Comparison of Frailty Models for Acute Leukemia Data under Gompertz Baseline Distribution

David D. Hanagal and Richa Sharma

Department of Statistics, University of Pune, Pune-411007, India.
Email: david_hanagal@yahoo.co.in; richa_stat@rediffmail.com

Abstract

In this paper, we consider two different shared frailty regression models under the assumption of Gompertz as baseline distribution. Mostly assumption of gamma distribution is considered for frailty distribution. To compare the results with gamma frailty model, we consider the inverse Gaussian shared frailty model also. We compare these two models to a real life bivariate survival data set of acute leukemia remission times (Freireich et al., 1963). Analysis is performed using Markov Chain Monte Carlo methods. Model comparison is made using Bayesian model selection criterion and a well fitted model is suggested for the acute leukemia data.

Key words: Bayesian estimation, Gamma distribution, Gompertz distribution, Inverse Gaussian distribution, Markov Chain Monte Carlo (MCMC), Model selection criterion, Predictive density, Shared frailty.

1 Introduction

Bivariate survival data arise when each study subject experiences two events. Particular examples include failure times of paired human organs, (e.g. kidneys, eyes) and first and second occurrences of a given disease. Bivariate survival data may consist of time to diagnosis or hospitalization and the time to eventual death from a fatal disease. In industrial applications, bivariate data may consist of survival times for two very similar components in a system. For example, the breakdown times of dual generators in a power plant or failure times of twin engines in a 2-engine airplane are illustrations of bivariate survival data.

Models fitted to these survival data may involve parametric or non-parametric forms for the hazard function. This depends on whether this form is defined (up to a small number of unknown parameters) as that of a known distribution, or whether it is completely undefined. In this paper, we shall be concerned only with parametric forms which fit well to acute leukemia data.
In particular, the distribution of the response should have positive support. Examples of such distributions are the exponential, Weibull, log-normal, log-logistic, Gompertz, and the generalized gamma, among others. Here, we consider Gompertz distribution as baseline distribution. The Gompertz distribution is one of the most important growth models. It has many applications in, for example, medical, biological, and actuarial studies. This distribution was first introduced by Gompertz (1825). It has been used as a growth model to fit the tumor growth.

Modeling dependence in multivariate survival data has received considerable attention in recent literature. A key development in modeling such data is to consider the frailty models. Frailty models for multivariate survival data are derived under a conditional independence assumption by specifying latent variables that act multiplicatively on the baseline hazard, see e.g., Clayton (1978), Oakes (1982), and Clayton and Cuzick (1985). These models formulate the variability of life times, coming from two distinct sources. The first source is natural variability and it is explained by the hazard function and the second is variability common to individuals of the same group or variability common to several events of an individual and it is explained by the frailty.

There are two important approaches in this field, the shared frailty model and the correlated frailty model. In a shared frailty model, the frailty is common to the individuals in the group, and is thus responsible for creating dependence. In the following, we will restrict our considerations to the bivariate case. Extensions to higher-dimensional models are straightforward in the shared frailty approach. We are interested mainly on paired data with common shared frailty. The key assumption is that the dependence between two individual life time variables $T_1$ and $T_2$ is caused by the frailty representing unobserved common risk factors and conditional on frailty $T_1$ and $T_2$ are independent. Because the frailty is not observed, it is assumed to follow some distribution, typically a gamma distribution and an inverse Gaussian distribution which we consider in this paper.

In this paper, we consider shared frailty model that allowed for a random effect or “frailty” in the hazard model which is common and shared by all individuals in the group ( see, for example, Vaupel et al. 1979; Lancaster and Nickell, 1980 who have considered univariate model as frailty model). Some interesting situations motivate to study this particular model, like survival times in genetic epidemiology, dental implants of patients and twin births (both
monozygotic and dizygotic) where genetic behavior of patients is unobserved and random. Because the frailty is not observed, it is assumed to follow some distribution, typically gamma and inverse Gaussian distributions which we consider in this paper.

The role of shared frailty parameter has some important features. This shared frailty (or unobserved risk) is due to some genetic factors or environmental factors which are unobserved in an individual and are common and shared by both paired organs in human or twins in a family. This shared frailty parameter is also is responsible for the dependence between the two components. If the frailty is significant (i.e., the variance of the frailty is large) in the model and is ignored, then the model will be inappropriate and the decision based on such models will be misleading. If frailty is ignored, the estimates of regression coefficients will underestimate or overestimate (Lancaster and Nickell, 1980) and AIC, BIC, DIC will be more as compared to the model with frailty. The decision based on the model without frailty will go wrong. Ignoring frailty effect with finite mean may result in a negative bias in the estimated time dependence (Lancaster and Nickell, 1980). In general terms, we let the heterogeneity go into the error term. This will, of course, lead to an increase in the variability of the response as compared to the case, when the frailty is included.

The objective of this paper is to compare two shared frailty models in which frailty is generated by gamma and inverse Gaussian distribution. We have considered Gompertz as a distribution of the life times of two components where each life time follows it. The dependence in $T_1$ and $T_2$ is induced by frailty variable. After integrating out frailty, $T_1$ and $T_2$ have a bivariate distribution. The two well-known methods for estimation are maximum likelihood (here denoted CML, referring to classical maximum likelihood) and Bayesian estimation. There exist a number of philosophical differences between the two approaches. As implied by the name, CML is concerned with finding that set of parameter values $\theta$ that maximizes the probability of observing the data $D$ given those parameters, $p(D|\theta)$; in ML estimation, this is usually rephrased as maximizing the likelihood of the parameters given the data, $l(\theta|D)$. In contrast, Bayesian approach has received much attention as it is concerned not merely with finding a single value for each parameter, but rather with returning a posterior (i.e., after the data) probability distribution for that parameter. That is, Bayesian approach combines the information obtained from the likelihood function with the information in the prior distribution to obtain a posterior estimate for the required population
parameter. This concept of updating probabilities requires some probability for the parameters to exist prior to data observation in the Bayesian approach, a computational difference between the Bayesian and maximum likelihood approaches. Since Bayesian approach has the benefits of making the analyst able to use prior information and giving results that are easier to interpret, here, we use Bayesian computation with Markov chain Monte Carlo (MCMC) technique to estimate parameters involved in these two models. MCMC methods enable the calculation of several features of the posterior distributions and help formulation of model choice criterion based on posterior predictive loss (Gelfand and Ghosh, 1998).

Wienke (2010) considered Halle Lung Cancer (Halluca) study data and apply three different parametric shared gamma frailty models with exponential, Gompertz, and Weibull baseline hazards. The results showed that the exponential hazard function is not flexible enough, the Gompertz and Weibull model show a significant better fit to the data with respect to the likelihood ratio test. Duchateau and Janssen (2008) fit the inverse Gaussian frailty model with Weibull hazard to the udder quarter infection data. In this paper, we fit our shared gamma and inverse Gaussian frailty models with Gompertz hazard to bivariate survival data set of acute leukemia remission times and compare these two models using Bayesian comparison techniques such as AIC, AICc, BIC, DIC, Bayes’ factor, and predicted model choice criterion. Anderson et al. (1993) have analyzed the same acute leukemia data set by means of gamma frailty model. But their interest did not concern the frailty parameter rather concerned with the estimation of the treatment effect $\beta$ or $\exp(\beta)$ in the presence of a possible within-pair correlation. They found that the association within pairs seems to be very small and the treatment effect is the same (Anderson et al., 1993).

The remaining paper is organized as follows: In Section 2, we introduce the notion of shared frailty model with Laplace transformation followed by the Section 3 in which we give the introduction of gamma distribution and inverse Gaussian distribution with unconditional bivariate survival function evaluated at the cumulative baseline hazard. We present the Gompertz distribution as a baseline model with different shared frailty in Section 4. The joint survival function of proposed model after integrating out frailty is also derived in this Section. In Section 5, we discuss Bayesian estimation and the issues of model adequacy and model choice criterion using predictive distributions. Section 6 illustrates the methodology with bivariate real data set. Finally, the paper ends with a discussion of our findings in
2 General Shared Frailty Model

The shared gamma frailty model was suggested by Clayton (1978) for the analysis of the correlation between clustered survival times in genetic epidemiology. An advantage is that without covariates its mathematical properties are convenient for estimation (see Oakes, 1982, 1986). By measuring some potentially important covariates, we can examine the influence of the covariates, and we can examine whether they explain the dependence, that is, whether the frailty has no effect (or more correctly, no variation), when the covariate is included in the model. The regression model is derived conditionally on the shared frailty (U).

It is assumed that there is independence between the observations from different clusters. If the variation of the frailty variable is zero, this implies independence between event times in the clusters; otherwise, there is positive dependence between event times within a cluster. The shared frailty model abounds in the literature on frailty models and was extensively studied in the books by Hougaard (2000) and Therneau and Grambsch (2000). A more detailed presentation of shared frailty models can be found in the books by Duchateau and Janssen (2008), Wienke (2010), and Hanagal (2011).

Suppose $n$ individuals are observed for the study and let a bivariate random variable $(T_{i1}, T_{i2})$ be the first and second survival times of $i^{th}$ individual ($i = 1, 2, 3, \ldots, n$). Also suppose that there are $p$ observed covariates collected in a vector $X_i = (X_{i1}, \ldots, X_{ip})'$ for $i^{th}$ individual where $X_{ik}$ ($k = 1, 2, 3, \ldots, p$) represents the value of $k^{th}$ observed covariate for $i^{th}$ individual. Here we assume that the first and second survival time $T_1$ and $T_2$ for each cluster share the same value of the covariates. Assuming that the frailties are acting multiplicatively on the baseline hazard function and both the survival times of individuals $T_1$ and $T_2$ are conditionally independent for given frailty $U_i = u_i$. The conditional hazard model for $i^{th}$ cluster at $j^{th}$, ($j = 1, 2$) survival time $t_{ij} > 0$, for given frailty $U_i = u_i$ has the form:

$$h(t_{ij}|U_i, X_i) = u_i h_0(t_{ij}) \exp(x_i'\beta); \quad i = 1, 2, \ldots, n; \quad j = 1, 2$$

(2.1)

where $U_i$ is the unobserved (random) common risk factor shared by all subjects in cluster $i,$
$h_0(t_{ij})$ is the common baseline hazard function, $X_i$ is a vector of observable covariates and \( \beta \) is a vector of unknown regression coefficients.

Here $\exp(x_i'\beta)$ is the factor that gives the subject specific contribution to the hazard. Model (2.1) is called the shared frailty model because subjects in the same cluster share the same frailty factor. This model induces correlation between survival times of subjects within the same cluster. The value of the frailty $U_i$ is common to the individuals in the group, and thus it is responsible for creating dependence. This dependence is always positive.

Under the assumption of independence, the conditional survival function in the bivariate case for given frailty $U_i = u_i$ at time $t_{i1} > 0$ and $t_{i2} > 0$ is,

$$S(t_{i1}, t_{i2} | U_i, X_i) = S(t_{i1} | U_i, X_i)S(t_{i2} | U_i, X_i)$$

$$= \exp[-u_i\{H_{01}(t_{i1}) + H_{02}(t_{i2})\}] \exp(x_i'\beta)$$

(2.2)

where $H_{0j}(t_{ij})$ is the integrated baseline hazard of $T_{ij}$, ($i = 1, 2, ..., n; j = 1, 2$). From this, we immediately derive the bivariate survival function by integrating out $U_i$ having the probability function $f(u_i)$, for $i^{th}$ individual.

$$S(t_{i1}, t_{i2} | X_i) = \int_{U_i} S(t_{i1}, t_{i2} | u_i, x_i) f(u_i) du_i$$

$$= L_{U_i}\left[(H_{01}(t_{i1}) + H_{02}(t_{i2})) \exp(x_i'\beta)\right]$$

(2.3)

where $L_{U_i}(\cdot)$ is the Laplace transform of the distribution of $U_i$. Thus, the bivariate survivor function is easily expressed by means of the Laplace transform of the frailty distribution, evaluated at the total integrated conditional hazard.

### 3 Frailty Distributions

One important problem in the area of frailty models is the choice of the frailty distribution. The frailty distributions most often applied are the gamma distribution (Clayton, 1978; Vaupel et al., 1979; Oakes, 1982; Hougaard, 2000; Wienke et al., 2002; Wienke et al., 2003a; Hanagal and Sharma, 2012), the positive stable distribution (Hougaard, 1986b), the power variance function (PVF) distribution (Hougaard, 1986a), the inverse Gaussian distribution (Hougaard, 1984), the compound Poisson distribution (Aalen, 1988) and the log-normal distribution (McGilchrist and Aisbett, 1991).
Gamma Frailty:

First we consider frailty distribution as gamma distribution. The cross ratio function of gamma distribution (Clayton, 1978) is constant and is independent of the life times. The popularity of the model is due to the fact that the model functions are very easy to derive because of the simplicity of the derivatives of the Laplace transform.

To make the model identifiable, we restrict that expectation of the frailty equals 1 and variance be finite which implies that scale parameter = shape parameter, so that only one parameter needs to be estimated. Thus, the distribution of frailty U is the one parameter gamma distribution i.e. $U \sim \text{Gamma}(\theta^{-1}, \theta^{-1})$

Under the restriction, the corresponding density function and Laplace transformation of gamma distribution result in the following simplified form,

$$f(u) = \begin{cases} \frac{u^{\theta^{-1}-1} \exp(-u/\theta)}{\theta^{\theta^{-1}/\Gamma(1/\theta)}} & ; \ u > 0, \ \theta > 0 \\ 0 & ; \ otherwise. \end{cases} \quad (3.1)$$

and Laplace transform is

$$L_U(s) = (1 + s\theta)^{-1/\theta}. \quad (3.2)$$

Note that there is heterogeneity if $\theta > 0$. So the large values of $\theta$ reflect a greater degree of heterogeneity among groups and a stronger association within groups. Replacing Laplace transformation in equation (3.2) using (2.3), we get the unconditional bivariate survival function for $i^{th}$ individual at time $t_{i1} > 0$ and $t_{i2} > 0$ as,

$$S_\theta(t_{i1}, t_{i2}) = \left[1 + \theta \{H_{01}(t_{i1}) + H_{02}(t_{i2})\} \exp(x_i'\beta)\right]^{-1/\theta} \quad (3.3)$$

where $H_{01}(t_{i1})$ and $H_{02}(t_{i2})$ are cumulative baseline hazard functions of life time random variables $T_{i1}$ and $T_{i2}$, respectively.

Inverse Gaussian Frailty:

In shared frailty models gamma distribution is the most commonly used frailty distribution because of its mathematical convenience. However, it has drawbacks (see Kheiri et al. (2007)), for example, it may weaken the effect of covariates. As an alternative to the gamma distribution, the inverse Gaussian distribution was introduced by Hougaard (1984) and has been used by Manton et al. (1986), Klein et al. (1992), Keiding et al. (1997), Price and Manatunga (2001), Kheiri et al. (2007), and Duchateau and Janssen (2008).
Hougaard (1984) remarked that survival models with gamma and inverse Gaussian frailties behave very differently, noting that the relative frailty distribution among survivors is independent of age for the gamma, but becomes more homogeneous with time for the inverse Gaussian.

For identifiability, we assume $U$ has expected value equal to one. Under this restriction, density function and Laplace transformation of inverse Gaussian distribution result in the following simplified form,

$$f(u) = \begin{cases} 
\left[\frac{1}{2\pi\theta}\right]^{\frac{1}{2}} u^{-\frac{3}{2}} e^{-\frac{(u-1)^2}{2\theta}} &; u > 0, \theta > 0 \\
0 &; otherwise.
\end{cases} \quad(3.4)$$

and

$$L_U(s) = \exp\left[\frac{1 - (1 + 2\theta s)^{\frac{1}{2}}}{\theta}\right] \quad(3.5)$$

with variance of $U$ is $\theta$. Note that there is heterogeneity if $\theta > 0$. Thus, the unconditional bivariate survival function expressed as the Laplace transform of the frailty distribution, evaluated at the cumulative baseline hazard is,

$$S(t_{i1}, t_{i2}) = \exp\left\{1 - [1 + 2\theta\{H_{01}(t_{i1}) + H_{02}(t_{i2})\} \exp(x'_i\beta)]^{\frac{1}{2}}\right\} \quad(3.6)$$

where $H_{01}(t_{i1})$ and $H_{02}(t_{i2})$ are cumulative baseline hazard functions of life time random variables $T_{i1}$ and $T_{i2}$, respectively.

4 Gompertz Baseline Distribution

In parametric proportional hazards model we assume a particular parametric function for the baseline hazard $h_0(t)$. One of the parametric choice for $h_0(t)$ is the hazard rate of Gompertz distribution. A vast literature on human mortality suggests the use of the Gompertz baseline hazard rate to describe the mortality and also to model the risk of disease.

The cumulative hazard function and survival function of Gompertz distribution are

$$H_0(t) = \lambda \gamma^{-1}(\exp(\gamma t) - 1) \quad(4.1)$$

with $\lambda > 0, \gamma \in \mathbb{R}$.

$$S_0(t) = \exp[-\lambda \gamma^{-1}(\exp(\gamma t) - 1)] \quad(4.2)$$
We note that for $\gamma > 0$, $S_0(t) \to 0$ for $t \to \infty$. With $\gamma < 0$, $S_0(t) \to 0 < \exp(\lambda \gamma^{-1})$ for $t \to \infty$. Therefore the event never occurs for a proportion $\exp(\lambda \gamma^{-1})$ of the population. We therefore consider the case $\gamma > 0$.

In this paper, the two-parameter Gompertz distribution is considered. Let us assume that the independent random variables $T_1$ and $T_2$ have Gompertz distribution with parameters $\lambda_1, \gamma_1$ and $\lambda_2, \gamma_2$, respectively. The following unconditional joint survival function corresponding to (3.3) and (3.6), respectively,

$$S_\theta(t_{i1}, t_{i2} | X_i) = \left[1 + \theta \exp(x_i^T \beta) \{\lambda_1 \gamma_1^{-1}(\exp(\gamma_1 t_{i1}) - 1) + \lambda_2 \gamma_2^{-1}(\exp(\gamma_2 t_{i2}) - 1)\}\right]^{-1/\theta}$$ (4.3)

and

$$S_\theta(t_{i1}, t_{i2} | X_i) = \exp \left[\frac{1 - (1 + 2\theta \exp(x_i^T \beta) \{\lambda_1 \gamma_1^{-1}(\exp(\gamma_1 t_{i1}) - 1) + \lambda_2 \gamma_2^{-1}(\exp(\gamma_2 t_{i2}) - 1)\})^{1/2}}{\theta}\right]$$ (4.4)

Here onwards we call equations (4.3) and (4.4) as Model-I and Model-II, respectively.

Once we have unconditional survival function of bivariate random variable $(T_{i1}, T_{i2})$ we can obtain likelihood function and estimate the parameters of the model. Here onwards we represent $S_\theta(t_{i1}, t_{i2} | X_i)$ as $S_\theta(t_{i1}, t_{i2})$.

5 Bayesian Estimation and Model Comparisons

In the Bayesian framework, the parameters of the model are viewed as random variables with some distribution known as prior distribution. To apply MCMC methods, we assume that, conditional on explanatory variables and on the entire set of parameters, observations are independent and prior distributions for all parameters are mutually independent. Given the distribution and the priors, all full conditional distributions of the parameters can be calculated. These full conditional distributions are used in a Gibbs sampling procedure. The distribution of a parameter can be updated by combining its prior distribution and the likelihood function, called as posterior density of a parameter.

Model Comparison:

Model comparison and selection are among the most common problems of statistical practice, with numerous procedures for choosing among a set of models proposed in the literature (see, for example, Kadane and Lazar (2001) and Rao and Wu (2001) for recent
reviews). Most (but not all) selection methods are defined in terms of an appropriate information criterion, a mechanism that uses data to give each candidate model a certain score; this then leads to a fully ranked list of candidate models, from the best to the worst. Here, we use these criterion to determine the best model in a class of frailty models. Bayesian model comparison is commonly performed by computing posterior model probabilities. It is the most common method of Bayesian model assessment. In order to compare proposed models we use Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Deviance Information Criterion (DIC).

Raftery (1994), following Jeffreys (1961), proposes the rules of thumb for interpreting twice the logarithm of the Bayes factor. For two models of substantive interest, $M_j$ and $M_k$, twice the log of the Bayes factor is approximately equal to the difference in their BIC approximations.

Bayesian model examination for adequacy and model comparison can be proceed by predictive distribution. The posterior predictive density, $\pi(t_{rep}|t_{obs})$ is the predictive density of a new independent set of observable, $t_{rep}$, under the model, given the actual set of observable, $t_{obs}$. By marginalizing $\pi(t_{rep}|t_{obs})$ we obtain the posterior predictive density of one observation $t_{r,rep}$; $r = 1, \ldots, n$, as follows,

$$\pi(t_{r,rep}|t_{obs}) = \int \pi(t_{r,rep}|\theta)\pi(\theta|t_{obs})d\theta$$ (5.1)

Let $\mu_r$ and $\sigma_r^2$ denote the posterior predictive mean and variance of $t_{r,rep}$ under the density (6.7). We can easily estimate $\mu_r$ and $\sigma_r^2$ by Monte Carlo integration as follows. Suppose that $\theta^{(1)}, \ldots, \theta^{(B)}$ denote $B$ Gibbs-sampled values from $\pi(\theta|t_{obs})$. Then, a random sample $t_{r}^{(j)}$ drawn from $\pi(t_{r}|\theta^{(j)})$, is a sample from the above predictive density. See for example Gelfand (1996) for more details.

Recently Gelfand and Ghosh (1998) have proposed a model choice criterion by studying utility functions. They consider loss functions which reward an action for its closeness to the predictive value and penalizes the action if it is too far from the observed value. The criterion is then obtained by minimizing this posterior predictive loss. As they claim, the criterion emerges approximately as a form partitioned into a goodness-of-fit term and a penalty term for a wide range of models. For censored data the criterion must be modified. The modified
criterion (Gelfand and Ghosh, 1998) with squared error loss is:

\[ D_\omega = \sum_{r=1}^{n} \sigma_r^2 + \frac{\omega}{\omega + 1} \sum_{r=1}^{n} (\mu_r - \nu_r)^2 \]  

(5.2)

where \( \nu_r = t_{r,obs} \) if the \( r^{th} \) observation is a failure time and \( \nu_r = \max(\mu_r, w_r) \) if the \( r^{th} \) observation is censored at \( w_r; \omega > 0 \) is some constant. The first term is a penalty term which penalizes both under-fitted and over-fitted models, since the predictive variances in such cases will tend to be larger. The second term without the factor involving \( \omega \) is a goodness-of-fit measure. Model selection using \( D_\omega \) is usually not sensitive to \( \omega \). A model with minimum value of \( D \) is selected as the best model among all the models considered.

6 Analysis of Acute Leukemia Data

We illustrate the proposed model with one well-known example. The proposed methods are applied to the data set of acute leukemia remission times of patients, given by Freireich et al. (1963). The data are reported in Hougaard (2000). We examine the effect of a clinical trial of a drug 6-mercaptopurine (6-MP) versus a placebo in 42 children with acute leukemia.

Here we demonstrate the method using the well-known leukemia data, consisted of 21 pairs matched of leukemia patients, analyzed by Cox (1972), Hougaard, (2000) (subsection 1.5.4), Ibrahim et al., (2001) (example 3.4), Spiegelhalter et al., (2004), among others. The leukemia remission times data was first given by Freireich et al. (1963).

The random variable of interest consists of remission times (in weeks) of the patients assigned to treatment with a 6-MP drug or a placebo during remission maintenance therapy. After having been judged to be in a state of partial or complete remission for the primary treatment with prednisone, a patient was paired with a second patient in the same state. One randomly chosen patient in each pair received the maintenance treatment 6-MP and the other a placebo. It was assumed that deaths at a given time always preceded censoring at the same time, and other ties were broken by randomization. Success (failure) was defined to occur in the \( i^{th} \) pair if the time from remission to relapse or censoring for the patient on 6-MP (placebo) exceeded the time to relapse for the patient on placebo (6-MP). The trial was stopped once the number of successes or failures had reached significance. Out of 21 patients in treatment group, 9 failed during the study period and 12 were censored.
In contrast, none of the data are censored in placebo group; that is, all 21 patients in the placebo group went out of remission during the study period. The data set contains a single covariate $x$ with value 0 or 1 indicating remission status (0=partial, 1=complete).

In the analysis, we have used the R program. Given the model assumptions, this program performs the Gibbs sampler by simulating from the full conditional distributions. The Bayesian estimators were obtained through the implementation of the Metropolis-Hastings algorithm within Gibbs sampling scheme based on normal transition kernels. We implemented 95,000 iterations of the algorithm and described the first 3,500 and 3,000 iterations as a burn-in for Model-I and Model-II, respectively. To generate the Gibbs posterior samples, we choose to use two parallel chains. The chains should start from over-dispersed initial values to ensure good converge of parameter space.

We have taken the independent prior as $\lambda_i \sim \Gamma(10^{-4}, 10^{-4})$, $\gamma_i \sim \Gamma(1, 10^{-4})$, $\beta \sim N(0, 10^5)$ and $\theta \sim Gamma(10^{-4}, 10^{-4})$ for both models. Here, results for chain I and chain II are similar so we present result for only one chain (i.e. chain I). Trace plots for all the parameters show zigzag pattern which indicate that parameters move and mix more freely. Thus, it seems that the Markov chain has reached the stationary state. Due to lack of space we are not presenting graphs for time series (or trace) plot and sample autocorrelation plot with thinning for the parameters.

Tables 4 and 5 present posterior summary along with Gelman-Rubin convergence statistic values and Geweke test statistic values with corresponding p-values for Model I and II, respectively. Monitoring convergence of the chains has been done via the Brooks and Gelman (1998) convergence-diagnostic. Hence, once convergence has been achieved, 91,500 and 92,000 observations, for Model-I and Model-II, respectively, are taken from each chain after the burn-in period. On inspection of the Brooks and Gelmans diagnostic, we find the BGR (Brooks and Gelman Ratio) convergent to one, this shows that the convergence for the coefficient of regression $\beta$, the variance of frailty $\theta$ and other parameters has obtained. Also, the Geweke test statistic values are quite small and corresponding p-values are large enough to say the chains attain stationary distribution.

The autocorrelation of parameters become almost negligible after the defined lag, given in Tables 4 and 5 for both models. There is not much difference in the posterior estimates of baseline parameters, presented in Tables 4 and 5 for both the models. On the average $\lambda_1$
is greater than $\lambda_2$ for both Model-I and Model-II. This is attributed to the effectiveness of 6-MP drug. Thus, it is concluded that 6-MP patients have significant less risk of leukemia relapse than placebo group patients. The posterior estimate of $\theta = 0.3536$ for Model-I and $\theta = 0.4034$ for Model-II, shows that there exists significant heterogeneity in population of patients even though each patient share the same value of the covariate. The posterior mean of $\beta$ is $0.3417$ with $95\%$ credible interval $(-0.5689, 1.3067)$ for Model-I and $0.1091$ with $95\%$ credible interval $(-0.7368, 1.0439)$ for Model-II. Thus, patients had no significant effect due to a partial or complete remission of their leukemia for both models. In other words, we can say that there is no difference in the risk of acute leukemia relapse for both partial or complete remission patients.

To compare both models we firstly use AIC, AICc, BIC, DIC and log-likelihood values which are given in Table 1. As it is clear from the Table, the difference between AIC, AICc, BIC, DIC and log-likelihood values for Model-I and Model-II is very small, so AIC, AICc, BIC, DIC and log-likelihood values are not worthy to take decision between the Model-I and Model-II. To take decision about better model between Model-I and Model-II, we use Bayes factor. From the Table 2, which represents Bayes factor for models we can observe that, between Model-I and Model-II, there is positive evidence against Model-I, so Model-II is better than Model-I for modeling acute leukemia data.

Finally, the values relating to the model choice criterion $D_\omega$ (Equation 6.8) for both the models are shown in Table 3. First and second column of the Table 3 gives penalty and goodness-of-fit term, next four columns gives $D_\omega$ values for different values of $\omega = 1, 5, 10$ and $\infty$. Penalty term and goodness-of-fit term are minimum for Model-II, also for all the values of $\omega$, $D_\omega$ is minimum for Model-II, so this criterion also suggest Model-II is better model. Thus, Model-II that is inverse Gaussian frailty regression model with Gompertz as baseline distribution is better model than gamma frailty regression model with same baseline for modeling acute leukemia data.

7 Conclusions

The present study focuses on parametric models, which implies parametric specification of the baseline hazard and the distribution of the frailty. This paper presents some results for
the shared frailty models. In this paper, we have considered two failure times by allowing for potential dependence in the random quantities corresponding to each failure time which is induced by frailty. In the literature, gamma distribution is mostly used as frailty distribution because of its simplicity. So, our aim of this study is to suggest a better shared frailty model to analyze acute leukemia data among two frailty models with Gompertz as baseline distribution. As discussed in the Section 3.2, Hougaard (1984) introduced inverse Gaussian distribution as a better frailty distribution. So, we selected inverse Gaussian distribution as the other frailty distribution.

We have discussed the Bayesian estimation procedure including Gibbs sampling for computing the estimation of the unknown parameters for real data example of Acute Leukemia Study. We have run two parallel chains from different starting points and considered the “burn-in” interval for each chain. We have provided 95,000 iterations to perform the analysis. We have clearly written the steps involved in the iteration procedure. The quality of convergence was checked by Gelman-Rubin statistics (see Brooks and Gelman, 1998). The values of the Gelman-Rubin statistics in this case are quite close to one. Thus, the sample can be considered to have arisen from stationary distribution. The type of remission status is not significant in reducing the risk of acute leukemia relapse for the patients. Also, there is a heterogeneity among the patients which is 0.3536 for gamma frailty and 0.4034 for inverse Gaussian frailty models. The real data analysis indicate that the performance of Bayesian estimation method is quite satisfactory.

We have compared the models using Bayesian model comparison criterion such as, log-likelihood, AIC, AICc, BIC, DIC, Bayes factor and prediction model choice criterion. From the value of all these criterion we conclude that the inverse Gaussian distribution provide a suitable choice for the life time model as compared to gamma distribution. Thus, through our study we have suggested inverse Gaussian frailty model as a better model to fit acute leukemia data than gamma frailty model with Gompertz as baseline distribution.

Acknowledgments

We thank the referee for the valuable suggestions and comments which improved the earlier version of the manuscript.
References


**Appendix: Tables**

**Table 1: AIC, BIC and DIC values for Models I and II fitted to Acute Leukemia Data Set.**

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>AICc</th>
<th>BIC</th>
<th>DIC</th>
<th>log – likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>ModelI</td>
<td>228.86</td>
<td>234.86</td>
<td>235.12</td>
<td>224.83</td>
<td>-108.429</td>
</tr>
<tr>
<td>ModelII</td>
<td>228.30</td>
<td>234.30</td>
<td>234.57</td>
<td>223.91</td>
<td>-108.152</td>
</tr>
</tbody>
</table>
Table 2: Bayes factor values and decision for Models fitted to Acute Leukemia Data Set.

<table>
<thead>
<tr>
<th>numerator model against denominator model</th>
<th>2log_e(Buv)</th>
<th>range</th>
<th>Evidence against model in denominator</th>
</tr>
</thead>
<tbody>
<tr>
<td>II against I</td>
<td>0.4978757</td>
<td>0 to 2</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Table 3: Model selection criterion (5.2) for Models I and II fitted to Acute Leukemia Data Set. $P$ is the penalty (first) term in (5.2), $G$ is the goodness-of-fit (second) term in (5.2). Model II is favored by the criterion.

<table>
<thead>
<tr>
<th></th>
<th>$P$</th>
<th>$G$</th>
<th>$D_1$</th>
<th>$D_5$</th>
<th>$D_{10}$</th>
<th>$D_{\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ModelI</td>
<td>30503.49</td>
<td>6267.512</td>
<td>33637.25</td>
<td>35726.42</td>
<td>36201.23</td>
<td>36771.0</td>
</tr>
<tr>
<td>ModelII</td>
<td>15410.13</td>
<td>4764.426</td>
<td>17792.34</td>
<td>19380.48</td>
<td>19741.43</td>
<td>20174.56</td>
</tr>
</tbody>
</table>

Table 4: Posterior Summary for Acute Leukemia Data Set for Model-I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>BGR stat. value</th>
<th>Geweke test value</th>
<th>ACF</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>test value</td>
<td>p-value</td>
<td></td>
<td></td>
<td>Lower limit</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>1.0012</td>
<td>0.0189</td>
<td>0.5076</td>
<td>300</td>
<td>0.0393</td>
<td>0.0154</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>1.0516</td>
<td>-0.0114</td>
<td>0.4954</td>
<td>100</td>
<td>0.0138</td>
<td>0.0084</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.0037</td>
<td>-0.0287</td>
<td>0.4885</td>
<td>250</td>
<td>0.1214</td>
<td>0.0447</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>1.0087</td>
<td>0.0005</td>
<td>0.5002</td>
<td>200</td>
<td>0.0383</td>
<td>0.0262</td>
</tr>
<tr>
<td>$\theta$</td>
<td>1.0013</td>
<td>-0.0223</td>
<td>0.4911</td>
<td>250</td>
<td>0.3536</td>
<td>0.2793</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.0513</td>
<td>0.0113</td>
<td>0.5045</td>
<td>150</td>
<td>0.3417</td>
<td>0.4772</td>
</tr>
</tbody>
</table>
Table 5: Posterior Summary for Acute Leukemia Data Set for Model-II.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>BGR stat. value</th>
<th>Geweke test stat. value</th>
<th>ACF</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Credible Interval Lower limit</th>
<th>Credible Interval Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>1.0039</td>
<td>0.0015</td>
<td>0.5006</td>
<td>100</td>
<td>0.0683</td>
<td>0.0282</td>
<td>0.0219</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>1.0045</td>
<td>-0.0095</td>
<td>0.4962</td>
<td>60</td>
<td>0.0175</td>
<td>0.0111</td>
<td>0.0042</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.0023</td>
<td>-0.0145</td>
<td>0.4942</td>
<td>80</td>
<td>0.0835</td>
<td>0.0301</td>
<td>0.0277</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.9999</td>
<td>-0.00073</td>
<td>0.4997</td>
<td>50</td>
<td>0.0379</td>
<td>0.0242</td>
<td>0.0022</td>
</tr>
<tr>
<td>$\theta$</td>
<td>1.00009</td>
<td>-0.0065</td>
<td>0.4974</td>
<td>130</td>
<td>0.4034</td>
<td>0.3235</td>
<td>0.0534</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.0026</td>
<td>0.0099</td>
<td>0.5039</td>
<td>120</td>
<td>0.1091</td>
<td>0.4657</td>
<td>-0.7368</td>
</tr>
</tbody>
</table>