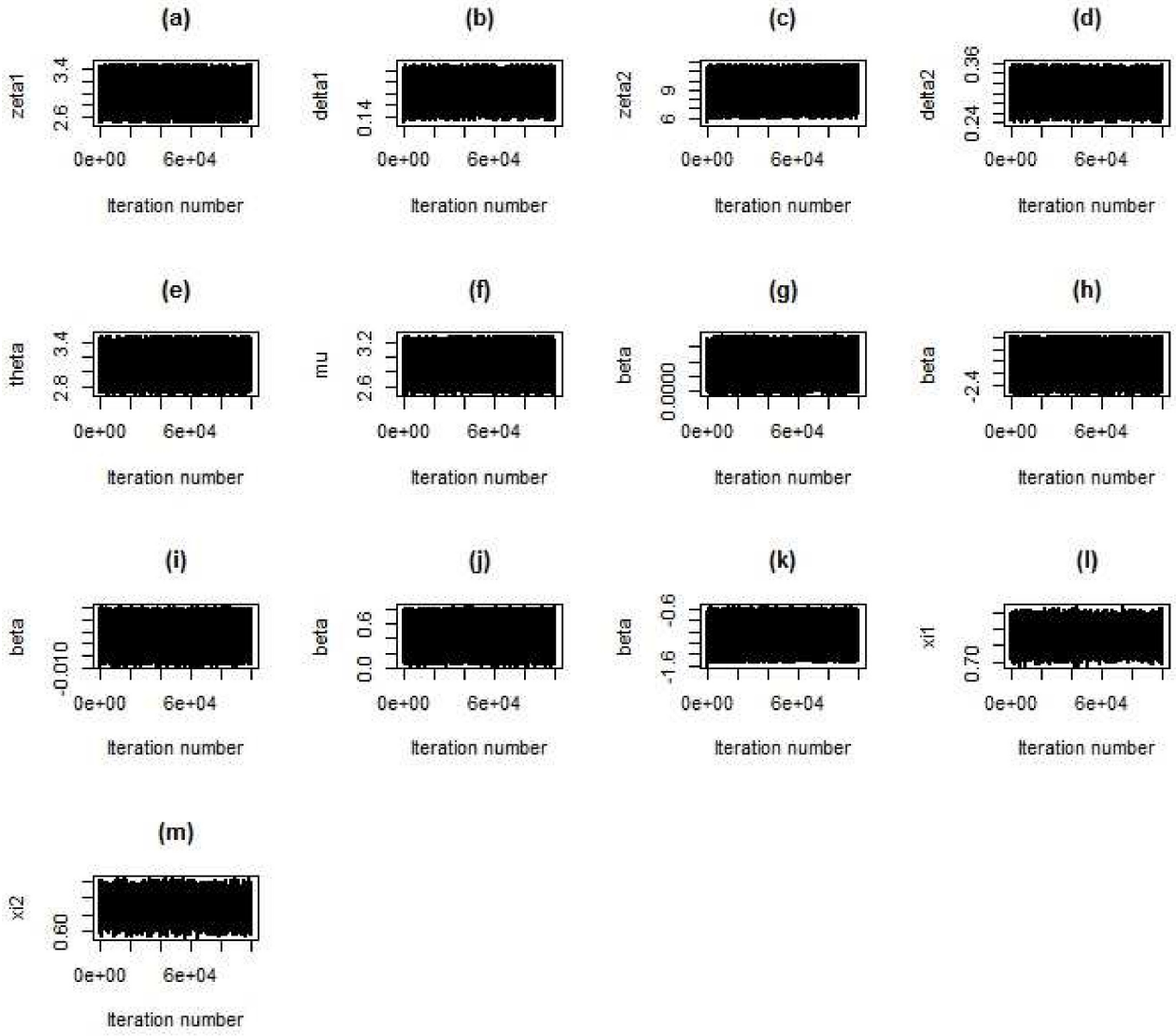

Figure 5: Trace plots for Model-I

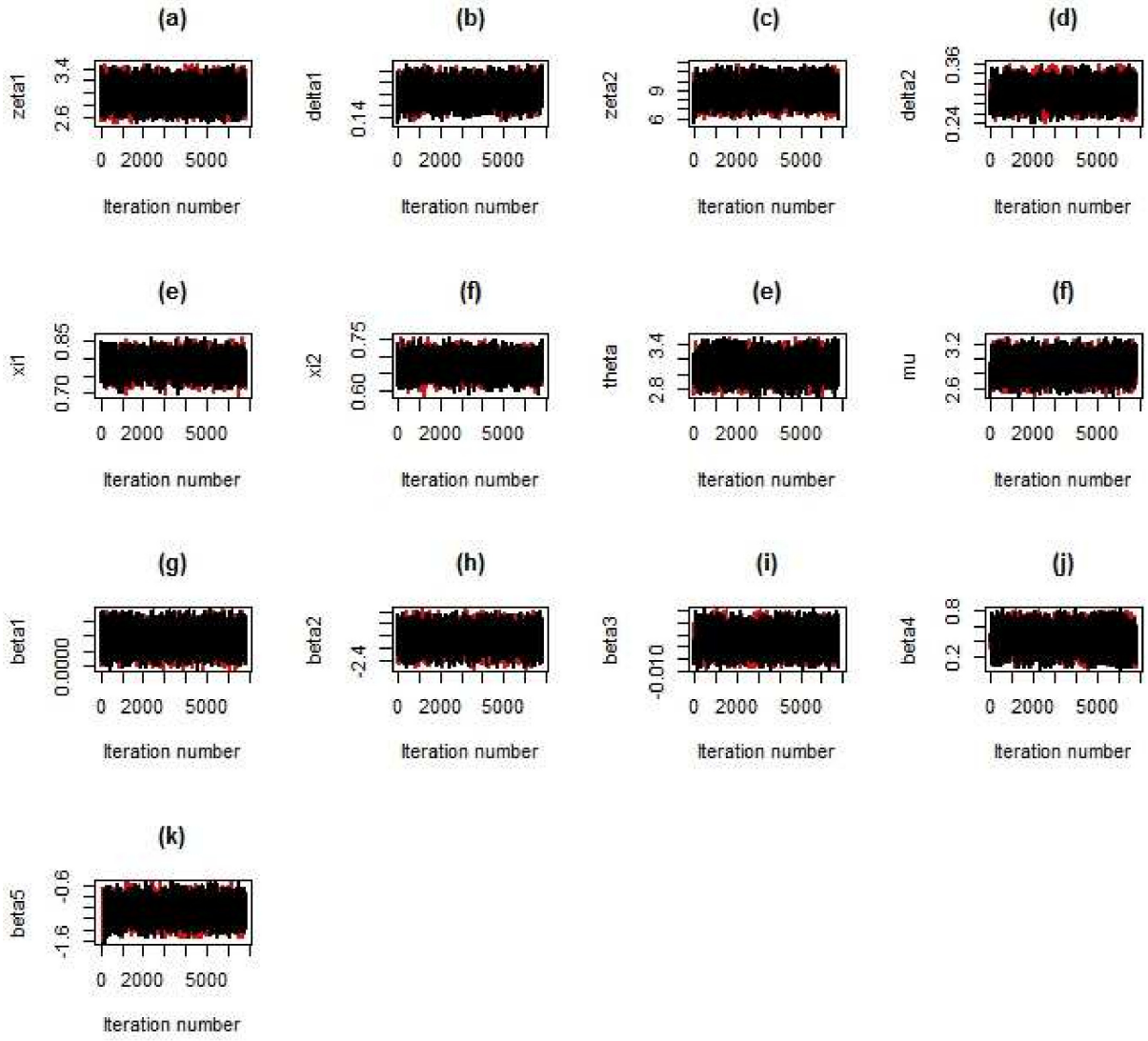

Figure 6: Coupling from the past plots for Model-I

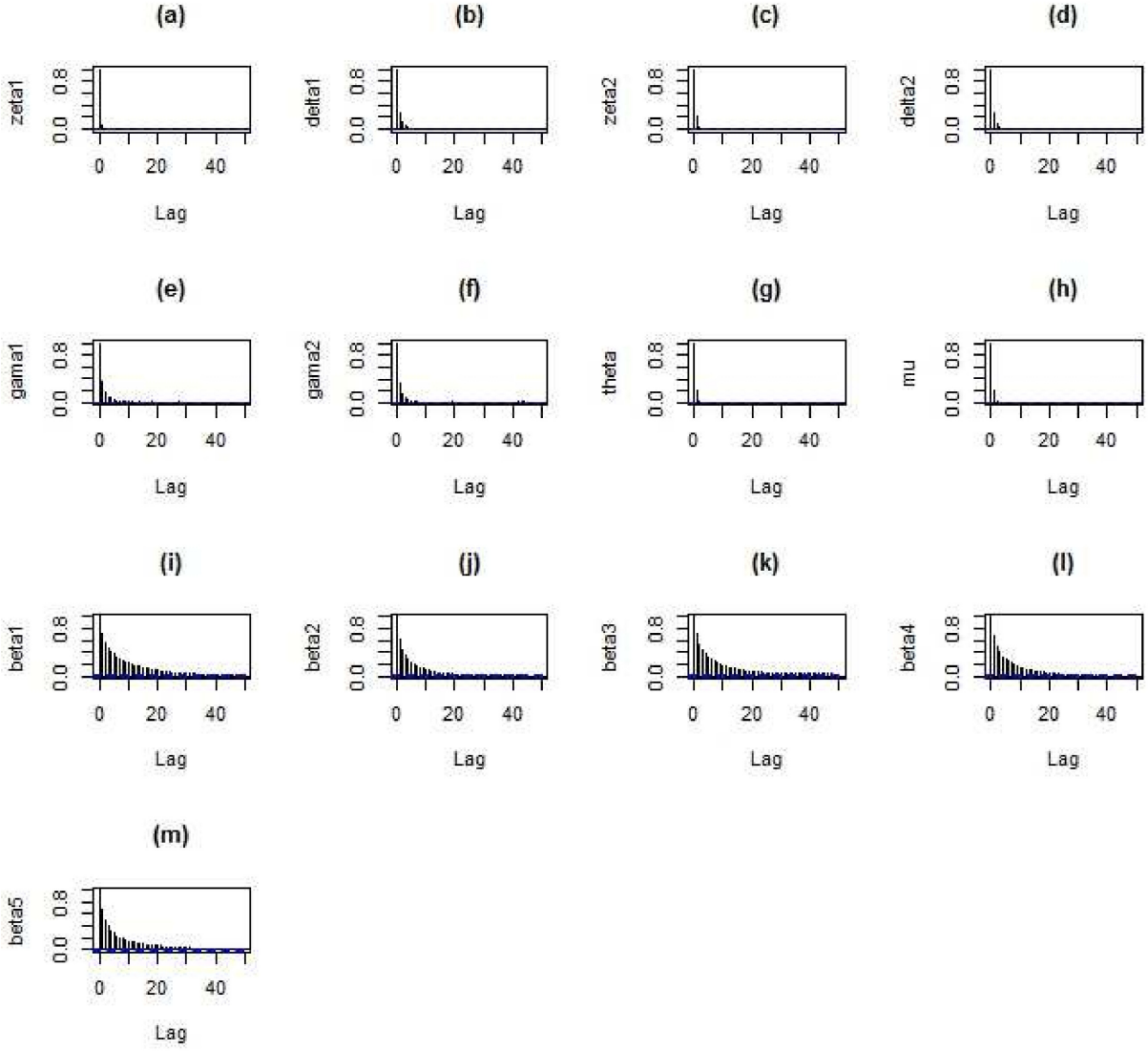

Figure 7: ACF plots for Model-I

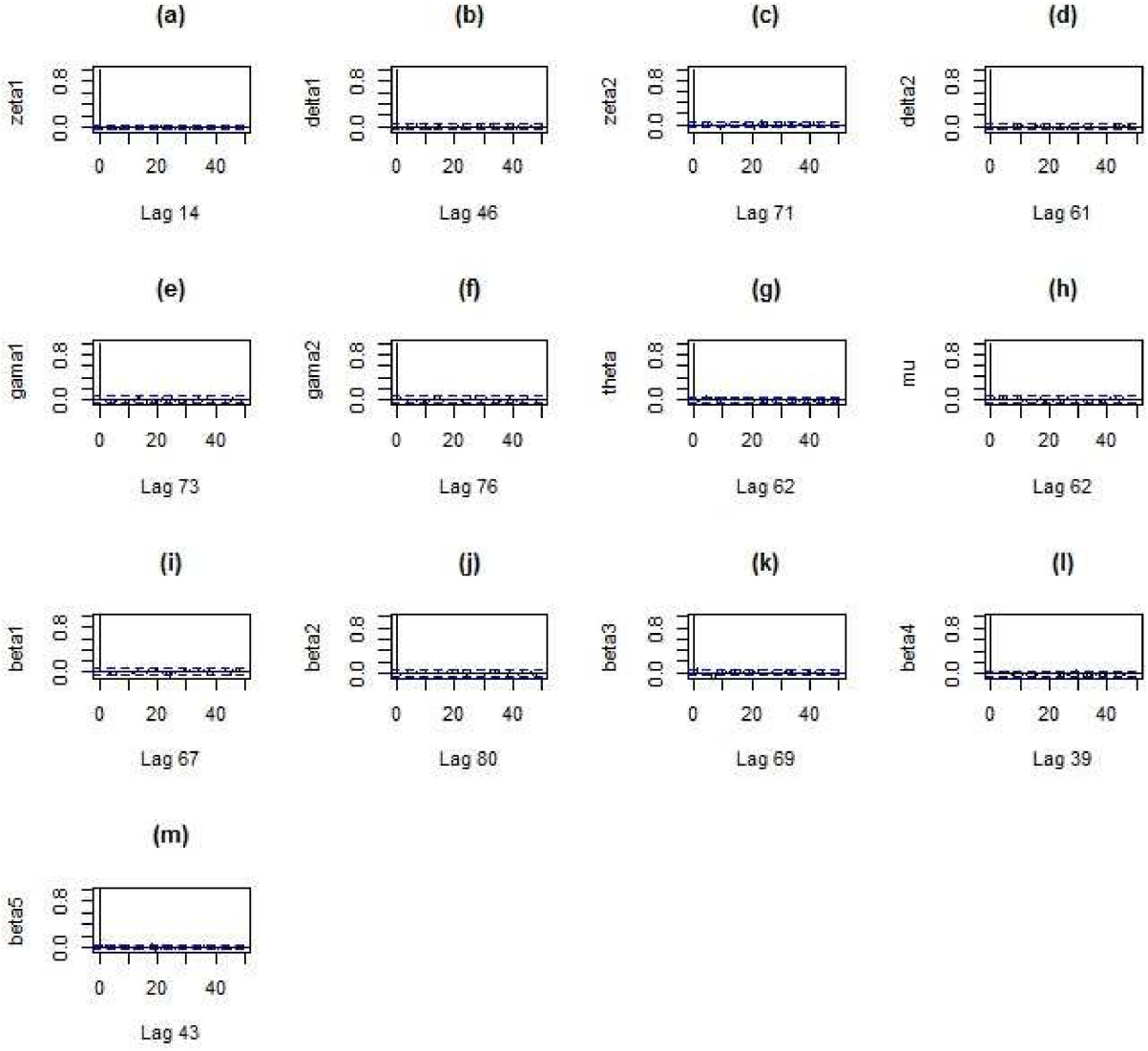

Figure 8: ACF plots After thinning for Model-I

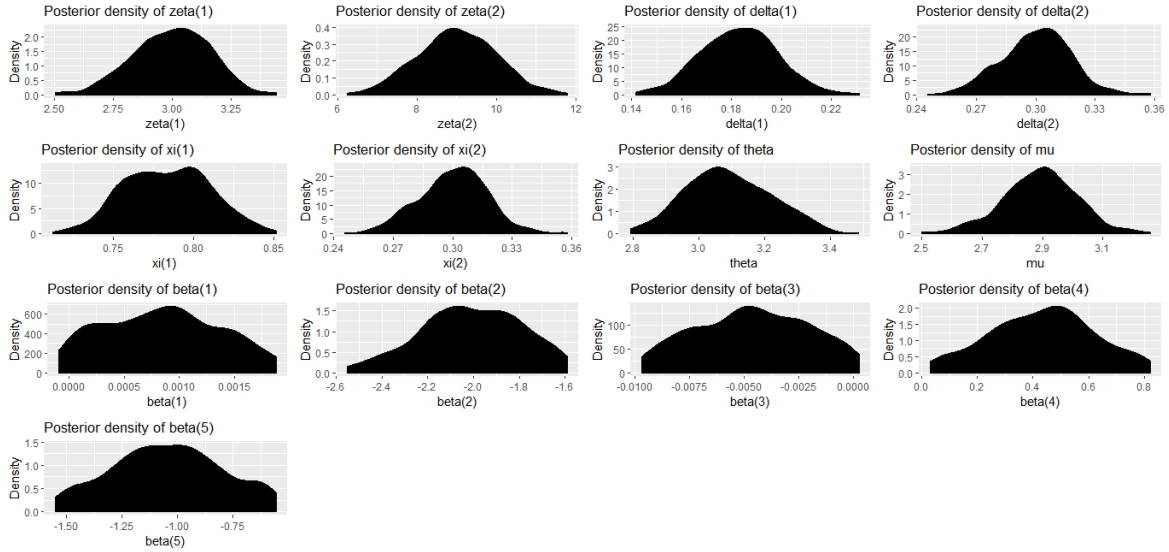

Figure 9: Posterior density plots for Model-I