

A Mixture Shared Inverse Gaussian Frailty Model under Modified Weibull Baseline Distribution

Lalpawimawha

PUC, Mizoram University, India

Arvind Pandey

Central University of Rajasthan, India

Abstract

Frailty models are used in the survival analysis to account for the unobserved heterogeneity in individual risks to disease and death. To analyze the bivariate data on related survival times (e.g. matched pairs experiments, twin or family data), the shared frailty models were suggested. In this manuscript, we propose a new mixture shared inverse Gaussian frailty model based on modified Weibull as baseline distribution. The Bayesian approach of Markov Chain Monte Carlo technique is employed to estimate the parameters involved in the models. In addition, a simulation study is performed to compare the true values of the parameters with the estimated values. A comparison with the existing model was done by using Bayesian comparison techniques. A better model for infectious disease data related to kidney infection is suggested.

Keywords: mixture frailty model, Bayesian approach, inverse Gaussian frailty, modified Weibull distribution, MCMC.

1. Introduction

While studying the effect of age at the onset of diseases and other related risk factors, it is not possible to assume all the individuals to have the same risk of getting a particular disease. Under this condition, the population is considered as a mixture of groups of individuals having different frailties that caused heterogeneity in the study population. Neglecting such type of frailty in real life experiment often lead to an incorrect conclusion especially in the study of survival analysis. However, most researchers in the field of medical sciences are unaware of the importance of the effects of heterogeneity in data analysis. Clayton (1978) first designed a model to account for such unobserved covariates. The term frailty was first introduced by Vaupel, Manton, and Stallard (1979) in the study of mortality.

Shepard and Zeckhauser (1977) showed that neglecting heterogeneity produces overestimates of the effects on life expectancy of a given medical improvement. Keyfitz and Littman (1979) also showed that ignoring heterogeneity will lead to an incorrect calculation of the life expectancy from known death rates. The same conclusion was drawn by Vaupel *et al.* (1979) using an infinite mixture model in which unobserved nonnegative random frailty represents all individuals in endowment for longevity. Aalen (1980) first suggested additive hazard model by assuming linear effects of the covariates term to the baseline hazard function. Lin and Ying

(1994) also proposed a new additive hazard model by assuming the failure times are independent. Hanagal and Pandey (2016) studied the shared frailty models under the assumption that frailty term acting additively to the baseline hazard function by taking gamma distribution as frailty distribution. Hanagal and Pandey (2017) also discussed the additive hazard model associated with the frailty term by considering frailty distributions as gamma and inverse Gaussian under generalized log-logistic, exponential power and generalized Weibull as baseline distributions. In the frailty model, the frailty term V and baseline hazard function $r_0(y)$ cannot be separated since in a cluster level, the frailty is incorporated with the baseline hazard in a multiplicative manner. Therefore, the hazard function at the individual level of a new mixture frailty model for the failure time Y and observable covariate X is expressed as

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

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

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

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

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

When $v = 0$, frailty distribution is degenerated for all individuals. Under this condition, a mixture frailty model reduces to proportional hazard model.

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

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

where $L_v(\cdot)$ is the Laplace transform of the frailty variable v .

Due to the simplicity of mathematical expression, Gamma distribution is the most commonly used frailty distribution. The shared gamma frailty models were first recommended by Clayton (1978) for the investigation of the relationship between clustered survival times in the study of disease transmission due to hereditary. An advantage is that without covariates its scientific properties are helpful for estimation (Oakes 1982). However, this frailty model also presents certain limitations (Kheiri, Kimber, and Meshkani 2007). The inverse Gaussian distribution is popularly used as frailty distribution for parametric model when the Gamma frailty model presents such limitations. 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 lifetime distribution and non-monotonic failure rate is expected, the inverse Gaussian model is more appropriate for the lifetime model. Gamma and inverse Gaussian distribution are more attractive as unconditional survival function and hazard functions can be expressed as simple closed form.

In this article, we consider right censored data with inverse Gaussian as the frailty distribution and modified Weibull distribution as the baseline distribution to explore the salient features of the shared inverse Gaussian frailty model. Here the dependency between survival times is due to the inverse Gaussian distributed frailty variable. The degrees of heterogeneity of the study population depends on variance value of the frailty distribution. Larger variance of frailty distribution implies more heterogeneity. When frailty distribution has zero variance, it is said to have degenerate distribution. The modified Weibull distribution was chosen as baseline distribution since the hazard function has flexible property, which is one of the important properties in the real-life data analysis.

Generally, the classical approach and the Bayesian approach are two commonly used techniques. Here, we adopted Markov Chain Monte Carlo (MCMC) method under Bayesian technique to estimate the model parameters as 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 the models. All the estimation procedure and models comparison were illustrated with infectious disease data related to kidney infection.

In sections 2 and 3, we give the introduction of a mixture shared inverse Gaussian frailty model and shared inverse Gaussian frailty model, followed by baseline distribution and proposed models in section 4 and section 5. In section 6, we discuss estimation strategies. We present simulation study and application to real-life data in sections 7 and 8 and section 9 for discussion of the results.

2. A mixture shared inverse Gaussian frailty model

Suppose n individuals are observed for the study and let a bivariate random variable (y_{1q}, y_{2q}) present the first and second survival times of the p^{th} ($p = 1, 2$) component of q^{th} ($q = 1, 2, \dots, n$) individual. Then, the conditional hazard function and conditional survival function for (y_{1q}, y_{2q}) given unobserved covariates v_q are respectively,

$$\begin{aligned} r(y_{pq}/v_q, X) &= r_0(y_{pq})(v_q + \eta_q) \\ S(y_{pq}/v_q, X) &= e^{-[R_0(y_{pq})(v_q + \eta_q)]} \end{aligned}$$

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

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

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

$$\begin{aligned} S(y_{1q}, y_{2q} | X_q) &= \int_{V_q} e^{-[(R_{01}(y_{1q}) + R_{02}(y_{2q}))(v_q + \eta_q)]} f_v(v_q) dv_q \\ &= e^{-(R_{01}(y_{1q}) + R_{02}(y_{2q}))\eta_q} L_{V_q}[R_{01}(y_{1q}) + R_{02}(y_{2q})] \end{aligned}$$

where $L_{V_q}(\cdot)$ 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 inverse Gaussian distribution as frailty distribution with parameters ζ and ξ having probability density function as

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

For the identifiability of the distribution, the expected value of the distribution is assumed to be one and have finite variance. By using Laplace transformation, the unconditional bivariate survival function of mixture shared inverse Gaussian frailty model for the q^{th} individual becomes

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

where $R_{01}(t_{1q})$ and $R_{02}(t_{2q})$ are the cumulative baseline hazard functions of the life time random variables Y_{1q} and Y_{2q} respectively.

3. Shared inverse Gaussian frailty model

The unconditional survival function for shared gamma frailty model, where the bivariate frailty distribution (V_1, V_2) depicted by $V = V_1 = V_2$ is given as

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

where ξ stand for the variance of V .

And the model without frailty becomes

$$S(y_{1q}, y_{2q}) = e^{-[(R_{01}(y_{1q}) + R_{02}(y_{2q}))\eta_q]} \quad (3)$$

4. Baseline distribution

Generally, in the parametric model, it is assumed that baseline hazard $r_0(y)$ is a parametric function. Here, modified Weibull distribution is considered as the baseline distribution proposed by [Sarhan and Zain-Din \(2008\)](#), which is the generalization of Weibull distribution and applied in data related to fatigue life and having the hazard function for time Y as

$$r(y) = \alpha + \lambda\gamma y^\gamma \quad ; y > 0, \alpha > 0, \lambda > 0, \gamma > 0$$

The corresponding cumulative hazard function and survival functions are respectively,

$$\begin{aligned} R(y) &= \alpha y + \lambda y^\gamma \quad ; y > 0, \alpha > 0, \lambda > 0, \gamma > 0 \\ S(y) &= e^{-\alpha y - \lambda y^\gamma} \quad ; y > 0, \alpha > 0, \lambda > 0, \gamma > 0 \end{aligned} \quad (4)$$

where α , λ and γ are the parameters of the modified Weibull distribution. The distribution generalized different distributions such as exponential, Rayleigh, linear failure rate and Weibull distributions depends on the value of the parameters. The hazard function is either constant as $\gamma = 1$ or increasing as $\gamma > 1$ or decreasing as $\gamma < 1$.

5. Proposed models

The unconditional survival functions are obtained by replacing the cumulative hazard function of modified Weibull distribution in equations (1), (2) and (3) respectively. Then,

$$S(y_{1q}, y_{2q}) = e^{-(\alpha_1 y_{1q} + \lambda_2 y_{2q}^{\gamma_2} + \alpha_2 y_{2q} + \lambda_2 y_{2q}^{\gamma_2})\eta_q} \exp\left[\frac{1 - (1 + 2\xi(\alpha_1 y_{1q} + \lambda_2 y_{2q}^{\gamma_2} + \alpha_2 y_{2q} + \lambda_2 y_{2q}^{\gamma_2})^{1/2})}{\xi}\right] \quad (5)$$

$$S(y_{1q}, y_{2q}) = \exp\left[\frac{1 - (1 + 2\xi\eta_q(\alpha_1 y_{1q} + \lambda_2 y_{2q}^{\gamma_2} + \alpha_2 y_{2q} + \lambda_2 y_{2q}^{\gamma_2})^{1/2})}{\xi}\right] \quad (6)$$

$$S(y_{1q}, y_{2q}) = \exp\left[-\xi\left(\alpha_1 y_{1q} + \lambda_2 y_{2q}^{\gamma_2} + \alpha_2 y_{2q} + \lambda_2 y_{2q}^{\gamma_2}\right)\eta_q\right] \quad (7)$$

The equations (5) and (6) are mixture shared inverse Gaussian frailty model and shared inverse Gaussian frailty model under modified Weibull baseline distribution, called as model-I and model-II. Equation (7) is without frailty model under the same baseline distribution and called as model-III.

6. Estimation strategies

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

and individuals lifetimes and is given by

$$\mathcal{L}(\underline{\Psi}, \underline{\beta}, \xi) = \prod_{q=1}^{n_1} f_1(y_{1q}, y_{2q}) \prod_{q=1}^{n_2} f_2(y_{1q}, d_{2q}) \prod_{q=1}^{n_3} f_3(d_{1q}, y_{2q}) \prod_{q=1}^{n_4} f_4(d_{1q}, d_{2q}) \quad (8)$$

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

$$\mathcal{L}(\underline{\Psi}, \underline{\beta}) = \prod_{q=1}^{n_1} f_1(y_{1q}, y_{2q}) \prod_{q=1}^{n_2} f_2(y_{1q}, d_{2q}) \prod_{q=1}^{n_3} f_3(d_{1q}, y_{2q}) \prod_{q=1}^{n_4} f_4(d_{1q}, d_{2q}) \quad (9)$$

where n_1, n_2, n_3 and n_4 are the random number of observations observed to 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

$$\begin{aligned} f_1(y_{1q}, y_{2q}) &= \frac{\partial^2 S(y_{1q}, y_{2q})}{\partial y_{1q} \partial y_{2q}} \\ f_2(y_{1q}, d_{2q}) &= -\frac{\partial S(y_{1q}, d_{2q})}{\partial y_{1q}} \\ f_3(d_{1q}, y_{2q}) &= -\frac{\partial S(d_{1q}, y_{2q})}{\partial y_{2q}} \\ f_4(d_{1q}, d_{2q}) &= S(d_{1q}, d_{2q}) \end{aligned} \quad (10)$$

Putting equation (10) in equations (8) and (9), we get the likelihood functions for mixture shared frailty model, shared frailty model and for without frailty model respectively.

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

$$\begin{aligned} \pi(\alpha_1, \lambda_1, \alpha_2, \lambda_2, \xi, \underline{\beta}) &\propto L(\alpha_1, \lambda_1, \gamma_1, \alpha_2, \lambda_2, \gamma_2, \xi, \underline{\beta}) \\ &\times g_1(\alpha_1) g_2(\lambda_1) g_3(\gamma_1) g_4(\alpha_2) g_5(\lambda_2) g_6(\gamma_2) g_7(\xi) \prod_{i=1}^5 p_i(\underline{\beta}_i) \end{aligned}$$

where $g_i(\cdot)$ ($i = 1, 2, \dots, 7$) indicates the prior density function with known hyperparameters of corresponding arguments for baseline parameters and frailty variance; $p_i(\cdot)$ is prior density function for regression coefficient β_i ; $\underline{\beta}_i$ represents a vector of regression coefficients except β_i , $i = 1, 2, \dots, a$ and likelihood function $\mathcal{L}(\cdot)$ is given by equation (8) or (9). Here it is assumed that all the parameters are independently distributed.

The expression of the likelihood functions in equations (8) and (9) are not easy to solve by using the Newton Raphson method. MLEs failed to converge as it involves a large number of parameters. Thus, in our problem a Bayesian approach, which does not suffer from these difficulties, is a natural one, even though it is relatively computationally intensive.

Prior distributions are used as follows - gamma distribution with mean 1 and large variance $\Gamma(\Psi, \Psi)$ is used as a prior distribution for frailty parameter with a small value of Ψ . Normal distribution with mean zero and large variance is used as prior for the regression coefficient, say φ^2 . The same type of prior distributions considered in [Ibrahim, Chen, and Sinha \(2001\)](#) and [Sahu, Dey, Aslanidou, and Sinha \(1997\)](#) 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 $\Psi, \varphi, a_1, a_2, b_1$ 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 $\Psi = 0.0001, \varphi^2 = 1000, a_1 = 1, b_1 = 0.0001, a_2 = 0$, and $b_2 = 100$.

To estimate the parameters in the models fitted with the above prior density function and likelihood equations (8) and (9), 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 and Rubin (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.

It is important to decide which model provides the best fit to the dataset, the 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 Π parameters in a model and n observations in a dataset. AIC, BIC and DIC are elucidated as

$$\begin{aligned} AIC &= -2\log\mathcal{L}(y|\hat{\theta}) + 2\Pi \\ BIC &= -2\log\mathcal{L}(y|\hat{\theta}) + \log(n)\Pi \\ DIC &= -2\log\mathcal{L}(y|\hat{\theta}) + 2\Pi_D \end{aligned}$$

where $\Pi_D = \mathcal{E}[2\log\mathcal{L}(y|\theta)] - 2\log\mathcal{L}(y|\hat{\theta})$.

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_r against Model M_v and defined as

$$\mathcal{BF}_{rv} = \frac{P(y|M_r)}{P(y|M_v)}$$

where

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

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

In spite of the fact that $2\log\mathcal{BF}_{rv}$ 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.

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

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

A value of $2\log\mathcal{BF}_{rv}$ more than 10 indicates an extremely strong positive evidence to favor model M_r over model M_v while a value between 0 and 2 is insufficient evidence to favor either model. A value between 2 and 6 or 6 and 10 indicates a mild or moderately strong evidence respectively, to favor the numerator model.

7. Simulation study

To evaluate the performance of the Bayesian estimation procedure we carried out a simulation study. For the simulation purpose we have considered only one covariate X_1 which assume to follow normal distribution. As the Bayesian strategies are time consuming, fifty sets of lifetimes were generated utilizing the inverse transform procedure. Both the chains were iterated for 100000 times. There is no effect of prior distribution on posterior summaries because estimates of parameters are nearly same and convergence rate of Gibbs sampler for both the prior sets is also not greatly different. Also for both the chains the results were somewhat similar, so we present here the analysis for only one chain with $\Gamma(a_1, b_1)$ as prior to baseline distribution for all the models. The Gelman-Rubin convergence statistic values are nearly equal to 1, the Geweke test values are quite small, and the corresponding p-values are

Table 1: Simulation under mixture shared inverse Gaussian frailty model

Actual Parameter values	Estimates	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6300; autocorrelation lag = 185							
$\alpha_1(0.0049)$	0.0049	0.0003	0.0044	0.0056	-0.0013	0.4994	1.0005
$\alpha_2(0.0051)$	0.0051	0.0004	0.0041	0.0059	0.0007	0.5003	1.0021
$\lambda_1(0.0028)$	0.0029	0.0005	0.0020	0.0039	-0.0101	0.4959	1.0000
$\lambda_2(0.0011)$	0.0011	3.4e-05	0.0010	0.0012	0.0012	0.5004	1.0004
$\gamma_1(0.9150)$	0.9123	0.0495	0.8426	1.0238	-0.0019	0.4992	0.9999
$\gamma_2(0.2400)$	0.2399	0.0557	0.1443	0.3354	-0.0004	0.4998	1.0005
$\xi(0.4350)$	0.4341	0.0541	0.3457	0.5315	-0.0015	0.4993	1.0018
$\beta(0.0298)$	0.0297	0.0054	0.0204	0.0393	-0.0025	0.4989	1.0000

Table 2: Simulation under shared inverse Gaussian frailty model

Actual Parameter values	Estimates	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6600; autocorrelation lag = 210							
$\alpha_1(0.0490)$	0.0494	0.0031	0.0432	0.0554	0.0044	0.5017	1.0004
$\alpha_2(0.0421)$	0.0424	0.0050	0.0319	0.0496	-0.0058	0.4976	1.0037
$\lambda_1(0.0070)$	0.0069	0.0005	0.0060	0.0079	0.0035	0.5014	1.0057
$\lambda_2(0.0011)$	0.0011	3.35e-05	0.0010	0.0012	-0.0007	0.4997	1.0000
$\gamma_1(0.9510)$	0.9438	0.0566	0.8539	1.0450	-0.0003	0.4998	1.0008
$\gamma_2(0.4410)$	0.4401	0.0563	0.3462	0.5331	-0.0125	0.4950	1.0007
$\xi(0.8220)$	0.8220	0.0540	0.7247	0.9137	0.0019	0.5007	0.9999
$\beta(0.0090)$	0.0089	0.0005	0.0080	0.0099	-0.0061	0.4975	1.0000

large enough to say that the chain attains stationary distribution. Tables 1 and 2 present the posterior summaries of mixture shared inverse Gaussian frailty and shared inverse Gaussian frailty with modified Weibull baseline distribution. It provides estimates (posterior means), standard error and upper and lower credible limits with 95% confidence intervals.

8. Application in real life data

To illustrate the Bayesian estimation procedure we use kidney infection data of (McGilchrist and Aisbett 1991). It consists of recurrent infection times from the use of catheters for 38 patients using portable dialysis machine. The two infection times per patient are grouped together in a cluster. Other relevant informations are censoring time of infection, the 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 is used to check the goodness of fit for kidney infection data and it is found that the p-values obtained for the first and second recurrences are large enough and hence there is no reason to reject the hypothesis that the first and second recurrence time to follow the distributions with survival function as given in equation (4) in univariate case and thus it is assumed to be fitted for bivariate case. The corresponding p-values are given in Table 3. 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 priors, so estimates are not dependent on the different prior distributions. The

Table 3: p-values of K-S Statistics for the goodness of fit test for Kidney Infection data set under Modified Weibull baseline distribution

Distribution	Recurrence times	
	first	second
Mixture shared inverse Gaussian frailty	0.8687	0.8430
Shared inverse Gaussian frailty	0.3573	0.4571

convergence rate of the Gibbs sampler for both the prior sets is almost the same. Also, both the chains show somewhat similar results, so we present here the analysis for only one chain with $\Gamma(a_1, b_1)$ as prior for the baseline parameters, for all the models. We also calculate

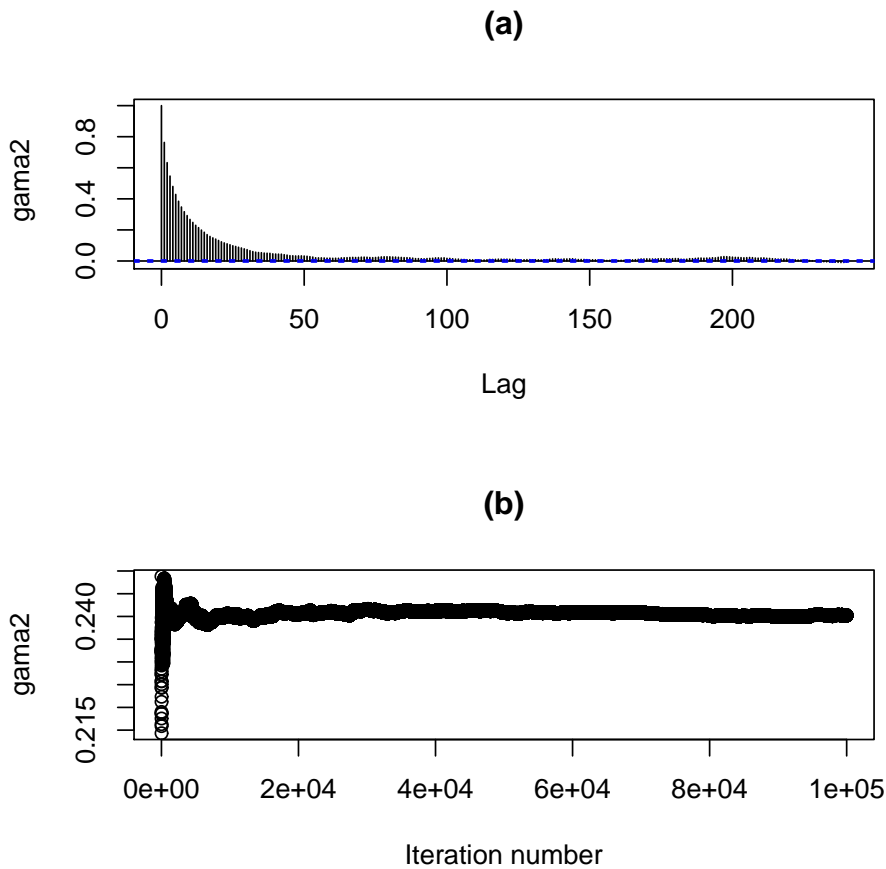


Figure 1: (a) ACF plot (b) Running mean plot of model-I.

Gelman Rubin statistic values and Geweke test statistics values to check the convergence of the Markov chain to a stationary distribution. Gelman Rubin statistics values are closed to one, Geweke test statistics values are closed to zero and the p-values are large enough to say that the chains attain stationary distribution. The largest values of GR statistics are 1.0072, 1.0038 and 1.0083, which indicating that 5900, 6100 and 5900 iterations would be satisfactory burn-in-period. Autocorrelation plot is used to decide autocorrelation lag given in figure 1(a) and the convergence of the chain also confirm by running mean plot given in figure1(b). The estimate of parameters (posterior mean), standard error, credible limits are given in Tables 4, 5 and 6. Since the credible intervals do not contain zero, all the factors are significant. The positive value of β_1 indicates that age is a significant factor for infection of the kidney, as age increases the chance of infection also increase. The negative value of β_2 shows that

Table 4: Posterior results under mixture shared inverse Gaussian frailty model

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 5900; autocorrelation lag = 240							
α_1	0.0050	0.0003	0.0043	0.0056	0.0065	0.5026	1.0000
α_2	0.0051	0.0005	0.0041	0.0059	-0.0167	0.4933	1.0068
λ_1	0.0029	0.0006	0.0020	0.0039	-0.0094	0.4962	0.9999
λ_2	0.0011	3.3e-05	0.0010	0.0012	0.0023	0.5009	1.0001
γ_1	0.9768	0.0563	0.8855	1.0730	0.0112	0.5044	1.0000
γ_2	0.2418	0.0556	0.1467	0.3366	-0.0028	0.4988	1.0013
ξ	0.4257	0.0521	0.3443	0.5331	0.0099	0.5039	1.0010
β_1	0.0300	0.0049	0.0210	0.0390	-0.0021	0.4991	1.0072
β_2	-4.1150	1.3291	-6.9925	-2.2882	0.0019	0.4991	1.0062
β_3	0.0029	0.0005	0.0020	0.0039	0.0083	0.5033	1.0002
β_4	0.5060	0.0547	0.4137	0.6023	0.0027	0.5010	0.9999
β_5	-6.8937	2.5580	-11.4292	-1.9047	0.0042	0.5016	1.0000

Table 5: Posterior results under shared inverse Gaussian frailty model

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 6100 ; autocorrelation lag = 190							
α_1	0.0480	0.0030	0.0414	0.0540	0.0083	0.5033	0.9999
α_2	0.0359	0.0051	0.0282	0.0471	0.0057	0.5023	1.0038
λ_1	0.0069	0.0005	0.0060	0.0079	-0.0011	0.4995	1.0001
λ_2	0.0009	0.0005	6.02e-05	0.0019	-0.0078	0.4968	1.0002
γ_1	0.6915	0.0560	0.5955	0.7822	-0.0025	0.4989	1.0021
γ_2	0.4265	0.0555	0.3341	0.5236	-0.0063	0.4974	1.0039
ξ	0.3910	0.0508	0.3065	0.4865	0.0058	0.5023	1.0002
β_1	0.0079	0.0005	0.0070	0.0089	0.0054	0.5021	1.0004
β_2	-2.0598	0.2317	-2.5027	-1.6181	-0.0006	0.5021	1.0000
β_3	0.0040	0.0005	0.0031	0.0049	-0.0131	0.4947	0.9999
β_4	0.0020	0.0005	0.0011	0.0029	-0.0036	0.4985	1.0000
β_5	-1.5227	0.4460	-2.3664	-0.6725	-0.0004	0.4998	0.9999

female has a lower chance of infection than male. Other significant factors, which have a higher chance of infection is disease types such as Glomerulo Nephritis(β_3), Acute Nephritis (β_4) and Polycystic Kidney Disease (β_5).

A 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 under modified Weibull 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 exceptionally little, so AIC, BIC, and DIC values are not recommendable to take a choice between the models. Presently we consider Bayes factor for comparing a pair of models r and v . The values in Table 8 shows that model-I is better than model-II and model-III as the corresponding value of $2\log(B_{rv})$ is greater than 2 and 6 indicating that there is positive and strong positive to favor model-I than model-II and model-III for the given dataset, which affirms our earlier result given in Table 7. Hence from all the demonstrated comparison criteria, we can say model-I that is mixture shared inverse Gaussian frailty model is better than model-II and model-III (shared inverse Gaussian frailty model and without frailty model) under modified Weibull distribution as a baseline for modeling kidney infection data. $2\log(B_{rv})$ value of model-II against model-III is 3.6679, which means that frailty models are better than without the frailty model under modified Weibull as baseline distribution.

Table 6: Posterior results under without frailty model

Parameter	Estimate	Standard Error	Lower Credible Limit	Upper Credible Limit	Geweke values	p values	Gelman & Rubin values
burn in period = 5900; autocorrelation lag = 240							
α_1	0.0050	0.0003	0.0043	0.0056	0.0044	0.5017	1.0007
α_2	0.0053	0.0004	0.0041	0.0059	0.0047	0.5018	1.0083
λ_1	0.0069	0.0005	0.0060	0.0079	-0.0052	0.4978	1.0208
λ_2	0.0031	0.0005	0.0022	0.0040	0.0012	0.5005	1.0061
γ_1	0.9938	0.0472	0.9232	1.0912	0.0020	0.5008	1.0074
γ_2	0.5042	0.0521	0.4123	0.5910	0.0046	0.5018	1.0005
β_1	0.0201	0.0046	0.0109	0.0285	0.0033	0.5013	1.0001
β_2	-1.4733	0.2793	-2.0170	-0.9352	-0.0041	0.5013	1.0007
β_3	0.0032	0.0005	0.0022	0.0041	-0.0031	0.4987	1.0003
β_4	0.6469	0.0510	0.5580	0.7423	-0.0096	0.4961	1.0010
β_5	-1.1094	0.4041	-1.8762	-0.2903	-0.0050	0.4979	1.0000

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

Model No.	AIC	BIC	DIC	log-likelihood
Model I	684.7071	704.3582	663.4949	-330.3536
Model II	688.2823	707.9334	668.3010	-332.1412
Model III	688.0174	706.0308	672.2273	-333.0087

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

Numerator model against denominator model	$2\log_e(B_{rv})$	Range	Evidence against model in denominator
M_I against M_{II}	5.0672	≥ 2 and < 6	Positive
M_I against M_{III}	8.7352	≥ 6 and < 10	Strong Positive
M_{II} against M_{III}	3.6679	≥ 2 and < 6	Positive

9. Discussion

In this paper, we discuss the results for a new mixture shared inverse Gaussian frailty model and existing shared inverse Gaussian frailty model under the same modified Weibull baseline distribution.

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

All the demonstrated comparison criteria exhibit that a mixture shared inverse Gaussian frailty model demonstrated with modified Weibull baseline is better for modeling of kidney infection data rather than the existing shared inverse Gaussian a frailty model under the same baseline distribution. The estimates of frailty parameters were high in all models which are 0.4257 and 0.3910 for mixture shared inverse Gaussian frailty model and shared inverse Gaussian frailty model respectively. This demonstrated that there is strong evidence of a high degree of heterogeneity among the patients in the population. Only few patients are anticipated to be exceptionally inclined to infection compared to others with the same covariate values. We further established 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 inverse

Gaussian frailty model to analyze kidney infection data.

References

- Aalen OO (1980). “A Model for Non-Parametric Regression Analysis of Counting Processes.” In W Klonecki, A Kozek, J Rosinski (eds.), *Lecture Notes on Statistics 2: Mathematical Statistics and Probability Theory*. Springer, New York.
- Clayton DG (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. doi:<https://doi.org/10.1093/biomet/65.1.141>.
- Gelman A, Rubin DB (1992). “A Single Series from the Gibbs Sampler Provides a False Sense of Security.” In JM Bernardo, JO Berger, AP Dawid, AFM Smith (eds.), *Bayesian Statistics 4*, pp. 625–632. Oxford University Press, Oxford.
- Geweke J (1992). “Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments.” In JM Bernardo, J Berger, AP Dawid, AFM Smith (eds.), *Bayesian Statistics 4*, pp. 169–193. Oxford University Press, Oxford.
- Hanagal DD, Pandey A (2016). “Shared Gamma Frailty Models Based on Additive Hazards.” *Journal of the Indian Society for Probability and Statistics*, **17**(2), 161–184. doi:<https://doi.org/10.1007/s41096-016-0011-7>.
- Hanagal DD, Pandey A (2017). “Shared Inverse Gaussian Frailty Models Based on Additive Hazards.” *Communications in Statistics. Theory and Methods*, **46**(22), 11143–11162. doi:<https://doi.org/10.1080/03610926.2016.1260740>.
- Ibrahim JG, Chen MH, Sinha D (2001). *Bayesian Survival Analysis*. Springer,Verlag.
- Kass RE, Raftery AE (1995). “Bayes Factors.” *Journal of the American Statistical Association*, **90**(430), 773–795. doi:<http://dx.doi.org/10.2307/2291091>.
- Keyfitz N, Littman G (1979). “Mortality in a Heterogeneous Population.” *Population Studies*, **33**, 333–342. doi:<https://doi.org/10.1080/00324728.1979.10410448>.
- Kheiri S, Kimber A, Meshkani MR (2007). “Bayesian Analysis of an Inverse Gaussian Correlated Frailty Model.” *Computational Statistics and Data Analysis*, **51**, 5317–5326. doi:<https://doi.org/10.1016/j.csda.2006.09.026>.
- Lin DY, Ying Z (1994). “Semiparametric Analysis of the Additive Risk Model.” *Biometrika*, **81**(1), 61–71. doi:<https://doi.org/10.1093/biomet/81.1.61>.
- McGilchrist CA, Aisbett CW (1991). “Regression with Frailty in Survival Analysis.” *Biometrics*, pp. 461–466. doi:<https://doi.org/10.2307/2532138>.
- Oakes D (1982). “A Model for Association in Bivariate Survival Data.” *Journal of Royal Statistical Society B*, **44**, 414–422. doi:<https://doi.org/10.1111/j.2517-6161.1982.tb01222.x>.
- 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:[DOI:https://doi.org/10.1023/A:1009605117713](https://doi.org/10.1023/A:1009605117713).
- Sarhan AM, Zain-Din M (2008). “New Generalized Weibull Distribution.” *Applied Science*, **In Press**.

Shepard D, Zeckhauser R (1977). *Interventionism Mixed Populations: Concepts and Applications*. Discussion Paper Series, Harvard University, JFK School of Government.

Vaupel JW, Manton KG, Stallard E (1979). "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography*, **16**, 439–454. doi:DOI:<https://doi.org/10.2307/2061224>.

Affiliation:

Lalpawimawha

Department of Statistics

Pachhunga University College, Mizoram University

College Veng, Aizawl, Mizoram-796001

E-mail: raltelalpawimawha08@gmail.com

URL: <https://www.linkedin.com/in/lalpawimawha-ralte-lalpawimawha-221026b7/>

Arvind Pandey

Department of Statistics

Central University of Rajasthan

Bandar Sindri, Ajmer, Rajasthan-305817

E-mail: arvindmzu@gmail.com