International Journal of Scientific Research in Mathematical and Statistical Sciences
Volume-5, Issue-6, pp.170-176, December (2018)
Research Paper
E-ISSN: 2348-4519

Correlated Gamma Frailty Model Based on Logistic Exponential Baseline Distribution

A.Pandey[1], Lalpawimawha[2*], S.Bhushan[3], P.K. Misra[4]

[1]Department of Statistics, Central University of Rajasthan, Ajmer, India
[2]Department of Statistics, PUC, Mizoram University, Aizawl, India
[3,4]Department of Mathematics and Statistics, Dr Shakuntala Misra National Rehabilitation University, Lucknow, India

[*]*Corresponding Author: raltelalpawimawha08@gmail.com, Tel.: +91-98623-07640*

Available online at: www.isroset.org

Received: 06/Dec/2018, Accepted: 21/Dec/2018, Online: 31/Dec/2018

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. Shared frailty models are used despite their limitations. To overcome their disadvantages correlated frailty models may be used. In this paper, we propose correlated gamma frailty model with logistic exponential distribution as baseline distribution to analyze real-life bivariate survival dataset of McGilchrist and Aisbett [9] related to kidney infection. The Bayesian approach of Markov Chain Monte Carlo was employed to estimate the parameters involved in the models and modles comparison was done by using Bayesian comparison techniques such as Akaike information criteria (AIC), Bayesian information criteria (BIC), Deviance information criteria (DIC) and Bayes factor. Simulation study also carried out to compare the true values of parameters and estimated values of the parameters. A better model suggested for the data.

Keywords—Bayesian model comparison, Correlated gamma frailty,Logistic exponential distribution, MCMC.

I. INTRODUCTION

A correlated frailty model is the expansion of shared frailty model. In a shared frailty model, it is expected that the individuals within a group share common frailty since they are related to each other. For illustration, the same family members share the same hereditary variables or living within the same area share common natural variables. However, it does have some few demerits. Firstly, it powers that within the cluster the unobserved variables are to be the same, which may not be applicable in some real-life experiment. Secondly, within a cluster the survival times are dependent, which is based on marginal distributions of survival times. However, confounding is observed between dependence parameter and population heterogeneity when covariates are shown in a proportional hazards model with gamma-distributed [1]. Which infers that from marginal distributions the joint distribution can be identified [4]. Thirdly, within the cluster, one-dimensional frailty can only bring out a positive association. But, negative association too observed between survival times within the cluster. For illustration, in heart transplantation study, generally, the longer the patient must wait for an available heart, the shorter the patient is likely to survive after the transplantation. Subsequently, a negative association appeared between the waiting time and survival time of the patient.

To avoid these limitations, correlated frailty models are being outlined for the examination of multivariate failure time data, in which related random factors are utilized to clarify the frailty impact for each cluster. Unlike shared frailty models, correlated frailty models give not only variance parameters of the frailties, extra parameter to account for the relationship between the survival times. The correlated frailty models permit us to add additional correlation parameters, which can address the questions about associations between event times. When the association between the event times is extraordinarily intrigued, the correlated frailty model is more appropriate, for illustration, genetic studies of event times in families. The conditional survival function for the correlated frailties U_1 and U_2 in the bivariate case (here without observed covariates) looks like

$$S(t_1,t_2\,|\,U_1,U_2)=S_1(t_1\,|\,U_1)S_2(t_2\,|\,U_2)=e^{-U_1H_{01}(t_1)}e^{-U_2H_{02}(t_2)} \quad (1)$$

The distribution of the random vector (U_1,U_2) necessarily to be specified and determines the association structure of the event times in the model.

In the shared frailty model same frailties are assumed, but for correlated frailty two individuals in a pair, frailties are not necessarily the same. We assume that the frailties are acting multiplicatively on the baseline hazard function (proportional hazards model) and that the observations in a pair are conditionally independent, given the frailties. Hence, the hazard of the individual j (j = 1, 2) has the form

$$h(t \mid X_j, U_j) = U_j h_{oj}(t) e^{\beta' X_j} \qquad (2)$$

where t denotes age or time, X_i is a vector of observed covariates, β is a vector of regression parameters describing the effect of the covariates X_j, $h_{0j}(t)$ are baseline hazard functions, and U_j are frailties. Bivariate correlated frailty models are characterized by the joint distribution of a two-dimensional vector of frailties (U_1, U_2). If the two frailties are independent, the resulting lifetimes are independent, and no clustering is present in the model. The shared frailty model can be obtained as a special case of the correlated frailty model when both the frailties are equal [14].

In order to derive a marginal likelihood function, the assumption of conditional independence of lifespans, given the frailty, is used. Let d_{jk} be a censoring indicator for individual j(j = 1,2) in pair k(k = 1,…, n). Indicator d_{jk} is 1 if the individual has experienced the event of interest, and 0 otherwise. According to (2), the conditional survival function of the j^{th} individual in the k^{th} pair is

$$S(t \mid X_{jk}, U_{jk}) = e^{-U_{jk} H_{oj}(t)} e^{\beta_k X_{jk}} \qquad (3)$$

with $H_{0j}(t)$ denoting the cumulative baseline hazard function. The contribution of individual j(j = 1, 2) in pair k(k= 1,…, n) to the conditional likelihood is given by

$$[U_{jk} h_{oj}(t) e^{\beta_k Xjk}]^{d_{jk}} e^{-U_{jk} H_{oj}(t_{jk})} e^{\beta_k X_{jk}} \qquad (4)$$

where t_{jk} stands for an observation time of individual j from pair k. Assuming the conditional independence of lifespans, given the frailty, and integrating out the frailty, we obtain the marginal likelihood function

$$\prod_{k=1}^{n} \iint_{R^2} [u_1 h_{o1}(t_{1k}) e^{\beta_k X_{1k}}]^{d_{1k}} e^{-u_{1k} H_{01}(t_{1k})} e^{\beta_k X_{1k}} \qquad (5)$$

$$[u_2 h_{o2}(t_{2k}) e^{\beta_k X_{2k}}]^{d_{2k}} e^{-u_{2k} H_{02}(t_{2k})} e^{\beta_k X_{2k}} f(u_{1k}, u_{2k}) du_{1k} du_{2k}$$

where f(.,.) is the probability density function of the corresponding frailty distribution. All these formulas can be easily extended to the multivariate case, but need a specification of the correlation structure between individuals in a cluster in terms of the multivariate density function, which complicates analysis. Divya et al. [13] also explained the algorithm to take decision about the data where there is no known right path for the specific problem.

The remaining sections are categorized as follows-illustration of correlated frailty model is in section 2. Baseline distribution and proposed models described in sections 3 and 4. Sections 5 and 6 for estimation strategies and simulation study. Sections 7 and 8 for application to real life data and discussion of the results.

II. CORRELATED GAMMA FRAILTY MODEL

The correlated gamma-frailty model [10,15] is developed for the analysis of multivariate failure time data, in which two associated random variables are used to characterize the frailty effect for each cluster. For example, one random variable is assigned to twin 1 and another to twin 2 so that they are no longer constrained to having common frailty as in the shared frailty model.

To be more specific, let p_0, p_1 be some real positive variables. Set v = p_0 + p_1 and let Y_0,Y_1,Y_2 be independent gamma-distributed random variables with $Y_0 \sim G(p_0, v)$, $Y_1 \sim G(p_1, v)$, $Y_2 \sim G(p_1, v)$. Consequently,

$U_1 = Y_0 + Y_1 \sim G(p_0 + p_1, v) \sim G(v,v)$

$U_2 = Y_0 + Y_2 \sim G(p_0 + p_1, v) \sim G(v,v)$

are the frailties of individual 1 and 2 in a pair. The bivariate survival function in term of cumulative hazard function of this model is given by

$$S(t_1,t_2) = (1+\sigma_1^2 H_{01}(t_1) + \sigma_2^2 H_{02}(t_2))^{\frac{-\rho}{\sigma_1\sigma_2}} \times (1+\sigma_1^2 H_{01}(t_1))^{\frac{-1+\frac{\sigma_1}{\sigma_2}\rho}{\sigma_1^2}} (1+\sigma_2^2 H_{02}(t_2))^{\frac{-1+\frac{\sigma_2}{\sigma_1}\rho}{\sigma_2^2}} \qquad (6)$$

which results in the following representation of the gamma correlated frailty model

$$S(t_1,t_2) = \frac{S_1(t_1)^{1-\frac{\sigma_1}{\sigma_2}\rho} S_2(t_2)^{1-\frac{\sigma_2}{\sigma_1}\rho}}{(S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{\frac{\rho}{\sigma_1\sigma_2}}} \qquad (7)$$

where S(t) denotes the marginal univariate survival function, assumed to be equal for both partners in a twin pair and $0 \le \rho \le \min\left\{\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right\}$ holds. Furthermore, it holds that ρ = corr(U_1,U_2) and σ^2 = V(U_j), (j = 1, 2). For simplicity, we drop the dependence of the survival functions from observed covariates.

The bivariate distribution in the presence of covariates, when the frailty variable is degenerate is given by

$$S(t_1,t_2) = e^{-(\eta_j \{H_{01}(t_1) + H_{02}(t_2)\})} \qquad (8)$$

where $\eta_k = e^{\beta_k X_{jk}}$. According to different assumptions on the baseline distributions we get different correlated gamma frailty models.

III. BASELINE DISTRIBUTION

Logistic Exponential distribution

Generally, in a parametric model it is assumed that baseline hazard $r_0(t)$ is a parametric function. Here, the logistic exponential distribution is considered as baseline distribution proposed by Lan and Leemis [8] Logistic exponential distribution is useful in the characterization of the lifetime data analysis and having the hazard function for time t as

$$r(t)=\frac{\alpha\lambda\left(e^{\alpha t}-1\right)^{\lambda-1}e^{\alpha t}}{1+\left(e^{\alpha t}-1\right)^{\lambda}}\ ;t>0,\lambda>,\alpha>0 \qquad (9)$$

The corresponding cumulative hazard function and survival functions are respectively,

$$R(t)=\ln\left(1+\left(e^{\alpha t}-1\right)^{\lambda}\right) \qquad (10)$$

$$S(t)=\frac{1}{1+\left(e^{\alpha t}-1\right)^{\lambda}} \qquad (11)$$

The failure rate is constant if α and λ values are one. If λ value is 1, then the failure rate is increasing for $\alpha > 1$ and decreasing for $\alpha < 1$. If α value is one, the failure rate increasing for $\lambda < 0$ and decreasing for $\lambda > 0$. The failure rate is also constant for $\lambda = 0$ and $\alpha = 1$.

IV. PROPOSED MODELS

The unconditional survival function is obtained by replacing the cumulative hazard functions of logistic exponential distribution in equations (6) and (7). Then,

$$S(t_1,t_2)=\left(1+\sigma_1^2\left(\left(\frac{t_1}{\gamma_1}\right)^{\alpha_1}+\log\left[1-\lambda_1+\lambda_1 e^{-\left(\frac{t_1}{\lambda_1}\right)^{\alpha_1}}\right]\right)+\sigma_2^2\left(\left(\frac{t_2}{\gamma_2}\right)^{\alpha_2}+\log\left[1-\lambda_2+\lambda_2 e^{-\left(\frac{t_2}{\lambda_2}\right)^{\alpha_2}}\right]\right)\right)^{\frac{-\rho}{\sigma_1\sigma_2}}$$

$$\times\left(1+\sigma_1^2\left(\left(\frac{t_1}{\gamma_1}\right)^{\alpha_1}+\log\left[1-\lambda_1+\lambda_1 e^{-\left(\frac{t_1}{\lambda_1}\right)^{\alpha_1}}\right]\right)\right)^{\frac{-1+\frac{\sigma_1}{\sigma_2}\rho}{\sigma_1^2}}\left(1+\sigma_2^2\left(\left(\frac{t_2}{\gamma_2}\right)^{\alpha_2}+\log\left[1-\lambda_2+\lambda_2 e^{-\left(\frac{t_2}{\lambda_2}\right)^{\alpha_2}}\right]\right)\right)^{\frac{-1+\frac{\sigma_2}{\sigma_1}\rho}{\sigma_2^2}}$$

(12)

$$S(t_1,t_2)=e^{-\eta_j\left(\left(\left(\frac{t_1}{\gamma_1}\right)^{\alpha_1}+\log\left[1-\lambda_1+\lambda_1 e^{-\left(\frac{t_1}{\lambda_1}\right)^{\alpha_1}}\right]\right)+\left(\left(\frac{t_2}{\gamma_2}\right)^{\alpha_2}+\log\left[1-\lambda_2+\lambda_2 e^{-\left(\frac{t_2}{\lambda_2}\right)^{\alpha_2}}\right]\right)\right)}$$

(13)

The equation (12) is correlated gamma frailty model based on logistic exponential distribution and equation (13) is without frailty model based on the same baseline distribution and called as model-I and model-II.

V. BAYESIAN ESTIMATION OF PARAMETERS AND MODEL COMPARISONS

Suppose there are n individuals under study, whose first and second observed failure times are represented by (t_{1k}, t_{2k}). Let d_{1k} and d_{2k} be the observed censoring times for the k^{th} individual (k= 1, 2,..., n) for first and second recurrence times respectively. We also assume the independence between censoring scheme and life times of individuals.

The contribution of bivariate life time random variable of the k^{th} individual in likelihood function is given by,

$$L_k(t_{1k},t_{2k})=\begin{cases}f_1(t_{1k},t_{2k}),\ ;t_{1k}<d_{1k},t_{2k}<d_{2k},\\ f_2(t_{1k},d_{2k}),\ ;t_{1k}<d_{1k},t_{2k}>d_{2k},\\ f_3(d_{1k},t_{2k}),\ ;t_{1k}>d_{1k},t_{2k}<d_{2k},\\ f_4(d_{1k},d_{2k}),\ ;t_{1k}>d_{1k},t_{2k}>d_{2k}.\end{cases}$$

And the likelihood function is,

$$L(\psi,\beta,\theta)=\prod_{k=1}^{n_1}f_1(t_{1k},t_{2k})\prod_{k=1}^{n_2}f_2(t_{1k},d_{2k})\prod_{k=1}^{n_3}f_3(d_{1k},t_{2k})\prod_{k=1}^{n_4}f_4(d_{1k},d_{2k}) \qquad (14)$$

where θ, $\underline{\psi}$ and $\underline{\beta}$ are respectively the frailty parameter, the vector of baseline parameters and the vector of regression coefficients. For without frailty model likelihood function is

$$L(\psi,\beta)=\prod_{k=1}^{n_1}f_1(t_{1k},t_{2k})\prod_{k=1}^{n_2}f_2(t_{1k},d_{2k})\prod_{k=1}^{n_3}f_3(d_{1k},t_{2k})\prod_{k=1}^{n_4}f_4(d_{1k},d_{2k}) \qquad (15)$$

Where n_1, n_2, n_3 and n_4 are random number of observations of failure times (t_1,t_2) observed to lie in the ranges $t_{1k} < d_{1k}$, $t_{2k} < d_{2k}$; $t_{1k} < d_{1k}$, $t_{2k} > d_{2k}$; $t_{1k} > d_{1k}$, $t_{2k} < d_{2k}$ and $t_{1k} > d_{1k}$, $t_{2k} > d_{2k}$ respectively. The contributions of k^{th} individual in the likelihood function as

$$f_1(t_{1k},t_{2k})=\frac{\partial^2 S(t_{1k},t_{2k})}{\partial t_{1k}\partial t_{1k}}$$

$$f_1(d_{1k},t_{2k})=-\frac{\partial S(d_{1k},t_{2k})}{\partial t_{2k}}$$

$$f_1(d_{1k},t_{2k})=-\frac{\partial S(d_{1k},t_{2k})}{\partial t_{2k}}$$

$$f_1(d_{1k},t_{2k})=S(d_{1k},d_{2k}) \qquad (16)$$

Maximum likelihood method plays an important role in computing the estimators of models. Unfortunately computing the maximum likelihood estimators (MLEs) involves solving a eight-dimensional optimization problem

for Model-I and five-dimensional optimization problem for Model-II. As the method of maximum likelihood fails to estimate the parameters due to the convergence problem in the iterative procedure, so we use the Bayesian approach. The traditional maximum likelihood approach to estimation is commonly used in survival analysis, but it can encounter difficulties with frailty models. Moreover, standard maximum likelihood-based inference methods may not be suitable for small sample sizes or situations in which there is heavy censoring (see [7]). 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.

Bayesian approach is now popularly used to estimate the parameters in the models because the computation of the Bayesian analysis becomes feasible due to advances in computing technology. Several authors have discussed Bayesian approach for the estimation of parameters of the frailty models. Some of them are, Ibrahim *et al.* [5] and references therein, Santos and Achcar [11]. Santos and Achcar [11] considered parametric models with Weibull and generalized gamma distribution as baseline distributions and gamma, log-normal as frailty distributions. Ibrahim *et al.* [5] and references therein considered Weibull model and piecewise exponential model with gamma frailty.

The joint posterior density function of parameters for given failure times is obtained as,

$\pi(\alpha_1, \lambda_1, \alpha_2, \lambda_2, \sigma_1,\sigma_2,\rho, \underline{\beta}) \propto L(\alpha_1, \lambda_1, \alpha_2, \lambda_2, \sigma_1,\sigma_2,\rho, \underline{\beta})\times$

$$g_1(\alpha_1)g_2(\gamma_1)g_3(\alpha_2)g_4(\lambda_2)g_5(\sigma_1)g_6(\sigma_2)g_7(\rho)\prod_{i=1}^{5} p_i(\underset{\rightarrow i}{\beta})$$

where $g_i(.)$ (i = 1, 2,…, 7) indicates the prior density function with known hyperparameters of corresponding arguments for baseline parameters and frailty variance; $p_i(.)$ is prior density function for regression coefficient β_i; β represents a vector of regression coefficients except β_j, j = 1, 2,…, q and likelihood function L(.) is given by equation (14) or (15). Here we assume that all the parameters are independently distributed.

To estimate the parameters of the models, we used Metropolis-Hastings algorithm and Gibbs sampler. We monitored the convergence of a Markov chain to a stationary distribution by Geweke test [3] and Gelman-Rubin Statistics [2]. Trace plots, coupling from the past plots and sample autocorrelation plots are used to check the behavior of the chain, to decide burn-in period and autocorrelation lag respectively. The algorithm consists in successively obtaining a sample from the conditional distribution of each of the parameter given all other parameters of the model. These distributions are known as full conditional distributions. In our case, full conditional distributions are not easy to integrate out. So full conditional distributions are obtained by considering that they are proportional to the joint distribution of the parameters of the model.

In order to compare the proposed models, we have used several Bayesian model selection criteria such as Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC) and Deviance Information Criteria (DIC). Also, we have used the Bayes factor B_{uv} for comparison of the model M_u against M_v. To compute the Bayes factor we have used MCMC approach given in Kass and Raftery [6].

VI. 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 $X= X_1$ which we assume to follow the normal distribution. The frailty variable U is assumed to have gamma distribution. Lifetimes (T_{1k},T_{2k}) for k^{th} pair are conditionally independent for given frailty $U_j = u_j$. We assume that T_{jk} (k = 1,2,…,n , j = 1, 2) follows the logistic exponential baseline distribution. As the Bayesian methods are time consuming, we generate only fifty pairs of lifetimes using inverse transform technique. A widely used prior for frailty parameter θ is the gamma distribution G(0.0001, 0.0001). In addition, we assume that the regression coefficients are normal with mean zero and large variance say 1000. Similar types of prior distributions are used in Ibrahim et al. [5], Sahu *et al.* [12] and Santos and Achcar [11]. So in our study, we also use the same non-informative prior for frailty parameter θ and regression coefficient β_1. Since we do not have any prior information about baseline parameters, λ_1, α_1, λ_2, and α_2 prior distributions are assumed to be flat. We consider two different non-informative prior distributions for baseline parameters, one is G(a_1, a_2) and another is U (b_1, b_2). All the hyper-parameters φ, ε^2, a_1, a_2, b_1 and b_2 are known. Here G(a, b) is gamma distribution with shape parameter a and scale parameter b and U (b_1, b_2) represents uniform distribution over the interval (b_1, b_2). We assume the value of the hyper-parameters as a_1 = 1, a_2 = 0.0001, b_1 = 0 and b_2 = 100.

We run two parallel chains for model one using two sets of prior distributions with the different starting points using Metropolis-Hastings algorithm and Gibbs sampler based on normal transition kernels. We iterate both the chains for 100000 times. There is no effect of a prior distribution on posterior summaries because estimates of parameters are nearly same. Also for both the chains the results were somewhat similar. Due to lack of space trace plot, coupling from the past plot, autocorrelation plot and running mean plot are not provided. Table 1 present estimates, credible intervals, Gelman-Rubin convergence statistic and Geweke test for all the parameters of the Model I based on simulation study for the parameters.

Table 1: Simulation study for model-I

Parameter (value)	Estimate	SE	LCL	UCL	Geweke values	P values	GR values
burn in period = 5900; autocorrelation lag = 290							
α_1 (0.017)	0.0171	0.0027	0.0119	0.0231	7.4e-05	0.5000	1.0001
α_2 (0.009)	0.0089	0.0005	0.0080	0.0099	-0.0065	0.4973	1.0024
λ_1 (1.085)	1.0845	0.0503	0.9849	1.1623	-0.0125	0.4950	1.0004
λ_2 (1.233)	1.2321	0.0536	1.1440	1.3329	-0.0022	0.4990	0.9999
σ_1 (0.265)	0.2672	0.0517	0.1707	0.3524	0.0105	0.5042	1.0002
σ_2(0.561)	0.5622	0.0528	0.4671	0.6494	0.0028	0.5011	1.0016
ρ(0.489)	0.4900	0.0530	0.3954	0.5808	-0.0053	0.4978	1.0033
β (0.002)	0.0019	0.0004	0.0011	0.0029	0.0120	0.5048	1.0032

From Table 1 it is observe that estimated values are close to real value, standard error is quite small. Gelman-Rubin convergence statistic values are nearly equal to one and also Geweke test values are quite small and corresponding p-values are large enough to say the chain attains stationary distribution.

VII. APPLICATION TO REAL LIFE DATA

We illustrate the two proposed models by applying to infectious disease data related to kidney infection kidney infection due insertion of catheter [9]. It consists of the first and recurrence time of infection at point of insertion of catheter by using a portable dialysis machine in 38 kidney patients and five risk variables age, sex (0=male and 1=female) and disease type GN, AN and PKD where GN, AN and PKD are short forms of Glomerulo Nephritis, Acute Nephritis and Polycystic Kidney Disease.

To check goodness of fit of kidney data set, we consider Kolmogrove-Smirnov (K-S) test for frailty distributions. Table 2 gives the p values of goodness of fit test for model I. 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 logistic exponential distribution. Figure 1 show the parametric plot vs non parametric plot.

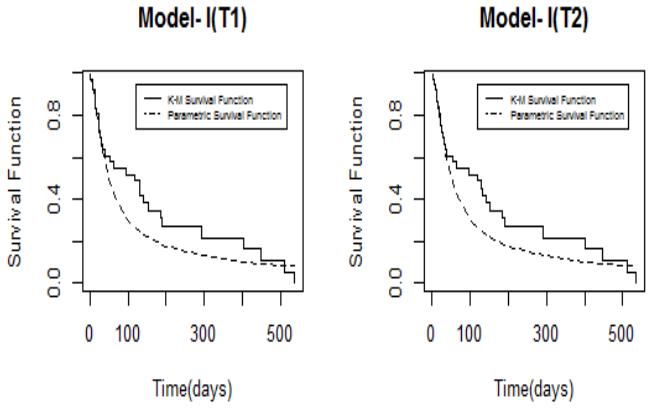


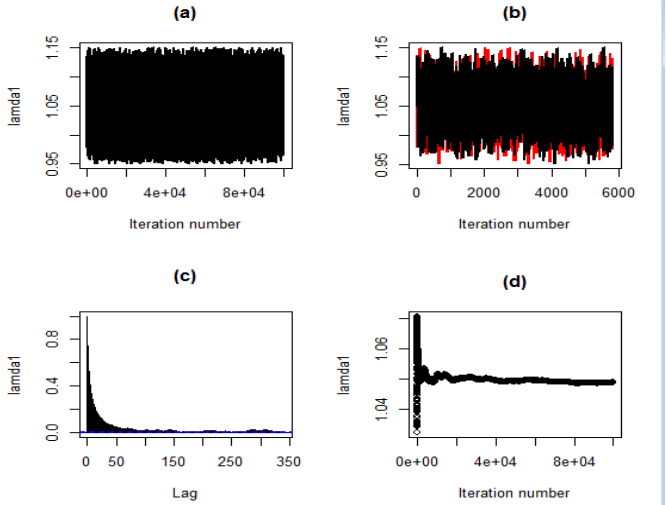
Figure 1: Survival function plots for (K-M survival and parametric survival).

Table 2: p-values of K-S Statistics for goodness of fit test for Kidney Infection data set recurrence time

Distribution	Recurrence time	
	first	second
Correlated Gamma	0.85751	0.96225

As in case of simulation, here also we assume same set of prior distributions. We run two parallel chains for both models using two sets of prior distributions with the different starting points using Metropolis-Hastings algorithm and Gibbs sampler based on normal transition kernels. We iterate both the chains for 100000 times. As seen in simulation study here also we got nearly same estimates of parameters for both the set of prior, so estimates are not dependent on the different prior distributions. Also both the chains shows somewhat similar results, so we present here the analysis for only one chain with $G(a_1,a_2)$ as prior for baseline parameters, for the model. Gelman-Rubin convergence statistic values are nearly equal to one and Geweke test statistic values are quite small and corresponding p-values are large enough to say the chains attain stationary distribution. The posterior mean and standard error with 95% credible intervals, Gelman-Rubin statistics values and Geweke test values with p values for Model I and II are presented in Table 3 and 4.

The trace plot, coupling from the past plot, autocorrelation plot and running mean plot are show in figures (2(a)-2(d)). The trace plot for all the parameters shows zigzag pattern which indicates that parameters move and mix more freely. Thus, it seems that the Markov chain has reached the stationary state. Burn in period is decided by using coupling from the past plot. However, a sequence of draws after burn-in period may have the autocorrelation. Because of the autocorrelation, consecutive draws may not be random, but values at widely separated time points are approximately independent. So, a pseudo random sample from the posterior distribution can be found by taking values from a single run of the Markov chain at widely spaced time points (autocorrelation lag) after burn-in period. The autocorrelation of the parameters become almost negligible after the certain lag. ACF plot after thinning show that observations are independent. We can also use running mean plots to check how well our chains are mixing. A running mean plot is a plot of the iterations against the mean of the draws up to each iteration. In fact running mean plots display a time series of the running mean for each parameter in each chain. These plots should be converging to a value. Running mean plot for each parameter is converging to the posterior mean of the parameter, thus, represents a good mixing of chain. Thus, our diagnostic plots suggest that the MCMC chains are mixing very well.



Figures 2: (a)Trace plot (b)Coupling from the past plot (c)ACF plot and (d)Running mean plot

Table 3: Posterior summary for Kidney Infection data set for Model I

Parameter	Estimate	SE	LCL	UCL	Geweke values	P values	GR values
burn in period = 5800; autocorrelation lag = 340							
α_1	0.0217	0.0027	0.0161	0.0270	0.0083	0.5033	1.0009
λ_1	1.0494	0.0576	0.9538	1.1430	-0.0083	0.4966	1.0008
α_2	0.0209	0.0042	0.0137	0.0290	0.0110	0.5043	1.0021
λ_2	1.2449	0.0524	1.1477	1.3360	0.0105	0.5042	1.0000
σ_1	0.2513	0.0528	0.1571	0.3427	-0.0047	0.4981	0.9999
σ_2	0.5635	0.0515	0.4629	0.6463	0.0020	0.5007	1.0004
ρ	0.4908	0.0495	0.3975	0.5795	-0.0046	0.4981	1.0064
β_1	0.0020	0.0005	0.0010	0.0029	-0.0007	0.4996	1.0079
β_2	-1.4838	0.2441	-1.918	-1.010	-0.0075	0.4996	1.0007
β_3	0.0039	0.0005	0.0030	0.0048	0.0091	0.5036	1.0013
β_4	0.4079	0.0519	0.3111	0.4946	0.0008	0.5003	1.0022
β_5	0.0019	0.0004	0.0011	0.0029	-0.0034	0.4986	1.0003

Table 4: Posterior summary for Kidney Infection data set for Model II

Parameter	Estimate	SE	LCL	UCL	Geweke values	P values	GR values
burn in period = 5600; autocorrelation lag = 345							
α_1	0.0222	0.0023	0.0176	0.0269	0.0060	0.5024	0.9999
λ_1	2.0798	0.0459	2.0218	2.1905	-0.0004	0.4998	1.0002
α_2	0.0144	0.0021	0.0106	0.0190	0.0015	0.5006	1.0000
λ_2	2.1153	0.0467	2.0440	2.2176	0.0021	0.5008	1.0016
β_1	0.0018	0.0004	0.0010	0.0027	-0.0102	0.4958	1.0000
β_2	-1.971	0.1937	-2.403	-1.652	-0.0004	0.4958	1.0001
β_3	0.0040	0.0004	0.0031	0.0049	0.0041	0.5016	1.0054
β_4	0.4587	0.0499	0.3653	0.5443	0.0027	0.5010	1.0000
β_5	0.0020	0.0004	0.0010	0.0029	-0.0009	0.4996	1.0030

AIC, BIC and DIC values for two models are given in Table 5. Bayes factor for Model I and Model II are given in Table 6.

Table 5: AIC, BIC and DIC values for all the models fitted to kidney infection data set.

Model No.	AIC	BIC	DIC	Log-likelihood
Model I	685.8200	705.4710	666.3892	-330.9100
Model II	728.8585	743.5968	715.7782	-355.4293

Table 6: Bayes factor values and decision for test of significance for frailty under Model I fitted to Kidney Infection Data Set

numerator model against denominator model	$2\log_e(B_{uv})$	range	Evidence against model in denominator
M_I against M_{II}	50.34167	≥ 10	Very Strong Positive

The comparison between two models is done using AIC, BIC and DIC values given in Table 5. The smallest AIC, BIC and DIC value is smaller in Model I than Model II, but the values are nearly equal. To take the decision about model I and model II, we use Bayes factor. The Bayesian test based on the Bayes factors for Model I against Model II is 50.34167 which support Model I for kidney infection data set compared to their corresponding model without frailty ($\sigma_1 = \sigma_2 = 0$, $\rho = 0$) and frailty is significant in Model I. Some patients are expected to be varying prone to infection compared to others with same covariate value. This is not surprising, as seen in the data set there is a male patient with infection time 8 and 16, and there is also male patient with infection time 152 and 562.

VIII. CONCLUSION

In this paper we discuss results for correlated Gamma frailty model with logistic exponential baseline distribution. Main aim of our study is to check which distribution (with correlated gamma frailty or without frailty) fits better. We perform simulation study and also analyze kidney infection data by using R. For maximum likelihood estimate, likelihood equations do not convergent and method of maximum likelihood fails to estimate the parameters so we used Bayesian approach. The entire estimation procedure using Bayesian approach took large amount of computational time but the time was more or less the same for the two models.

Different prior gives the same estimates of the parameters. Convergence rate of Gibbs sampling algorithm does not depend on these choices of prior distributions in our models for kidney infection data. The estimate of σ's (Model-I σ_1 = 0.2513 and σ_2=0.5635) from model show that there is a strong evidence of high degree of heterogeneity in the population of patients.

Bayes factor is used to test the frailty parameter $\sigma_1=\sigma_2 = 0$ and it observed that frailty is present and model with frailty fit better than without frailty model. The covariates sex, GN, AN and PKD are the covariates which are significant for all models. Negative value of regression coefficient (β_2) of covariate sex indicates that the female patients have a slightly lower the risk of infection.

The comparison between two models is done using AIC, BIC and DIC values. The smallest AIC value is Model I (correlated gamma frailty model with logistic exponential baseline distribution). The same result holds for BIC and DIC value. But these differences are not much significant. To take the decision about Model I and Model II, we use the Bayes factor. We observe that, the Model I is better than Model II. In this case we can conclude that correlated frailty model is better than without frailty model. Also we observe that gamma frailty is better than without frailty based on the same baseline distribution. By referring all the above analysis now we are in a position to say that, we have suggested a new correlated gamma frailty model with logistic exponential distribution as baseline distribution which is better than without frailty model for modeling of kidney infection data.

REFERENCES

[1].D.Clayton, J.Cuzick, " Multivariate Generalizations of the Proportional Hazards Model", Royal Statistical Society A, Vol.48, pp.82-117,1985.

[2] A.Gelman, D.B.Rubin," 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 Univ. Press, Oxford, pp.625-632, 1992.

[3] J.Geweke, "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 University Press, Oxford, pp. 169-193, 1992.

[4] P.Hougaard,"Survival models for heterogeneous populations derived from stable distributions", Biometrika, Vol.73, pp.387-396, 1986.

[5] J.G.Ibrahim , C.Ming-Hui, D.Sinha,"Bayesian Survival Analysis", Springer, Verlag, 2001.

[6] R.E.Kass, A.E.Raftery,"Bayes Factor". Journal of the American Statistical Association, Vol.90,Issue. 430,pp. 773-795,1995.

[7] S.Kheiri, A.Kimber, M.R.Meshkani,"Bayesian analysis of an inverse Gaussian correlated frailty model". Computational Statistics and Data Analysis, Vo.51, pp.5317-5326,2007.

[8] Y.Lan, L.M.Leemis,"The logisticexponential survival distribution",Naval Research Logistics (NRL), Vol.55, Issue.3, pp.252-264,2008.

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

[10] J.H.Petersen, "An additive frailty model for correlated lifetimes", Biometrics, Vol.54,pp. 646,1998.

[11] C.A.Santos, J.A.Achcar, "A Bayesian analysis for multivariate survival data in the presence of covariates",Journal of Statistical Theory and Applications, Vol.9, pp.233-253,2010.

[12] S.K.Sahu, D.K.Dey, H.Aslanidou, D.Sinha,"A Weibull regression model with gamma frailties for multivariate survival data",Life Time Data Analysis, Vol.3, pp.123-137,1997.

[13] K. Sree Divya, P.Bhargavi, S.Jyothi,"Machine Learning Algorithms in Big data Analytics". International Journal of Computer Sciences and Engineering, Vol 6, Issue 1, pp.63-70, 2018.

[14]A.Wienke,"Frailty Models in Survival Analysis",Chapman & Hall/CRC, 2011.

[15] A.I.Yashin, I.A.Iachine, "Environment determines 50% of variability in individual frailty: Results from Danish twin study, Research Report, Population Studies of Aging, 10, Odense University, Odense, 1994.

AUTHORS PROFILE

Dr Arvind Pandey is currently working as assistant Professor , Department of Statistics at Central University of Rajasthan. He was working as assistant professor and head in the department of statistics, Pachhunga University College since 2005. He has more than 20 publications

Mr.Lalpawimawha is currently working as assistant professor, Department of Statistics, Pachhunga University College, Mizoram University, Aizawl, Mizoram. He was working in Centre for applied Mathematics, Central University of Jharkhand during 2011-2012. He has more than 15 publications.

Dr Shahshi Bhushan is head and associate professor, Department of Mathematics and Statistics, Dr Shakuntala Misra National Rehabilitation University, Lucknow , Uttar Pradesh. He has more than 13 years teaching experience in UG and PG level. He has more than 20 publications.

Dr Praveen Kumar Misra is assistant professor, Department of Mathematics and Statistics, Dr Shakuntala Misra National Rehabilitation University, Lucknow , Uttar Pradesh. He has more than 10 years teaching experience and has many publications

