International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1088-1097
© Research India Publications. https://dx.doi.org/10.37622/IJAER/15.11.2020.1088-1097

A Mixture Shared Frailty Model based on Log-logistic Baseline Distribution

Arvind Pandey[1], Praveen Kumar Misra[2] and Lalpawimawha[2,3]

[1]*Department of Statistics, Central University of Rajasthan -305817, India.*

[2]*Department of Mathematics and Statistics, Dr Shakuntala Misra National Rehabilitation University, Lucknow, Uttar Pradesh-226017, India.*

[3]*Department of Statistics, Pachhunga University College, Aizawl, Mizoram -796001, India.*

Abstract

Frailty model is a survival model designed to account for unobserved heterogeneity in the population. In this paper, we propose a new shared frailty model based on log-logistic as baseline distribution. The Bayesian approach of Markov Chain Monte Carlo (MCMC) technique was employed to estimate the parameters involved in the models. A simulation study was performed to compare the true values and the estimated values of the parameters. Comparison of different proposed models was done by using Bayesian comparison techniques. We apply to real life data set related to kidney infection due to insertion of catheter and the better model is suggested.

Keywords: Bayesian comparison, gamma frailty, inverse Gaussian frailty, log-logistic distribution, mixture frailty model, MCMC.

1. INTRODUCTION

In shared frailty model, the individuals in the same group suppose to have common frailty, which is share by an individuals or pair of organs. Frailty is an unobserved covariate, which is random and is responsible for the dependence between the failure times of the study subjects and heterogeneity in the population. But, most researchers used to neglect such type of unobserved covariates and does not know the important role plays in the interpretation of the covariates effects. That is why, the absence of frailty term may lead to the misinterpretation of the outcomes or results. Clayton(1978) proposed useful way to introduce such type of neglected covariates. Vaupel et al.(1979) first used frailty term as an unobserved quantity in the study of the mortality of the population.

According to Keyfitz and Littman (1979) in the study of mortality among heterogeneous population, it is observed that calculation of life expectancy from know death rate while ignoring heterogeneity may not give correct outcome. Vaupel et al.(1979) drawn to the same conclusion by using a continuous mixture model. In frailty model, the frailty term V and baseline hazard function $r_0(y)$ can not be separated since in a cluster level, the frailty is incorporated with the baseline hazard in multiplicative manner. Yin and Ibrahim (2005) proposed a new additive frailty model under the assumption that frailty has multiplicative effect on the baseline hazard. The hazard function for time $y > 0$ is given as

$$r(y|X, v) = r_0(y)v + X'\beta \quad (1.1)$$

where $r_0(y)$, β and X are the baseline hazard function, regression coefficients and known covariates. This can be expressed in another way as

$$r(y|X, v) = r_0(y)v + e^{X'\beta} \quad (1.2)$$

This is the special case of the more general class of shared frailty model (Clayton, 1978). The more interesting situation is that when there are infinite number of covariates, some heterogeneity is still present in the study population. Assuming that known covariates and frailty are additive in nature. The combined effect of known and unknown covariates has a multiplicative effect on the baseline hazard. The hazard function for a given frailty $V = v$ at time $y > 0$ is

$$r(y|X) = r_0(y)(v + e^{X'\beta}) \quad (1.3)$$

where β is the regression coefficient associated with known covariate. Then, the cumulative hazard function is written as

$$R(y|X) = R_0(y)(v + e^{X'\beta}) \quad (1.4)$$

Where $R_0(y)$ is the cumulative hazard function for the time $y > 0$. The conditional survival function given frailty V is given as

$$S(y|v) = e^{-[R_0(y)(v+e^{X'\beta})]} \quad (1.5)$$

The marginal survival function can be obtained by integrating over the range of frailty variable V having the probability density function as f(v) and is given by

$$S(y) = e^{-R_0(y)e^{X'\beta}} L_v[R_0(y)] \quad (1.6)$$

where $L_v(.)$ is the Laplace transformation of the frailty variable v.

When $v = 0$, frailty distribution is degenerated for all individuals. Under this condition, a mixture frailty model reduces to proportional hazard model. Since, the mathematical expression of Gamma distribution can be easily derived, it is most commonly used frailty distribution. The shared gamma frailty models were first recommended by Clayton (1978) for the examination of the relationship between clustered survival times in the study of disease transmission due to hereditary. The advantage is that, without covariates, its scientific properties are helpful for estimation (Oakes 1982). But it also has some limitations (Kheri 2007),whereby inverse Gaussian distribution is popularly used as frailty distribution for the parametric model. Similarities are also observed between frailty distribution and age of survivors as time increases in inverse Gaussian distribution. Further, inverse Gaussian distribution is more flexible than gamma for modeling of the

survival data. When there are more failures at the beginning of life time distribution and non-monotonic failures rate is expected, the inverse Gaussian model is more appropriate for the life time model. Gamma and inverse Gaussian distribution are more attractive because the unconditional survival function and hazard function can be expressed as simple closed form.

In this manuscript, we considered right censored data with gamma and inverse Gaussian as the frailty distributions and log-logistic distribution (LLD) as the baseline distribution to explore the salient features of the shared gamma frailty and shared inverse Gaussian frailty models. Here the dependency between the survival times is due to the frailty parameter of gamma and inverse Gaussian distributions. The degree of heterogeneity of the study population depends on the value of the frailty distribution variance. Larger variance of frailty distribution implies more heterogeneity in the population. When frailty distribution has zero variance, it is said to have degenerate distribution. The Log logistic distribution is chosen as baseline distribution due to flexibility property of the functional form.

Generally, classical approach and Bayesian approach are two commonly used techniques. Here, we adopt Markov Chain Monte Carlo(MCMC) method under Bayesian technique to estimate the model parameters as the prior distribution can be used, different properties of posterior distribution can be easily derived, interpretation of the results become easier and model choice criteria can be formulated. We also presented a simulation study to check the performance of models. All the estimation procedure and models comparison are illustrated with infectious disease data related to kidney infection.

In sections 2 and 3, introduction of a mixture shared frailty models and baseline distribution are given, followed by proposed models and estimation strategies in section 4 and section 5. In section 6, the proposed models are illustrated with simulation study. Application to real life data and discussion of the results are given in sections 7 and 8.

2. SHARED FRAILTY MODELS

Suppose there are n individuals under consideration in the study. Let (y_{1q}, y_{2q}) be the first and second survival time of the p^{th} (p = 1; 2) component of q^{th} (q = 1; 2; :::; n) individual. Then, the conditional hazard function and conditional survival function for (y_{1q}, y_{2q}) given unobserved covariates v_q are respectively,

$$r(y_{pq}|v_q, X) = r_0(y_{pq})(v_q + \eta_q) \quad (2.1)$$

$$S(y_{pq}|v_q, X) = e^{-[R0(ypq)(vq+\eta q)]} \quad (2.2)$$

where $\eta_q = e^{X_q\underline{\beta}}$. Under the assumption of independence, the bivariate survival function for the given frailty $V_q = v_q$ at time $y_{1q} > 0$ and $y_{2q} > 0$ is

$$S(y_{1q}, y_{2q}|v_q, X_q) = e^{-[(R01(y1q)+R02(y2q))(vq+\eta q)]} \quad (2.3)$$

By integrating the conditional survival function with respect to the frailty variable V_q having the probability density function f(v), we obtained the unconditional survival function as

$$S(y_{1q}, y_{2q} \mid Xq) = \int_{V_q} e^{-[R_{01}(y_{1q})+R_{02}(y_{2q})(v_q+\eta_q)]} f_v(v_q) dv_q$$

$$= e^{-[R_{01}(y_{1q})+R_{02}(y_{2q})]\eta_q} L_{Vq}[R_{01}(y_{1q}) + R_{02}(y_{2q})] \quad (2.4)$$

Where L_{Vq} (.) is the Laplace transform of the frailty variable of V_q for q^{th} individual. Here onwards $S(y_{1q},y_{2q}|X_q)$ expressed as $S(y_{1q}, y_{2q})$.

Here, we consider gamma distribution and inverse Gaussian distribution as frailty distributions with parameters ζ and ξ having probability density functions as

$$f(v) = \begin{cases} \frac{\frac{1}{\xi}^{\frac{1}{\xi}} v^{\frac{1}{\xi}-1} e^{\frac{-v}{\xi}}}{\Gamma\frac{1}{\xi}} & ; v > 0, \xi > 0 \\ 0 & ; Otherwise \end{cases} \quad (2.5)$$

$$f(v) = \begin{cases} \left[\frac{1}{2\pi\zeta}\right]^{\frac{1}{2}} v^{-\frac{3}{2}} e^{\frac{(v-\xi)^2}{2v\zeta\xi^2}} & ; v > 0, \zeta > 0, \xi > 0 \\ 0 & ; Otherwise \end{cases} \quad (2.6)$$

For the identifiability of the distributions, the expected value of the distribution is assumed to be one and having finite variance. Under this condition and by using Laplace transformation, the unconditional bivariate survival functions of mixture shared frailty models for the q^{th} individual becomes

$$S(y_{1q}, y_{2q}) = e^{-[R_{01}(y_{1q})+R_{02}(y_{2q})\eta_q]}[1 + \xi(R_{01}(y_{1q}) + R_{02}(y_{2q}))]^{-1/\xi} \quad (2.7)$$

$$S(y_{1q}, y_{2q}) = e^{-(R_{01}(y_{1q})+R_{02}(y_{2q}))\eta_q} exp\left[\frac{1-(1+2\xi(R_{01}(y_{1q})+R_{02}(y_{2q})))^{1/2}}{\xi}\right] \quad (2.8)$$

where $R_{01}(y_{1q})$ and $R_{02}(y_{2q})$ are the cumulative baseline hazard functions of the lifetime Y_{1q} and Y_{2q}.

When the frailty is absent in the model and is called as without frailty model. In this case, the model becomes

$$S(y_{1q}, y_{2q}) = e^{-(R_{01}(y_{1q})+R_{02}(y_{2q}))\eta_q} \quad (2.9)$$

3. BASELINE DISTRIBUTION

The baseline consider is log-logistic distribution because of the important properties. Generally it is used when the rate of events of interest increases initially, after reaching some peak values it declines afterward. Langlands et al. (1979) given example in the study of breast cancer where highest mortality observed three years later and having the hazard function for time Y as

$$r(y) = \frac{\frac{\lambda}{\alpha}\left(\frac{y}{\alpha}\right)^{\lambda-1}}{1+\left(\frac{y}{\alpha}\right)^{\lambda}} \quad ; y>0, \alpha > 0, \lambda > 0$$

The corresponding cumulative hazard function and survival functions are respectively,

$$R(y) = \left[1 - \left(\frac{y}{\alpha}\right)^{\lambda}\right] ; y> 0, \alpha > 0, \lambda > 0 \quad (3.1)$$

$$S(y) = \left[1 - \left(\frac{y}{\alpha}\right)^{\lambda}\right]^{-1} ; y>0, \alpha > 0, \lambda > 0 \quad (3.2)$$

where α and λ are the parameters of the log-logistic distribution. The reality of the distribution is that the cumulative distribution function can be expressed in a closed form, which is especially beneficial for evaluation of survival data with censoring (Bennett, 1983). The shape of parameter of log-logistic distribution is very akin to the log-normal distribution but is more suited for the evaluation of the survival data. This is due to the fact of its greater mathematical tractability when dealing with the censored observations which show up regularly in such data.

4. PROPOSED MODELS

The unconditional survival function is obtained by replacing the cumulative hazard functions of log-logistic distribution in equations (2.7), (2.8) and (2.9). Then,

$$S(y_{1q}, y_{2q}) = e^{-\left(ln\left[1-\left(\frac{y_{1q}}{\alpha_1}\right)^{\lambda_1}\right]+ln\left[1-\left(\frac{y_{2q}}{\alpha_2}\right)^{\lambda_2}\right]\right)\eta_q}\left[1+\xi\left(ln\left[1-\left(\frac{y_{1q}}{\alpha_1}\right)^{\lambda_1}\right]+ln\left[1-\left(\frac{y_{2q}}{\alpha_2}\right)^{\lambda_2}\right]\right)\right]^{-1/\xi} \quad (4.1)$$

$$S(y_{1q}, y_{2q}) = e^{-\left(ln\left[1-\left(\frac{y_{1q}}{\alpha_1}\right)^{\lambda_1}\right]+ln\left[1-\left(\frac{y_{2q}}{\alpha_2}\right)^{\lambda_2}\right]\right)\eta_q} e^{\left[\frac{1-\left(1+2\xi\left(ln\left[1-\left(\frac{y_{1q}}{\alpha_1}\right)^{\lambda_1}\right]+ln\left[1-\left(\frac{y_{2q}}{\alpha_2}\right)^{\lambda_2}\right]\right)\right)^{1/2}}{\xi}\right]} \quad (4.2)$$

$$S(y_{1q}, y_{2q}) = e^{-\left(ln\left[1-\left(\frac{y_{1q}}{\alpha_1}\right)^{\lambda_1}\right]+ln\left[1-\left(\frac{y_{2q}}{\alpha_2}\right)^{\lambda_2}\right]\right)\eta_q} \quad (4.3)$$

The equations (4.1) and (4.2) are mixture shared gamma frailty model and mixture shared inverse Gaussian frailty model under log-logistic baseline distribution, called as model-I and model-II. Equation (4.3) is a model without frailty and called as model-III.

5. ESTIMATION STRATEGIES

The likelihood function can be obtained by blending the failure times of the q^{th} individuals (q = 1,2,3,…,n) and censoring times (d_{1q}, d_{2q}) by assuming independence between censoring scheme and individuals lifetimes and is given by

$$L\left(\underline{\psi}, \underline{\beta}, \xi\right) = \prod_{q=1}^{n_1} f_1(y_{1q}, y_{2q}) \prod_{q=1}^{n_2} f_1(y_{1q}, d_{2q}) \prod_{q=1}^{n_3} f_1(d_{1q}, y_{2q}) \prod_{q=1}^{n_4} f_1(d_{1q}, d_{2q}) \quad (5.1)$$

where $\underline{\psi}$, $\underline{\beta}$ and ξ are vectors of baseline parameters, regression coefficients and frailty distribution parameter. The likelihood function for without frailty is given as

$$L\left(\underline{\psi}, \underline{\beta}\right) = \prod_{q=1}^{n_1} f_1(y_{1q}, y_{2q}) \prod_{q=1}^{n_2} f_1(y_{1q}, d_{2q}) \prod_{q=1}^{n_3} f_1(d_{1q}, y_{2q}) \prod_{q=1}^{n_4} f_1(d_{1q}, d_{2q}) \quad (5.2)$$

where n_1, n_2, n_3 and n_4 are the random number of observations observed to lie in the range (y_{1q}, y_{2q}) lie in the ranges $y_{1q} < d_{1q}$; $y_{2q} < d_{2q}$; $y_{1q} < d_{1q}$; $y_{2q} > d_{2q}$; $y_{1q} > d_{1q}$; $y_{2q} < d_{2q}$ and $y_{1q} > d_{1q}$; $y_{2q} > d_{2q}$ respectively and the contribution of the q^{th} individual in the likelihood function as

$$f_1(y_{1q}, y_{2q}) = \frac{\partial^2 S(y_{1q}, y_{2q})}{\partial y_{1q} \partial y_{2q}}$$

$$f_2(y_{1q}, d_{2q}) = -\frac{\partial S(y_{1q}, d_{2q})}{\partial y_{1q}}$$

$$f_3(d_{1q}, y_{2q}) = -\frac{\partial S(d_{1q}, y_{2q})}{\partial y_{2q}}$$

$$f_4(d_{1q}, d_{2q}) = S(d_{1q}, d_{2q}) \quad (5.3)$$

Replacing the survival function in equation (5.3) and differentiating it, we get the likelihood functions given in equations (5.1) and (5.2). The first expression is a likelihood function for a mixture shared frailty model and second expression is likelihood function for without frailty model.

The joint posterior density of the parameters given failure times is given as

$$\pi\left(\alpha_1, \lambda_1, \alpha_2, \lambda_2, \xi, \underline{\beta}\right) \propto L\left(\alpha_1, \lambda_1, \alpha_2, \lambda_2, \xi, \underline{\beta}\right) \times g_1(\alpha_1) g_2(\lambda_1) g_3(\alpha_2) g_4(\lambda_2) g_5(\xi) \prod_{i=1}^{5} p_i(\underline{\beta_i})$$

where $g_i(.)$ (i = 1, 2,…, 5) indicates the prior density function with known hyper parameters of corresponding arguments for baseline parameters and frailty variance; $p_i(.)$ is prior density function for regression coefficient β_i; $\underline{\beta_i}$ represents a vector of regression coefficients except β_i, i = 1, 2,…, a and likelihood function L(.) is given by equations (5.1) or (5.2). Here it is assumed that all the parameters are independently distributed.

The expression of the likelihood function in equations (5.1) and (5.2) are not easy to solve by using Newton-Raphson method. MLEs fail to converge as it involved large number of parameters. Therefore, Bayesian approach was utilized to estimate the parameters involved in the models, which does not endure any such kind of troubles.

Prior distributions are used as follows - gamma distribution with mean 1 and large variance $\Gamma(\Psi, \Psi)$ is used as prior distribution for frailty parameter with a small value of Ψ. Normal distribution with mean zero and large variance say Φ^2 is used as prior for the regression coefficient. The same type of prior distributions considered in Ibrahim et al. (2001) and Sahu et al. (1997) and non-informative prior assumed as the baseline parameters since we do not have any information about the baseline parameters. $\Gamma(a_1, b_1)$ and $U(a_2, b_2)$ are used as non-informative prior distributions. All the hyper-parameters Ψ, Φ, a_1, a_2, b_1 and b_2 are assumed to be known. Here $\Gamma(a_1, b_1)$ represents gamma distribution with shape parameter a_1 and scale parameter b_1 and $U(a_2, b_2)$ is the uniform distribution over the interval a_2 to b_2. We set hyper-parameters as Ψ = 0:0001, Φ^2 =1000, a_1 = 1, b_1 = 0:0001, a_2 = 100, and b_2 = 100.

To estimate the parameters in the models fitted with the above prior density function and likelihood equations (5.1) and (5.2), Metropolis Hasting Algorithm and Gibbs Sampler was utilized. The convergence of the Markov chain to a stationary distribution is also observed by Geweke test and Gelman-Rubin Statistics as suggested by Geweke (1992) and Gelman et al. (1992). To check the behavior of the chain, to decide burn-in period and autocorrelation lag, we used trace plots, coupling from the past plots and sample autocorrelation plots respectively.

To decide the model which provides the best fit to the dataset, model comparison was done by using Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and Deviance Information Criteria (DIC) and Bayes factor.

Suppose there are P parameters in a model and n observations in a dataset. AIC, BIC and DIC are elucidated as

$$\text{AIC} = -2\log L\left(y|\tilde{\theta}\right) + 2P \quad (5.4)$$

$$\text{BIC} = -2\log L\left(y|\tilde{\theta}\right) + \log(n)\, P \quad (5.5)$$

$$\text{DIC} = -2\log L\left(y|\tilde{\theta}\right) + 2P_D \quad (5.6)$$

where P_D = E[2log L(y|θ)] -2log $L\left(y|\tilde{\theta}\right)$

Smaller values of AIC, BIC and DIC for the models are considered as better models than higher values.

Bayes factor also employed for selection of Model M_u against Model M_v and defined as

$$BF_{uv} = \frac{P(y|M_u)}{P(y|M_v)} \quad (5.7)$$

Where

$$P(y|M_k) = \int P(y|\Omega, M_k)\pi(\Omega|M_k)d\Omega$$

where Ω represents the number of unknown parameters, $\pi(\Omega|M_k)$ is the density of prior distribution.

In spite of the fact that, $2\log BF_{uv}$ is roughly equal to the differences in the values of BIC for the given models, we utilized the strategy given by Kass and Raftery (1995), to compute P(y|M) from the MCMC sample gotten from each of the model parameters.

$$P(y|M) = \left(\frac{\sum_{k=1}^{N} L(y|\Omega^k)^{-1}}{N}\right)^{-1}$$

where Ω^k and N symbolize sample and sample size of posterior distribution.

A value of $2\log BF_{uv}$ more than 10 shows that a very strong confirmation to favour model M_u over model M_v, whereas a value between 0 and 2 is adequate to prove to favour any of the model. A value between 2 and 6 or 6 and 10 shows a mild or strong confirmation respectively, to favor the numerator model.

6. SIMULATION STUDY

To examine the performance of the Bayesian estimation method a simulation study was carried out. Only one covariate X1 was considered for the simulation purpose and was assumed to follow binomial distribution for Model I and Model II. Since we do not have any prior information about the baseline parameters, α_1, λ_1, α_2 and λ_2, the prior distributions are assumed to be at. We consider two different non-informative prior distributions for the baseline parameters, one is $G(a_1, a_2)$ and another is $U(b_1, b_2)$. All the hyper-parameters a_1, a_2, b_1, and b_2 are known. Here G(a, b) is the gamma distribution with the shape parameter a and the scale parameter b and $U(b_1, b_2)$ represent the uniform distribution over the interval b_1 to b_2. For Model I, we set α_1 = 65, λ_1 = 1.3, α_2 = 65, λ_2 = 1.3, ξ = 0.12, and β = - 0.13 and censoring distribution as the exponential distribution with the parameter 0.05 each. For Model II, we set α_1 = 65, λ_1 = 1.4, α_2 = 60, λ_2 = 1.3, ξ = 0.11, and β = -0.22 and censoring distribution as the exponential distribution with the parameter 0.05 each. We assume the value of the hyper-parameters as a_1 = 1, a_2 = 0:0001, b_1 = 0, and b_2 = 100. As the Bayesian strategies are time consuming, fifty sets of lifetimes were generated utilizing inverse transform procedure. Both the chains were iterated for 100000 times. Gelman-Rubin scale reduction factor values were very close to one and p-values for Geweke test values were huge, which sufficiently demonstrates that the chains achieve stationary distribution for both the prior sets. Further the convergence rate was not enormously diverse. There was no impact of prior distribution on posterior summaries because estimates of parameters were

about the same. Table 2, and 3 present the posterior summaries of mixture shared gamma frailty and mixture shared inverse Gaussian frailty models with log-logistic baseline distribution. It provides estimates (posterior means), standard errors and upper and lower credible limits.

7. APPLICATION IN REAL LIFE DATA

The applicability of the models was checked by applying them to the infectious disease data related to infection of kidney that happens during insertion of catheter(McGilchrist and Aisbett, 1991). It consists of the first and second recurrence time of infection from the use of catheters using portable dialysis equipment for 38 patients. These two times of infection are grouped together for each patient in a cluster. The other relevant information are censoring time of infection, age of patients, gender (0 for male and 1 for female), disease types such as Glomerulo Nephritis(GN), Acute Nephritis (AN) and Polycystic Kidney Disease (PKD).

First, Kolmogorov Smirnov test was used to check the goodness of fit for kidney infection data and the p-values obtained for the first and second recurrences are large enough to say that there is no reason to reject the hypothesis that the first and second recurrence time to follow the log-logistic distribution in the univariate case, we assume to be valid for the bivariate case also. The corresponding p-values are given in Table 1.

As in simulation study, we run two parallel chains for all models using two sets of prior distributions with the different starting points using the Metropolis-Hastings algorithm and the Gibbs sampler based on normal transition kernels. We iterate both the chains for 100000 times. As seen in the simulation study here also we got nearly the same estimates of parameters for both the set of prior, so estimates are not dependent on the different prior distributions. The convergence rate of the Gibbs sampler for both the prior sets is almost the same. Also both the chains shows somewhat similar results, so we present here the analysis for only one chain with $G(a_1, b_1)$ as prior for the baseline parameters, for all the models. We are also calculate Gelman Rubin statistic values and Geweke test statistics values to check the convergence of Markov chain to a stationary distribution. Gelman Rubin statistics values are closed to one, Geweke test statistics values are closed to zero and their corresponding p-values are large enough to say that the chains attains stationary distribution. Trace plots and coupling from the past plots given in Figures 1 and 2. The largest values of GR statistics are 1.0045, 1.0003 and 1.0013, which indicating that 6400, 8200 and 7700 iterations would be satisfactory burn-in-period. Autocorrelation plot used to decide autocorrelation lag and the convergence of the chain to a stationary distribution also confirm by running mean plot given in Figures 3 and 4. The estimate of parameters (posterior mean), standard error, credible limits are given in Tables 4, 5 and 6. Since the credible intervals does not contains zero, all the factors are significant. The positive value of β_1 indicates that age is significant factors for infection of kidney, as age increases chance of infection also increase. Negative value of β_2 shows that female has a lower chance of infection than male. Another significant factors, which have higher chance of infection are disease types such as Glomerulo Nephritis(β_3), Acute Nephritis (β_4) and Polycystic Kidney Disease (β_5). When the variances of shared frailty term under gamma distribution are compare for both the models, it appears that the estimate of the variances tend to be overestimated for the existing shared gamma frailty model.

The better model as per AIC, BIC and DIC values is model-I, since it has lower values of AIC, BIC and DIC than model-II and model-III under log-logistic baseline distribution which is given in Table 7. But the distinction between AIC, BIC, and DIC values for model-I, model-II and model-III are small, AIC, BIC, and DIC values may not be enough to differentiate the models. Presently, we consider Bayes factor for comparing a pair of models u and v. The values in Table 8 shows that model-I is better than model-II and model-III as the corresponding value of $2\log(B_{uv})$ are greater than 2 and 10 indicating that there is mild and very strong confirmation to favor model-I than model-II and model-III for the given dataset, which affirm our earlier results given in Table 7. Hence from all the demonstrated comparison criteria we can say model-I (mixture shared gamma frailty model) is better than model-II and model-III that is mixture shared inverse Gaussian frailty model and without frailty model under log-logistic distribution as baseline for modeling kidney infection data.

8. DISCUSSION

In this study, we examine a new mixture shared gamma and inverse Gaussian frailty models and existing shared frailty models under the same log-logistic baseline distribution.

The Metropolis-Hastings and Gibbs sampler was utilized to fit all the proposed models. Kidney infection data was analyzed using the proposed models and the finest model is suggested. We have utilized self-composed programs in R statistical software to perform the analysis.

All the demonstrated comparison criteria exhibits that a mixture shared gamma frailty model with log-logistic baseline is better for modeling of kidney infection data than mixture shared inverse Gaussian frailty model and without frailty model under the same baseline distribution. The estimates of frailty parameters are high in all models which are 0.0186 and 0.0190 for mixture shared gamma frailty model and mixture shared in verse Gaussian frailty model respectively. This indicates that there is a strong evidence that heterogeneity is present among the patients. A few patients are anticipated to be exceptionally inclined to infection compared to others with the same covariate values. We can further establish that there is a strong positive relationship between the two infection times for the same patient. So, we have a new model called a mixture shared gamma frailty model to analyse kidney infection data. We compare with shared gamma frailty model and shared inverse frailty model based on the same baseline distribution. It is observe and worth to mention that our proposed models perform better than under log-logistic as baseline distribution proposed by Hanagal and sharma (2012) and Hanagal and Sharma (2015).

REFERENCES

[1] Bennett,S., 1983, "Log-logistic Regression Models for Survival Data," J. R. Statist. Soc. Ser. C. Appl. Statist, 32(2), pp. 165-171.

[2] 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," Biometrica, 65, pp. 141-151.

[3] Gelman,A. , and Donald,B.R., 1992, "A single series from the Gibbs sampler provides a false sense of security. In Bayesian Statistics 4(J.M.Bernardo,J.0.Berger,A.P.Dawid and A.F.M.Smith, eds.)," Oxford, Oxford Univ.Press, pp. 625-632.

[4] 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.

[5] Hanagal,D.D., and Richa,S., 2012, "Modeling heterogeneity for bivariate survival data by shared gamma frailty regression model," Model Assisted Statistics and Applications, 8(2), pp. 85-102.

[6] Hanagal,D.D., and Richa,S., 2015, "Analysis of bivariate survival data using shared inverse Gaussian frailty model,"Communication in Statistics-Theory and Methods, 44(7), pp. 1351-1380.

[7] Ibrahim,J.G., Ming-Hui,C., and Debajyoti,S., 2001, "Bayesian Survival Analysis," Springer, Verlag.

[8] Kass,R.E., and Adrian,E.R., 1995, "Bayes Factors," Journal of the American Statistical Association, 90(430), pp. 773-95.

[9] Key_tz,N., and Littman,G., 1979,"Mortality in a Heterogeneous Population," Population Studies, 33, pp. 333-342.

[10] Kheiri,S., Alan,K., and Mohammad,R.M., 2007, "Bayesian analysis of an inverse Gaussian correlated frailty model," Computational Statistics and Data Analysis, 51, pp. 5317-5326.

[11] Kleiber C., and Kotz S., 2003, "Statistical Size Distributions in Economics and Actuarial Sciences," Hoboken NJ, Wiley-Interscience.

[12] Langlands,A.O., Pocock,S.J., Kerr,G.R., and Gore,S.M., 1979, "Long-Term Survival of Patients with Breast Cancer: A Study of the Curability of the Disease," British Medical Journal, 2, pp. 1247-1251.

[13] McGilchrist,C.A., and Aisbett,C.W., 1991,"Regression with frailty in survival analysis," Biometrics, 47, pp.461-466.

[14] Yin,G., and Ibrahim,J.G., 2005, "A Class of Bayesian Shared Gamma Frailty Models with Multivariate Failure Time Data," Biometrics, 61, pp. 208-216.

[15] Oakes,D., 1982, "A model for association in bivariate survival data, "J.R.Statist.Soc.B, 44, pp. 414-22.

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

[17] Sahu,S.K., Dey,.K., Aslanidou,H., and Sinha,., 1997, "A Weibull regression model with gamma frailties for multivariate survival data," Life time data analysis, 3, pp. 123-137.

Appendix : Tables and Figures

Table 1: Simulation under mixture shared gamma frailty model

Parameter	Estimates	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke Values	P Values	Gelman and Rubin values
burn in period = 5800; autocorrelation lag = 210							
α_1(66.82)	66.8165	0.3266	66.1731	67.4570	-0.0013	0.4994	1.0000
λ_1(98.66)	98.6500	0.5755	97.7302	99.6378	-0.0192	0.4923	1.0010
α_2(1.12)	1.1167	0.0547	1.0180	1.2053	-0.0192	0.4929	1.0005
λ_2(1.40)	1.4045	0.0557	1.3155	1.5036	0.0018	0.5007	1.0049
ξ(0.018)	0.0183	0.0060	0.0101	0.0290	-0.0035	0.4985	1.0065
β(-0.13)	-0.1272	0.0602	-0.2261	-0.0298	0.0078	0.5031	1.0046

Table 2: Simulation under mixture shared inverse Gaussian frailty model

Parameter	Estimates	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke Values	P Values	Gelman and Rubin values
burn in period = 5300; autocorrelation lag = 215							
α_1(62.64)	62.6490	0.3320	61.9894	63.3069	0.0116	0.5046	1.0008
λ_1(92.40)	92.4043	0.5315	91.4744	93.3411	0.0081	0.5032	1.0000
α_2(1.81)	1.8146	0.2995	1.2649	2.4986	0.0021	0.5008	1.0003
λ_2(1.17)	1.1703	0.1886	0.8289	1.5774	0.0153	0.5061	0.9999
ξ(0.019)	0.0192	0.0050	0.0105	0.0286	-0.0031	0.4987	1.0007
β(-0.12)	-0.1177	0.0547	-0.2036	0.0021	0.0133	0.5053	1.0017

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

Model	Recurrences	
	First	Second
Mixture shared gamma frailty	0.5722	0.8412
Mixture shared inverse Gaussian frailty	0.3516	0.7356

Table 4: Posterior results under mixture shared gamma frailty model

Parameter	Estimates	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke Values	P Values	Gelman and Rubin values
burn in period = 6400; autocorrelation lag = 245							
α_1	69.4135	0.3110	68.8223	70.0489	0.0020	0.5008	0.9999
λ_1	98.7955	0.6180	97.8261	99.7407	0.0080	0.5032	1.0043
α_2	1.2127	0.0564	1.1162	1.3035	0.0007	0.5002	1.0045
λ_2	1.4250	0.0535	1.3360	1.5231	0.0079	0.5031	1.0027
ξ	0.0186	0.0052	0.0106	0.0288	-0.0122	0.4951	0.9999
β_1	-0.1320	0.0160	-0.1650	-0.1064	-0.0036	0.4985	1.0000
β_2	-12.5289	3.5262	-18.8690	-6.4163	0.0054	0.4985	1.0024
β_3	7.6491	0.5179	6.7530	8.6226	0.0017	0.5006	1.0004
β_4	7.1690	0.5143	6.2625	8.1155	0.0105	0.5042	0.9999
β_5	0.2059	0.0498	0.1215	0.2957	-0.0021	0.4991	0.9999

Table 5: Posterior results under mixture shared inverse Gaussian frailty model

Parameter	Estimates	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke Values	P Values	Gelman and Rubin values
burn in period = 8200; autocorrelation lag = 85							
α_1	69.1259	0.3339	68.4878	69.7877	-0.0050	0.4979	1.0003
λ_1	92.3842	0.5421	91.4749	93.3407	0.0072	0.5029	1.0000
α_2	1.2373	0.1688	0.9283	1.5676	-0.0012	0.4994	1.0000
λ_2	1.4736	0.2318	1.0526	1.9346	-0.0020	0.4991	1.0000
ξ	0.0190	0.0046	0.0107	0.0282	-0.0026	0.4989	1.0001
β_1	-0.1062	0.0192	-0.1567	-0.0787	-0.0068	0.4972	1.0002
β_2	-12.197	3.6825	-19.147	-5.7491	-0.0061	0.4972	1.0000
β_3	6.1190	0.4784	5.2417	6.9979	0.0087	0.5034	1.0000
β_4	6.0600	0.4636	5.2192	6.9731	0.0130	0.5052	1.0001
β_5	1.1958	0.4680	0.3175	2.0715	0.0015	0.5006	1.0003

Table 6: Posterior results under without frailty model

Parameter	Estimates	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke Values	P Values	Gelman and Rubin values
burn in period = 7700; autocorrelation lag = 140							
α_1	50.0306	0.3333	49.3699	50.6603	0.0027	0.5010	1.0000
λ_1	58.5419	0.5529	57.5641	59.4531	-0.0015	0.4993	1.0001
α_2	1.1960	0.1590	0.9022	1.5193	-0.0010	0.4995	1.0006
λ_2	1.1960	0.2262	1.0491	1.9633	0.0028	0.5011	1.0013
β_1	-0.1347	0.0294	-0.2087	-0.0954	0.0001	0.5000	1.0001
β_2	-10.053	2.4872	-14.6149	-5.6510	0.0122	0.5000	1.0000
β_3	6.1265	0.4836	5.2182	6.9830	-0.0159	0.4936	1.0000
β_4	6.3321	0.4571	5.4939	7.2916	0.0061	0.5024	1.0000
β_5	0.2094	0.0512	0.1181	0.3011	-0.0004	0.4998	1.0000

Table 7: AIC, BIC and DIC values for all models

Model No.	AIC	BIC	DIC	Log-likelihood
Model I	691.7195	708.0953	674.7153	-335.8597
Model II	691.5905	707.9664	678.4154	-335.7953
Model III	697.6049	712.3432	684.9765	-339.8025

Table 8: Bayes factor values and decision for test of significance for frailty fitted to Kidney Infection Data Set

Numerator model	$2\log_e(B_{uv})$	Range	Evidence against against denominator model
M_I against M_{II}	4.877261	≥ 2 and ≤ 10	Positive
M_I against M_{III}	10.09704	≥ 10	Very strong Positive
M_{II} against M_{III}	5.219776	≥ 2 and ≤ 10	Positive

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1088-1097
© Research India Publications. https://dx.doi.org/10.37622/IJAER/15.11.2020.1088-1097

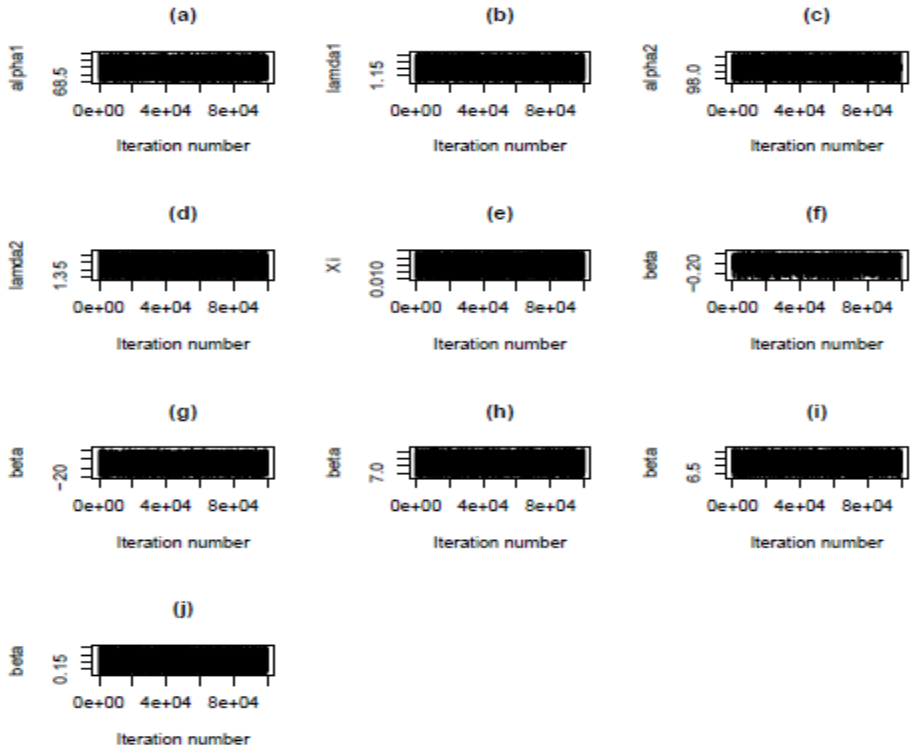

Figure 1: Trace plots of model-I

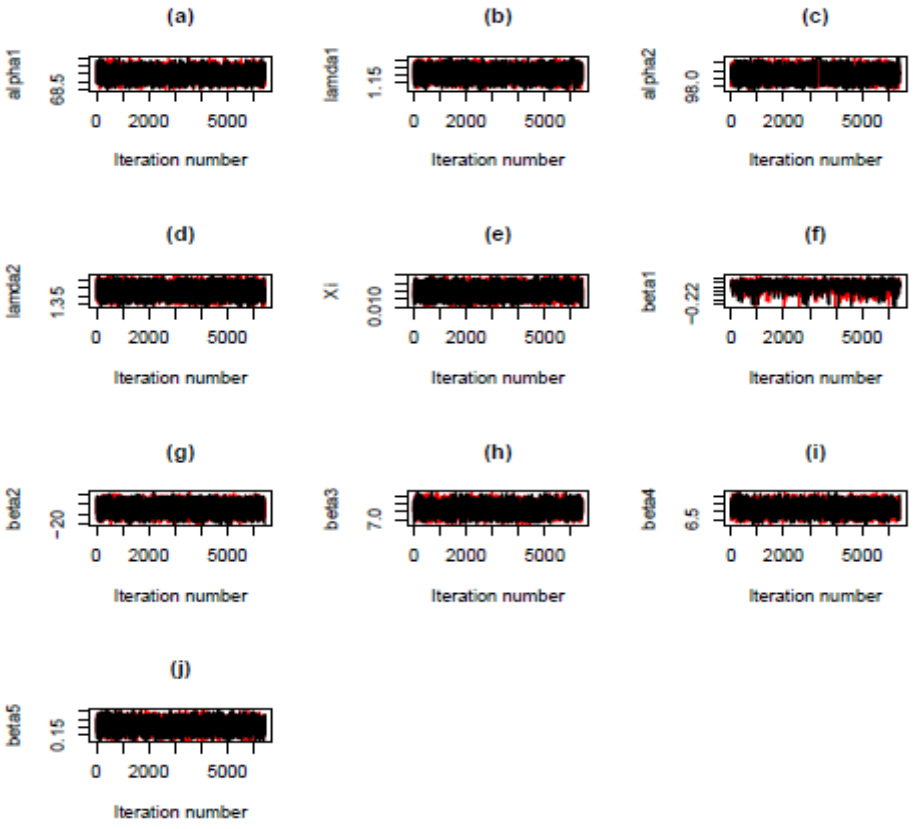

Figure 2: Coupling from the past plots of model-I

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1088-1097
© Research India Publications. https://dx.doi.org/10.37622/IJAER/15.11.2020.1088-1097

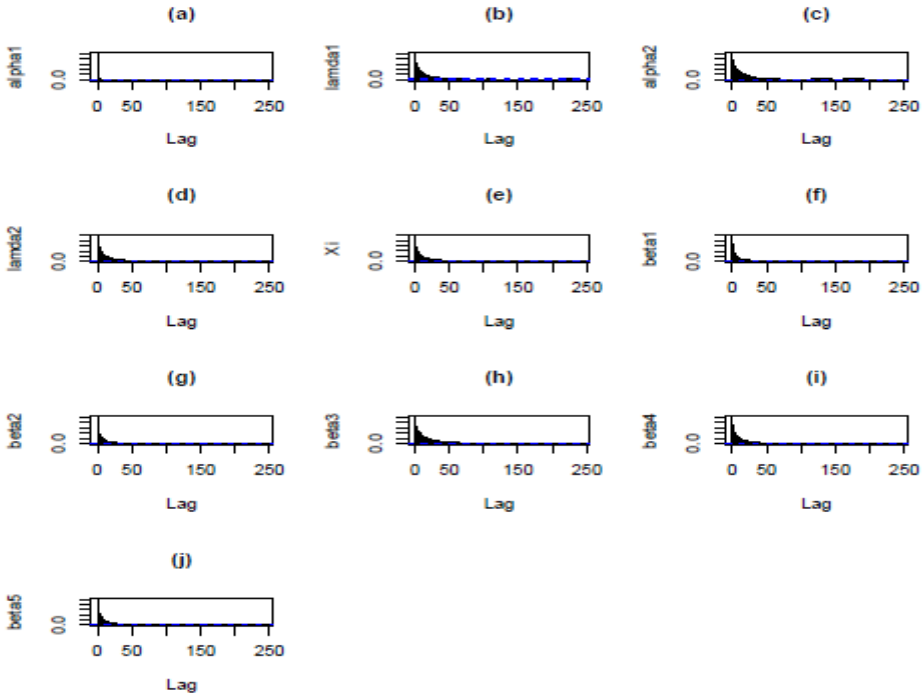

Figure 3: ACF plots of model-I

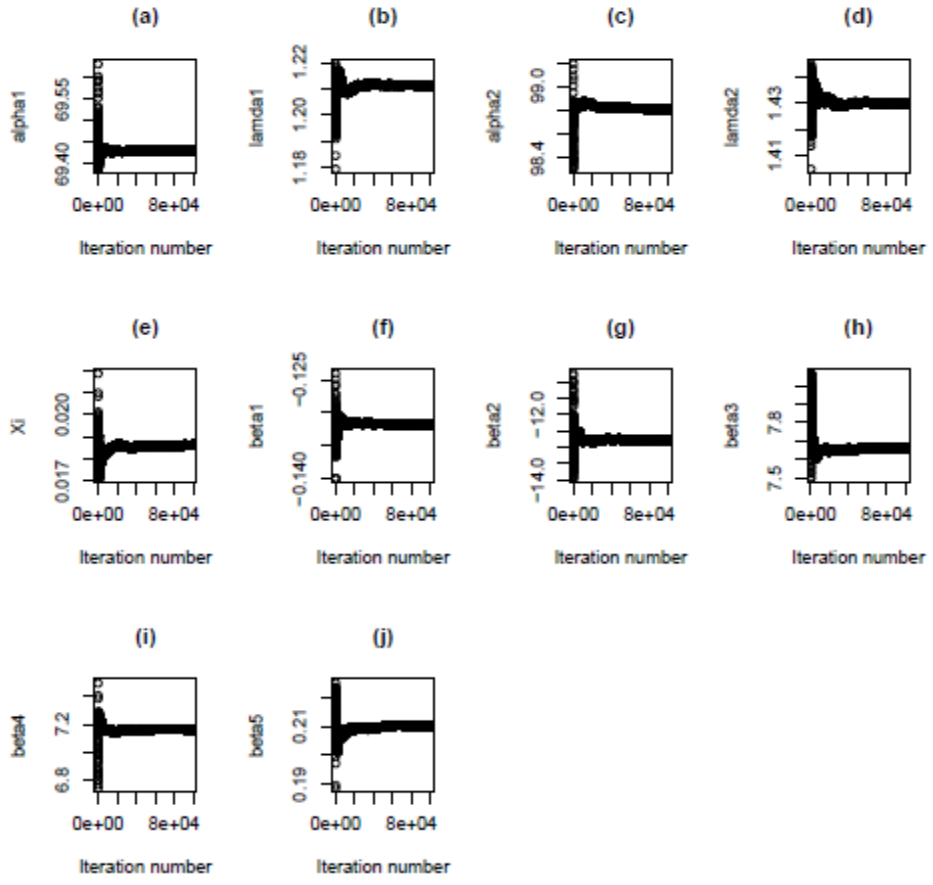

Figure 4: Running mean plots of model-I