Bayesian Analysis for Partially Complete Time and Type of Failure Data

Debasis Kundu *

Abstract

In this paper we consider the Bayesian analysis of competing risks data, when the data are partially complete both in time and type of failures. It is assumed that the latent cause of failures have independent Weibull distributions with the common shape parameter but different scale parameters. When the shape parameter is known, it is assumed that the scale parameters have Beta-Gamma priors. In this case the Bayes estimates and the associated credible intervals can be obtained in explicit forms. When the shape parameter is also unknown, it is assumed that it has a very flexible log-concave prior density functions. When the common shape parameter is unknown, the Bayes estimates of the unknown parameters and the associated credible intervals cannot be obtained in explicit forms. We propose to use Markov Chain Monte Carlo sampling technique to compute Bayes estimates and also to compute associated credible intervals. We further consider the case when the covariates are also present. The analysis of two competing risks data sets, one with covariate and the other without covariates, have been performed for illustrative purposes. It is observed that the proposed model is very flexible, and the method is very easy to implement in practice.

Key Words and Phrases: Latent failure time model; Competing risks model; Importance sampling; Weibull distribution; Prior distribution; Credible interval.

1 Introduction

In medical studies or in reliability analysis an investigator is often interested in the assessment of a specific risk factor in presence of other risks. It is well known as the competing risks

*Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. e-mail: kundu@iitk.ac.in. The author would like to thank the referees for their constructive comments.

problem in the statistical literature. Usually the competing risks data consists of failure time and an indicator denoting the cause of failure. But in practice, often data may be incomplete both in failure time and in failure type. In many situations, the failure type can be determined even when the failure time is censored. For example, consider an analysis of pro-static cancer in older man. Since pro-static cancer is not lethal, and can be diagnosed long before death, a man known to have the pro-static cancer, and who is alive at the time of analysis provides complete information about the failure type, but incomplete information about failure time. Similarly, the time of death provides a complete information on the failure time, but no information on the failure type, without an autopsy, see for example Dinse [7].

In presence of complete information both on failure time and failure type, several studies have been carried out both under parametric and non-parametric set up, see for example the monograph of Crowder [5]. Note that, the analysis of competing risks data can be performed usually by two different methods, (a) latent failure times modeling as suggested by Cox [4]; (b) cause specific hazard function modeling as suggested by Prentice *et al.* [12]. In this paper, we mainly consider the latent failure times model formulation, although similar analysis can be performed for cause specific hazard function model formulation also.

In this paper, we make similar assumptions as of Dinse [7]. It is assumed that every member of a certain target population dies of a particular disease, say, cancer or by other causes. A proportion π of the population die of cancer and proportion (1-π) die due to other causes. Suppose that an individual can experience one of J failure types, and let us define T be the time of failure and δ be the failure type. At the end of the study, we have the following types of observations;

$$1.\ \{T = t,\ \ \delta = j\}, \qquad 2.\ \{T > t\}, \qquad 3.\ \{T > t,\ \ \delta = j\}, \qquad 4.\ \{T = t\}. \tag{1}$$

In this paper we provide Bayesian inference of the unknown parameters, under latent failure times model formulations. If T_i denotes the lifetime of the i-th individual then

$$T_i = \min\{T_{i1}, \cdots, T_{iM}\},$$

where $T_{i1}, \cdots, T_{iM}$ are the latent failure times of the M different causes for the i-th individual. According to the latent failure time model assumptions, $T_{i1}, \cdots, T_{iM}$ are independently distributed. Moreover, $T_{i1}, \cdots, T_{iM}$ are not observable, only T_i is observable, and the indicator J, such that $T_{iJ} = \min\{T_{i1}, \cdots, T_{iM}\}$ is observable. Moreover, it is assumed that T_{im} for $m = 1, \cdots, M$, follows a Weibull distribution with the probability density function (PDF)

$$f(t; \alpha, \lambda_m) = \begin{cases} \alpha\lambda_m e^{-\lambda_m t^{\alpha}} t^{\alpha-1} & \text{if} \quad t > 0 \\ 0 & \text{if} \quad t < 0, \end{cases} \tag{2}$$

here $\alpha > 0$ and $\lambda_m > 0$ are the shape and scale parameters of the Weibull distribution.

For developing the Bayesian inference of the unknown parameters, we need to assume some priors on the unknown parameters. When the common shape parameter α is known, it is quite natural to assume the convenient but quite general conjugate priors on the scale parameters, namely the Beta-Gamma (for $M > 2$ it is Dirichlet-Gamma) priors, see Pena and Gupta [11] or Kundu and Pradhan [10] in this respect. In this case, the explicit Bayes estimates of the unknown parameters can be obtained mainly under squared error loss functions. But when the common shape parameter is not known, we assume that the shape parameter has a log-concave prior density function, similarly as in Berger and Sun [2]. Based on the above prior distributions, we obtain the joint posterior density function of the unknown parameters. In this case, the Bayes estimators of the unknown parameters cannot be obtained in explicit forms, as expected. We propose to use Monte Carlo sampling technique to compute the Bayes estimates, and also to compute approximate highest posterior density (HPD) credible intervals. We analyze one data set using the proposed Bayesian technique.

It is further assumed that the underlying latent survival times of an individual depend on a set of covariate vector $\boldsymbol{x} = (x_1, \ldots, x_p)^T$, which does not depend on time. There are several ways to incorporate covariates in the model. In this case we introduce covariates in the model through the scale parameters of the latent survival distributions. In presence of covariate vector $\boldsymbol{x}$, we assume the following regression equation

$$\lambda_{1i} = \theta_1 exp(\boldsymbol{\beta}^T \boldsymbol{x}_i), \ldots, \lambda_{Mi} = \theta_M exp(\boldsymbol{\beta}^T \boldsymbol{x}_i),$$

where $\theta_1 > 0, \ldots, \theta_M > 0$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the regression parameter vector, and $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{pi})^T$, for $i = 1, \ldots, n$. Based on the Beta-Gamma (for $M > 2$ it is Dirichlet-Gamma) prior on θ_i, independent normal priors on β_i and log-concave prior on the common shape parameter α, the posterior density function can be obtained. It is observed that the Bayes estimates and the associated credible intervals of the unknown parameters cannot be obtained in closed form, and we propose to use the Monte Carlo sampling technique to compute simulation consistent Bayes estimates and also to construct associated credible intervals. The analysis of one data set in presence of covariates has been performed, and it is observed that the proposed method is very easy to implement in practice.

Rest of the paper is organized as follows. In Section 2 we provide model formulations and prior assumptions. Posterior analysis and Bayesian inferences are provided in Section 3. In Section 4, we present the results based on covariates. The analysis of two data sets are presented in Section 5. Finally we conclude the paper in Section 6.

We use the following notations throughout.

T_i :	lifetime of the i-th individual
T_{ji} :	latent failure time of mode j of the i-th individual
$F(.)$:	cumulative distribution function of T_i
$F_j(.)$:	cumulative distribution function of T_{ji}
$f(.)$:	density function of T_i
$f_j(.)$:	density function of T_{ji}
$S_j(.)$:	$1 - F_j(.)$

$S(.)$:	$1 - F(.)$
$\delta_i(.)$:	indicator variable denoting the cause of failure of the i-th individual
$\lvert I \rvert$:	the number of elements in the set I
$exp(\lambda)$:	exponential random variable with density function $\lambda e^{-\lambda x}$
$Weibull(\alpha, \lambda)$:	Weibull random variable with density function $\alpha\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}}$
$Gamma(\alpha, \lambda)$:	gamma random variable with density function $\frac{\lambda^{\alpha}}{\Gamma\alpha}x^{\alpha-1}e^{-\lambda x}$
$\chi^2_{k,\alpha}$:	the lower α-th percentile point of the central χ^2 distribution with k-degrees of freedom
$N(a, b^2)$:	Normal distribution with mean a and variance b^2.

2 Models and Priors

2.1 Models

For notational simplicity, we assume that there are only two causes of failures. All the results presented here can be easily extended for any finite number of causes of failures. We assume here that $(T_{1i}, T_{2i}); i = 1, \ldots, n$, are n independent and identically distributed (*i.i.d.*) random variables and $T_i = \min\{T_{1i}, T_{2i}\}$. Without loss of generality, we further assume the following. The first r_1 observations have complete failure times and the corresponding cause of failure is 1 for all of them. We denote this set as I_1. Similarly, the next r_2 observations have complete failure times and the corresponding cause of failure is 2. The corresponding set is denoted by I_2. The next r_3 observations have complete failure times but the corresponding causes of failures are unknown, and the set is denoted by I_3. The set of next r_4 observations is denoted by I_4, where the failure type is known to be 1, but the failure times are censored. Similarly, the set of next r_5 observations is denoted by I_5 where the failure times are censored but the failure type is known to be 2 for all of them. Finally, the last r_6 right censored observations where the exact failure time and failure type both are unknown. The set is denoted by I_6. Therefore, it is assumed that any observation will be from one of the I_j's, for $j = 1, \ldots, 6$.

We further denote $r_1 + r_2 + r_3 = n$, $r_4 + r_5 + r_6 = m$, $r_1 + r_2 = n_1$, $r_4 + r_5 = m_1$ and $N = m + n$. In order to analyze the incomplete data it is assumed that the failure times are from the same population as the complete data, *i.e.*, the population remains unchanged irrespective of the cause of failure. Moreover, the likelihood contributions from the observations from different sets are

$$f_1(t)S_2(t), \quad f_2(t)S_1(t), \quad f(t), \quad \int_x^\infty f_1(y)S_2(y)dy, \quad \int_x^\infty f_2(y)S_1(y)dy, \quad S(t), \tag{3}$$

respectively. When the covariates are also present, it is assumed that

$$\lambda_{1i} = \theta_1 exp(\boldsymbol{\beta}^T \boldsymbol{x}_i) \quad \text{and} \quad \lambda_{2i} = \theta_2 exp(\boldsymbol{\beta}^T \boldsymbol{x}_i), \tag{4}$$

as mentioned before. Based on the above assumption, the likelihood contribution of an individual having an observation from one of those six sets, can be easily obtained as in (3).

2.2 Priors

Let us denote $\lambda = \lambda_1 + \lambda_2$, $p = \dfrac{\lambda_1}{\lambda}$, hence $1 - p = \dfrac{\lambda_2}{\lambda}$. Similarly as in Pena and Gupta [11], it is assumed that λ has a Gamma(a_0, b_0) prior, with $a_0 > 0, b_0 > 0$, and p has a Beta(a_1, a_2) prior, with $a_1 > 0, a_2 > 0$, *i.e.*

$$\pi_1(\lambda|a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)}\lambda^{a_0-1}e^{-b_0\lambda}; \quad \lambda > 0 \tag{5}$$

$$\pi_2(p|a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)}p^{a_1-1}(1-p)^{a_2-1}; \quad 0 < p < 1, \tag{6}$$

and they are independently distributed. After simplification, the joint PDF of λ_1 and λ_2 takes the following form:

$$\pi(\lambda_1, \lambda_2|a_0, b_0, a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_0)}(b_0\lambda)^{a_0-a_1-a_2} \times \frac{b_0^{a_1}}{\Gamma(a_1)}\lambda_1^{a_1-1}e^{-b_0\lambda_1} \times \frac{b_0^{a_2}}{\Gamma(a_2)}\lambda_2^{a_2-1}e^{-b_0\lambda_2}. \tag{7}$$

This is the Beta-Gamma distribution and it will be denoted by $\text{BG}(b_0, a_0, a_1, a_2)$. For known α, the prior (7) on (λ_1, λ_2) is a conjugate prior. It is a very flexible prior. Although, in general λ_1 and λ_2 are dependent, if $a_0 = a_1 + a_2$, λ_1 and λ_2 become independent. Moreover, if $a_0 > a_1 + a_2$, then the covariance between λ_1 and λ_2 become positive, and when $a_0 < a_1 + a_2$, it is negative. Note that if $(\lambda_1, \lambda_2) \sim \text{BG}(b_0, a_0, a_1, a_2)$, then $p \sim \text{Beta}(a_1, a_2)$ and $\lambda \sim \text{Gamma}(a_0, b_0)$.

When the shape parameter α is known, the above $\text{BG}(b_0, a_0, a_1, a_2)$ prior is a conjugate prior on (λ_1, λ_2). When the shape parameter α is not known, the conjugate priors on $(\alpha, \lambda_1, \lambda_2)$ do not exist. In this case it is assumed that the prior on α, $\pi_2(\alpha)$ is absolute continuous and it has a log-concave PDF. (λ_1, λ_2) has the prior $\pi_1(\cdot)$ as defined in (7) and it is independent of $\pi_2(\alpha)$.

In presence of covariates, when $(\lambda_{1i}, \lambda_{2i})$ has the form (4), we make the following prior assumptions on the unknown parameters

$$\begin{aligned} (\theta_1, \theta_2) &\sim \pi_1(\theta_1, \theta_2) = \text{BG}(b_0, a_0, a_1, a_2) \\ \alpha &\sim \pi_2(\alpha) \\ \beta_j &\sim \pi_3(\beta_j) = N(0, \sigma_j^2); \quad j = 1, \ldots, p. \end{aligned}$$

We assume the independence among the prior parameters, i.e. it is assumed that (θ_1, θ_2), α and β_j for $j = 1, \ldots, p$ are apriori independent. Here $\pi_2(\alpha)$ has the same form as defined above.

The following result related to Beta-Gamma distribution will be useful for further development, and the proof can be easily obtained from Theorem 2 of Pena and Gupta [11].

LEMMA 1: If $(\lambda_1, \lambda_2) \sim \text{BG}(b_0, a_0, a_1, a_2)$, then for $i = 1,2$

$$\text{(a)} \quad E(\lambda_i) = \frac{a_0 a_i}{b_0(a_1 + a_2)} \quad \text{and} \quad \text{(b)} \quad V(\lambda_i) = \frac{a_0 a_i}{b_0^2(a_1 + a_2)} \times \left\{ \frac{(a_i + 1)(a_0 + 1)}{a_1 + a_2 + 1} - \frac{a_0 a_i}{a_1 + a_2} \right\}.$$

3 Posterior Analysis

In this section we consider the Bayesian inferences of the unknown parameters based on the prior assumptions provided in Section 2. In developing the Bayes estimates it is assumed that the loss function is squared error loss function, although other loss functions also can be easily incorporated. We consider two cases separately namely when (i) shape parameter is known, (ii) shape parameter is unknown.

3.1 Shape Parameter is Known

Based on the observations as described in Section 2, and using the same notations as has been mentioned, the likelihood function can be written as

$$l(Data|\lambda_1, \lambda_2, \alpha) = \lambda_1^{r_1+r_4} \lambda_2^{r_2+r_5} (\lambda_1 + \lambda_2)^{r_3-m_1} e^{-(\lambda_1+\lambda_2)\sum_{i\in I} t_i^{\alpha}} \times \alpha^{n_1} \prod_{i\in I_1\cup I_2\cup I_3} \left\{t_i^{\alpha-1}\right\}. \quad (8)$$

If we use $\lambda = \lambda_1 + \lambda_2$ and $p = \dfrac{\lambda_1}{\lambda}$, the likelihood function (8) can be written as follows;

$$l(Data|p, \lambda, \alpha) = p^{r_1+r_4}(1-p)^{r_2+r_5} \lambda^{r_1+r_2+r_3} e^{-\lambda \sum_{i\in I} t_i^{\alpha}} \times \alpha^{n_1} \prod_{i\in I_1\cup I_2\cup I_3} \left\{t_i^{\alpha-1}\right\}. \quad (9)$$

Now based on the priors on λ and p as mentioned in Section 2, the joint posterior density of λ and p can be written as

$$\pi(p, \lambda|Data, \alpha) \propto p^{r_1+r_4+a_1-1}(1-p)^{r_2+r_5+a_2-1} \lambda^{r_1+r_2+r_3+a_0-1} e^{-\lambda(b_0+\sum_{i\in I} t_i^{\alpha})}. \quad (10)$$

It is interesting to see that for known α, aposteriori also λ and p are independently distributed. Therefore, for known α, the posterior distribution of λ_1 and λ_2 becomes,

$$\pi(\lambda_1, \lambda_2|Data, \alpha) \sim \text{BG}(b_0 + T(\alpha), r_1 + r_2 + r_3 + a_0, r_1 + r_4 + a_1, r_2 + r_5 + a_2), \quad (11)$$

here $T(\alpha) = \sum_{i\in I} t_i^{\alpha}$. Therefore, the Bayes estimates of λ_1 and λ_2 under squared error loss functions are

$$\widehat{\lambda}_{1B}(\alpha) = \frac{(n + a_0)(r_1 + r_4 + a_1)}{(b_0 + T(\alpha))(n_1 + m_1 + a_1 + a_2)} \quad \text{and} \quad \widehat{\lambda}_{2B}(\alpha) = \frac{(n + a_0)(r_2 + r_5 + a_2)}{(b_0 + T(\alpha))(n_1 + m_1 + a_1 + a_2)},$$

respectively. The corresponding posterior variances are

$$V(\widehat{\lambda}_{1B}(\alpha)) = A_1 \times B_1 \qquad \text{and} \qquad V(\widehat{\lambda}_{2B}(\alpha)) = A_2 \times B_2,$$

where A_i and B_i for $i = 1, 2$, are given below

$$\begin{aligned}
A_1 &= \frac{(r_1 + r_2 + r_3 + a_0)(r_1 + r_4 + a_1)}{(b_0 + T(\alpha))^2(r_1 + r_2 + r_4 + r_5 + a_1 + a_2)} \\
B_1 &= \frac{(r_1 + r_4 + a_1 + 1)(r_1 + r_2 + r_3 + a_0 + 1)}{r_1 + r_2 + r_4 + r_5 + a_1 + a_2 + 1} - \frac{(r_1 + r_2 + r_3 + a_0)(r_1 + r_4 + a_1)}{r_1 + r_2 + r_4 + r_5 + a_1 + a_2} \\
A_2 &= \frac{(r_1 + r_2 + r_3 + a_0)(r_2 + r_5 + a_2)}{(b_0 + T(\alpha))^2(r_1 + r_2 + r_4 + r_5 + a_1 + a_2)} \\
B_2 &= \frac{(r_2 + r_5 + a_2 + 1)(r_1 + r_2 + r_3 + a_0 + 1)}{r_1 + r_2 + r_4 + r_5 + a_1 + a_2 + 1} - \frac{(r_1 + r_2 + r_3 + a_0)(r_2 + r_5 + a_2)}{r_1 + r_2 + r_4 + r_5 + a_1 + a_2}.
\end{aligned}$$

Under the assumptions of non-informative priors, *i.e.* when $a_0 = b_0 = a_1 = b_1 = 0$, the Bayes estimates of λ_1 and λ_2 become

$$\widehat{\lambda}_{1B}(\alpha) = \frac{n(r_1 + r_4)}{T(\alpha)(n_1 + m_1)} \quad \text{and} \quad \widehat{\lambda}_{2B}(\alpha) = \frac{n(r_2 + r_5)}{T(\alpha)(n_1 + m_1)}, \tag{12}$$

respectively, and they coincide with the corresponding maximum likelihood estimators (MLEs), see for example Kundu [8].

Note that although, the Bayes estimates of the unknown parameters can be obtained in explicit form, the corresponding credible intervals of the unknown parameters cannot be obtained in explicit forms. One way to construct individual credible intervals of λ_1 and λ_2 is by direct sampling from the joint posterior density functions, as it has been suggested by Kundu and Pradhan [10]. In this case posterior samples from the joint posterior density function (11) can be easily generated, and in turn they can be used to construct approximate highest posterior density (HPD) credible intervals of λ_1 and λ_2.

Now we describe how to construct an exact 100(1-γ)% credible set of (λ_1, λ_2). Note that $C_{\alpha,1-\gamma}(\lambda_1, \lambda_2)$ is said to be a 100(1-γ)% credible set of λ_1 and λ_2, if

$$P[(\lambda_1, \lambda_2) \in C_{\alpha,1-\gamma}(\lambda_1, \lambda_2)] = 1 - \gamma, \tag{13}$$

when $(\lambda_1, \lambda_2) \sim \pi(\lambda_1, \lambda_2 | Data, \alpha)$. We need the following lemma for further development.

Lemma 2: If

$$(X, Y) \sim \text{BG}(b_0 + T(\alpha), r_1 + r_2 + r_3 + a_0, r_1 + r_4 + a_1, r_2 + r_5 + a_2),$$

then

$$\begin{aligned} Z = X + Y &\sim \text{Gamma}(r_1 + r_2 + r_3 + a_0, b_0 + T(\alpha)), \\ V = \frac{X}{X+Y} &\sim \text{Beta}(r_1 + r_4 + a_1, r_2 + r_5 + a_2), \end{aligned}$$

and Z and V are independently distributed.

Proof: It can be obtained by simple transformation technique. ∎.

Using Lemma 2, $C_{\alpha,1-\gamma}(\lambda_1, \lambda_2)$ can be constructed as follows:

$$C_{\alpha,1-\gamma}(\lambda_1, \lambda_2) = \{(\lambda_1, \lambda_2); \lambda_1 > 0, \lambda_2 > 0, A \le \lambda_1 + \lambda_2 \le B, C \le \frac{\lambda_1}{\lambda_1 + \lambda_2} \le D\}, \tag{14}$$

Here A, B, C and D are such that

$$P(A \le Z \le B) \times P(C \le V \le D) = 1 - \gamma, \tag{15}$$

and Z and V are two independent random variables as in Lemma 2. It simply follows that $C_{\alpha,1-\gamma}(\lambda_1, \lambda_2)$ is a trapezoid enclosed by the following four straight lines;

$$(i)\ x + y = A, \quad (ii)\ x + y = B, \quad (iii)\ x(1 - D) = yD, \quad (iv)\ x(1 - C) = yC.$$

Simple calculation shows that the area of the trapezoid $C_{\alpha,1-\gamma}(\lambda_1, \lambda_2)$ is $(B^2 - A^2)(D - C)/2$. Therefore, if we want to find the credible set with the smallest area, then we need to find A, B, C, D, so that they satisfy (15), and $(B^2 - A^2)(D - C)$ is minimum. This has to be performed numerically, and it is not pursued here further.

3.2 Shape Parameter is Unknown

Now we will discuss the more important case *i.e.* when the shape parameter α is also unknown. Based on the priors on α, λ and p, as defined in Section 2, we obtain the joint posterior density function of λ, p and α as follows;

$$\pi(p, \lambda, \alpha | Data) = \pi(p, \lambda | Data, \alpha) \times \pi(\alpha | Data), \tag{16}$$

here $l(p, \lambda | Data, \alpha)$ is same as (10), and

$$\pi(\alpha | Data) = k \times \pi_2(\alpha) \times \alpha^{n_1} \times \prod_{i \in I_1 \cup I_2 \cup I_3} t_i^{\alpha} \times \frac{1}{(b_0 + T(\alpha))^{r_1 + r_2 + r_3 + a_0 - 1}}, \tag{17}$$

here k is the normalizing constant. Therefore, the posterior density function of λ_1, λ_2 and α can be written as

$$\pi(\lambda_1, \lambda_2, \alpha | Data) = \pi(\lambda_1, \lambda_2 | Data, \alpha) \times \pi(\alpha | Data), \tag{18}$$

here $\pi(\lambda_1, \lambda_2 | Data, \alpha)$ and $\pi(\alpha | Data)$ are same as defined in (11) and (17) respectively. Therefore, the Bayes estimate of any function of $\alpha, \lambda_1, \lambda_2$, say $g(\alpha, \lambda_1, \lambda_2)$ with respect to the squared error loss function is

$$\widehat{g}_B(\alpha, \lambda_1, \lambda_2) = E_{\lambda_1, \lambda_2, \alpha | Data}(g(\alpha, \lambda_1, \lambda_2)). \tag{19}$$

Clearly in general (19) cannot be obtained in closed form and it will involve three dimensional integration. We propose to use direct sampling method to approximate (19), and they can be used to construct credible intervals also. We need the following result for further development.

Lemma 3: The posterior distribution λ_1 and λ_2 given α is given by (11).

Lemma 4: The posterior density function of α, *i.e.* $l(\alpha | Data)$ is log-concave.

Proof: Proof can be obtained similarly as the proof of Theorem 2 of Kundu [9]. ∎

Devroye [6] suggested a method to generate samples from a general log-concave density function. Therefore, samples from the posterior density function of α, namely $l(\alpha|Data)$, can be easily generated. Alternatively as suggested by Kundu [9], $l(\alpha|Data)$ can be very well approximated by a two-parameter gamma density function. Hence two-parameter gamma density function also can be used for generating samples from $l(\alpha|Data)$. Generation of samples from a Beta-Gamma distribution can be performed using the algorithm suggested by Kundu and Pradhan [10]. Therefore, using Lemma 3 and Lemma 4, generation from the posterior density function $l(\lambda_1, \lambda_2, \alpha|Data)$ can be easily performed. Once the posterior samples are obtained they can be used to compute the Bayes estimate of $g(\alpha, \lambda_1, \lambda_2)$ and also to construct credible interval of $g(\alpha, \lambda_1, \lambda_2)$.

The following algorithm is suggested to compute the Bayes estimates and the associated HPD credible intervals for the unknown parameters.

Algorithm 1:

- Step 1: Generate α from $\pi(\alpha|Data)$ as given in (17) using the the method suggested by Devroye [6], or the approximation method suggested by Kundu [9].

- Step 2: For a given α, generate λ_1 and λ_2 from the conditional posterior distribution of λ_1 and λ_2 given α as given in (11), as suggested by the method of Kundu and Pradhan [10].

- Step 3: Repeat Step 1 and Step 2, to generate $(\alpha_1, \lambda_{11}, \lambda_{21}), \cdots, (\alpha_M, \lambda_{1M}, \lambda_{2M})$.

- Step 4: The Bayes estimate of $g(\alpha, \lambda_1, \lambda_2)$ and the corresponding posterior variance can be obtained as

$$\begin{aligned}
\widehat{g}(\alpha, \lambda_1, \lambda_2) &= \frac{1}{M}\sum_{i=1}^{M} g(\alpha_i, \lambda_{1i}, \lambda_{2i}), \qquad \text{and} \\
\widehat{V}(g(\alpha, \lambda_1, \lambda_2)) &= \frac{1}{M}\sum_{i=1}^{M} \left(g(\alpha_i, \lambda_{1i}, \lambda_{2i}) - \widehat{g}(\alpha, \lambda_1, \lambda_2)\right)^2,
\end{aligned}$$

respectively.

- Step 5: To construct the HPD credible of $g(\alpha, \lambda_1, \lambda_2)$, first order g_i as $g_{(1)} < g_{(2)} < \cdots < g_{(M)}$, where $g_i = g(\alpha_i, \lambda_{1i}, \lambda_{2i})$. Then 100(1-2$\beta$)% credible interval of $g(\alpha, \lambda_1, \lambda_2)$ becomes
$$(g_{(j)}, g_{(j+M-2\beta)}), \quad \text{for} j = 1, \cdots, 2M\beta.$$
Therefore, 100(1-2β)% HPD credible interval becomes $(g_{(j^*)}, g_{(j^*+M-2\beta)})$, where j^* is such that
$$g_{(j^*+M-2\beta)} - g_{(j^*)} \leq g_{(j+M-2\beta)} - g_{(j)}$$
for all $j = 1, \cdots, 2M\beta$.

COMMENTS: Similar method as it has been mentioned in Section 3.1 can be used to construct 100(1-γ)% credible set of $(\lambda_1, \lambda_2, \alpha)$. In this case a set $C_{1-\gamma}(\lambda_1, \lambda_2, \alpha)$ is said to be a 100(1-γ)% credible set of $(\lambda_1, \lambda_2, \alpha)$ if
$$P((\lambda_1, \lambda_2, \alpha) \in C_{1-\gamma}(\lambda_1, \lambda_2, \alpha) = 1 - \gamma,$$
where $(\lambda_1, \lambda_2, \alpha) \sim \pi(\lambda_1, \lambda_2, \alpha|Data)$. Now choose β and δ such that $(1-\beta)\times(1-\delta) = (1-\gamma)$. Then a 100(1-$\gamma$)% credible interval of $(\lambda_1, \lambda_2, \alpha)$ can be obtained as follows:
$$C_{1-\gamma}(\lambda_1, \lambda_2, \alpha) = (\alpha_L, \alpha_U) \times C_{\alpha,1-\delta}(\lambda_1, \lambda_2),$$
here α_L and α_U are such that,
$$\int_{\alpha_L}^{\alpha_U} \pi(\alpha|Data) d\alpha = 1 - \beta,$$
and $C_{\alpha,1-\delta}(\lambda_1, \lambda_2)$ is same as defined in Section 3.1.

4 Presence of Covariates

In this section we consider the case when for each individual there exists a set of covariate vector $\boldsymbol{x}$. We use the following notation for providing the posterior distribution function.

Let us denote $T(\alpha,\boldsymbol{\beta}) = \sum_{i\in I} t_i^{\alpha} exp(\boldsymbol{\beta}^T \mathbf{x}_i)$, and $s_1 = r_1 + r_4 + a_1$, $s_2 = r_2 + r_5 + a_2$, $s_0 = r_1 + r_2 + r_3 + a_0$. Based on the model and prior assumptions as described in Section 2, the posterior distribution function of $\theta_1, \theta_2, \alpha, \boldsymbol{\beta}$ can be written as

$$\begin{aligned} l(\theta_1,\theta_2,\alpha,\boldsymbol{\beta}|Data) \;&\propto\; \theta_1^{s_1-1}\theta_2^{s_2-1}(\theta_1+\theta_2)^{s_0-s_1-s_2} exp(-(\theta_1+\theta_2)(T(\alpha,\boldsymbol{\beta})+b_0)) \times \\ &\pi(\alpha)\alpha^n \left\{ \prod_{i\in I_1\cup I_2\cup I_3} t_i^{\alpha} \right\} \times exp(\sum_{i\in I_1\cup I_2\cup I_3} \boldsymbol{\beta}^T \mathbf{x}_i) \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\sigma_i} exp(-\frac{\beta_i^2}{2\sigma_i^2}). \end{aligned} \tag{20}$$

Therefore, (20) can be written as

$$l(\theta_1,\theta_2,\alpha,\boldsymbol{\beta}|Data) = k \times \mathrm{BG}(T(\alpha,\boldsymbol{\beta})+b_0, s_0, s_1, s_2) g_1(\alpha|Data) g_2(\boldsymbol{\beta}|Data) h(\alpha,\boldsymbol{\beta}).$$

Here $g_1(\alpha|Data)$, $g_2(\boldsymbol{\beta}|Data)$ are proper density functions such that

$$g_1(\alpha|Data) \;\propto\; \pi_2(\alpha)\alpha^n \left\{ \prod_{i\in I_1\cup I_2\cup I_3} t_i^{\alpha} \right\}, \tag{21}$$

$$g_2(\boldsymbol{\beta}|Data) \;\propto\; exp(\sum_{i\in I_1\cup I_2\cup I_3} \boldsymbol{\beta}^T \mathbf{x}_i) \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\sigma_i} exp(-\frac{\beta_i^2}{2\sigma_i^2}), \tag{22}$$

$$h(\alpha,\boldsymbol{\beta}) \;=\; \frac{1}{(T(\alpha,\boldsymbol{\beta})+b_0)^{s_0}} \tag{23}$$

and k is the normalizing constant.

If we want to obtain the Bayes estimate of any function of the unknown parameters, say $g(\theta_1,\theta_2,\alpha,\boldsymbol{\beta})$, then for the squared error loss function it is

$$\widehat{g}_B = \frac{\int\cdots\int g(\theta_1,\theta_2,\alpha,\boldsymbol{\beta})\mathrm{BG}(T(\alpha,\boldsymbol{\beta})+b_0,s_0,s_1,s_2)g_1(\alpha|Data)g_2(\boldsymbol{\beta}|Data)h(\alpha,\boldsymbol{\beta})d\theta_1 d\theta_2 d\alpha d\boldsymbol{\beta}}{\int\cdots\int \mathrm{BG}(T(\alpha,\boldsymbol{\beta})+b_0,s_0,s_1,s_2)g_1(\alpha|Data)g_2(\boldsymbol{\beta}|Data)h(\alpha,\boldsymbol{\beta})d\theta_1 d\theta_2 d\alpha d\boldsymbol{\beta}}. \tag{24}$$

Clearly, (24) cannot be obtained in explicit form in most of the cases. We propose to use the importance sampling technique to compute Bayes estimate $\widehat{g}_B$ and also to construct the associated credible interval. The following observations will be useful. Note that $g_1(\alpha|Data)$ is log-concave and

$$g_2(\boldsymbol{\beta}|Data) = \prod_{j=1}^{p} N(u_j\sigma_j, \sigma_j^2), \tag{25}$$

here $u_j = \sum_{i \in I_1 \cup I_2 \cup I_3} x_{ji}$, for $j = 1, \ldots, p$.

The following algorithm can be used to compute Bayes estimate of $g(\theta_1, \theta_2, \alpha, \boldsymbol{\beta})$ and to construct associated credible interval.

Algorithm 2:

- Step 1: Generate $\alpha_1 \sim g_1(\alpha | Data)$, and $\beta_{j1} \sim \text{N}(u_j \sigma_j, \sigma_j)$, for $j = 1, \ldots, p$. Let $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{p1})$
- Step 2: For a given α_1 and $\boldsymbol{\beta}_1$, generate $(\theta_{11}, \theta_{21}) \sim \text{BG}(T(\alpha_1, \boldsymbol{\beta}_1) + b_0, s_0, s_1, s_2)$.
- Step 3: Repeat steps 1 and 2, and obtain $\{(\alpha_1, \theta_{11}, \theta_{21}, \boldsymbol{\beta}_1), \ldots, (\alpha_N, \theta_{1N}, \theta_{2N}, \boldsymbol{\beta}_N)\}$.
- Step 4: A simulation consistent Bayes estimate can be obtained as

$$\widehat{g}_B = \frac{\sum_{i=1}^{N} g(\theta_{1i}, \theta_{2i}, \alpha_i, \boldsymbol{\beta}_i) h(\alpha_i, \boldsymbol{\beta}_i)}{\sum_{j=1}^{N} h(\alpha_j, \boldsymbol{\beta}_j)}$$

- Step 5: Now to construct 100(1-γ)% HPD credible interval of $g(\theta_1, \theta_2, \alpha, \boldsymbol{\beta})$, first let us denote

$$g_i = g(\theta_{1i}, \theta_{2i}, \alpha_i, \boldsymbol{\beta}_i) \quad \text{and} \quad w_i = \frac{h(\alpha_i, \boldsymbol{\beta}_i)}{\sum_{j=1}^{N} h(\alpha_j, \boldsymbol{\beta}_j)}, \quad \text{for} \quad i = 1, \ldots, N.$$

  Rearrange, $\{(g_1, w_1), \ldots, (g_N, w_N)\}$ as $\{(g_{(1)}, w_{[1]}), \ldots, (g_{(N)}, w_{[N]})\}$, where $g_{(1)} < \ldots < g_{(N)}$. In this case $w_{[i]}$'s are not ordered, they are just associated with $g_{(i)}$. Let N_p be the integer satisfying

$$\sum_{i=1}^{N_p} w_{[i]} \leq p < \sum_{i=1}^{N_p+1} w_{[i]}$$

  for $0 < p < 1$. A 100(1-γ)% credible interval of $g(\theta_1, \theta_2, \alpha, \boldsymbol{\beta})$ can be obtained as $(g_{(N_\delta)}, g_{(N_{\delta+1-\gamma})})$, for $\delta = w_{[1]}, w_{[1]} + w_{[2]}, \ldots, \sum_{i=1}^{N_{1-\gamma}} w_{[i]}$. Therefore, a 100(1-$\gamma$)% HPD credible interval of $g(\theta_1, \theta_2, \alpha, \boldsymbol{\beta})$ becomes $(g_{(N_{\delta^*})}, g_{(N_{\delta+1-\gamma^*})})$, where

$$g_{(N_{\delta^*+1-\gamma})} - g_{(N_{\delta^*})} \leq g_{(N_{\delta+1-\gamma})} - g_{(N_\delta)} \quad \text{for all } \delta.$$

5 Data Analysis

Data Set 1: This data set was originally analyzed by Dinse [7] using non-parametric method. It was from a study of lymphocytic non-Hodgkins lymphoma, conducted by the Eastern Cooperative Oncology Group. It has the survival time (in weeks) and also indicates whether a patient is judged asymptomatic or symptomatic. Here symptoms monitored include weight loss, fever and night sweats. In this case, the observations fall into all six categories. Some patients die during the clinical trial and others are alive at the time of analysis. For some patients, the initial presence and absence of symptoms are absent. The data set indicates survival times of 79 male Stage 4 patients entered on the maintenance phase of the trial. Approximately 35% of the survival times are censored, and over 50% of the patients lack classifications on the symptoms indicator. Almost 50% of the observations are discordant. The summary of the data is as follows; $r_1 = 16$, $r_2 = 9$, $r_3 = 26$, $r_4 = 12$, $r_5 = 1$, $r_6 = 15$, $n_1 = 25$, $n = 51$, $m = 28$, $m_1 = 13$, $N = 79$. $\sum_{i \in I_1} t_i = 3314$, $\sum_{i \in I_2} t_i = 1244$, $\sum_{i \in I_3} t_3 = 3616$, $\sum_{i \in I_4} t_i = 4198$, $\sum_{i \in I_5} t_i = 362$, $\sum_{i \in I_6} t_i = 5207$. For computational purposes, we have divided all the observations by 100, it is not going to make any difference in the statistical inference.

To perform the Bayesian analysis we need to assume some specific form of $\pi_2(\alpha)$. We have assumed $\pi(\alpha) \sim \text{Gamma}(c, d)$. We assume all the hyper-parameters to be zero, *i.e.* $b_0 = a_0 = a_1 = a_2 = c = d = 0$. It may be noted that $\pi_2(\alpha)$, when $c = d = 0$, is not log-concave, but the posterior density function of α given data is log-concave. Based on the non-informative priors, using the gamma approximation suggested by Kundu [9], we have generated samples from the posterior distribution of $\pi(\alpha|Data)$, and for a given α, we have generated (λ_1, λ_2) from the posterior density function $\pi(\lambda_1, \lambda_2|Data, \alpha)$ In Figure 1 we provide the histogram of the generated α, and for comparison purposes, we have also provided the exact PDF of $\pi(\alpha|Data)$ in Figure 1. It is clear that the gamma approximation works very well. The Bayes estimates of α, λ_1 and λ_2 under squared error loss functions are 1.3427, 0.3719 and 0.1377 respectively. The corresponding 95% credible intervals are (0.9532, 1.8021), (0.2476,0.5213) and (0.0664,0.2348) respectively.

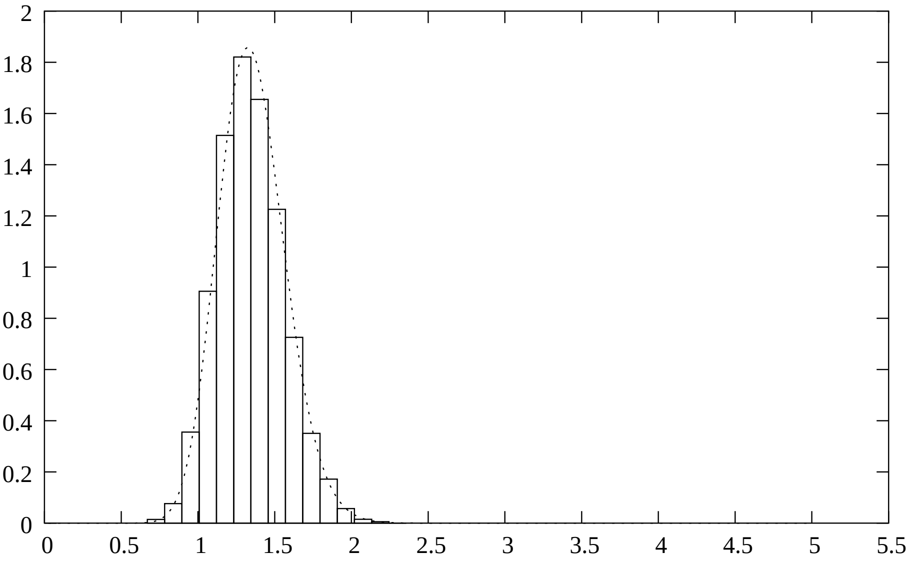

Figure 1: Histogram of the samples generated from the posterior distribution of α and the true posterior density function based on Prior 1.

Data Set 2: This data set is consisting of 65 patients with melanoma survival data collected at Odense Hospital Denmark by Drzewiecki, K.T, and it has been obtained from Anderson *et al.* [1]. In this case survival time in days and the corresponding cause of death, either cancer or other, are reported. For some patients the cause of death is not known. For each patient we have used the single covariate namely age (reported in years). Here observations fall into only three categories. The summary of the data set is as follows: $r_1 = 42$, $r_2 = 10$, $r_3 = 13$, $n = 65$, $m = 0$, $n_1 = 52$, $m_1 = 0$, $N = 65$, $\sum_{i \in I_1} t_i = 37230$, $\sum_{i \in I_2} t_i = 6509$, $\sum_{i \in I_3} t_3$ $= 19010$. For computational purposes we have divided all the survival times by 1000, and all the ages by 100.

To perform the Bayesian analysis, we have assumed as before that $\pi_2(\alpha) \sim \text{Gamma}(c, d)$. The hyperparameters are assumed to be zero as in the previous example. The prior variance of β is assumed to be 0.25. Based on importance sampling, with 10000 replications, we obtain the Bayes estimates of α, θ_1, θ_2 and β as 2.812, 0.448, 0.100 and -0.164 respectively. The associated 95% HPD credible intervals are (2.659, 3.034), (0.382,0.543), (0.081,0.128) and (-0.369,0.134) respectively. It is observed that the age does not have a significant effect on the survival time.

6 Conclusions

In this paper we consider the Bayesian inference of the competing risks data when the data might be incomplete both in time and type of failures. We have mainly restricted the attention when there are only two causes of failures, although our method can be easily generalized for more than two failures also. It is assumed that the lifetime of the latent failures distributions follow Weibull distributions with the same shape parameter but different scale parameter. Based on fairly general priors on the scale and shape parameters the Bayes estimates are obtained using Markov Chain Monte Carlo technique. We have further considered the case when each individual has a set of covariates. In this case the Bayes estimates and the associated credible intervals are obtained using importance sampling technique. The proposed model is very flexible, and the method is very easy to implement. Moreover, it is possible to obtain small sample results which is a clear advantage compared to the frequentist inference.

References

[1] Anderson, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993), *Statistical Models based on Counting Process*, Springer, New York.

[2] Berger, J.O. and Sun, D. (1993), "Bayesian analysis for the Poly-Weibull distribution", *Journal of the American Statistical Association*, vol. 88, 1412-1418.

[3] Chen, M.H. and Shao, Q.M. (1999), "Monte Carlo estimation of Bayesian Credible and HPD intervals", *Journal of Computational and Graphical Statistics*, vol. 8, 69 - 92.

[4] Cox, D. R. (1959), "The analysis of exponentially distributed lifetimes with two types of failures", *Journal of the Royal Statistical Society, Series B*, vol. 21, 411 - 421.

[5] Crowder, M. J. (2001), *Classical Competing Risks*, Chapman and Hall, Boca Raton, FL.

[6] Devroye, L. (1984), "A simple algorithm for generating random variables using log-concave density function", *Computing*, vol. 33, 247 - 257.

[7] Dinse, G. E. (1982), "Non-parametric estimation of partially incomplete time and type of failure data", *Biometrics*, vol. 38, 417 - 431.

[8] Kundu, D. (2004), "Parameter estimation for partially complete time and type of failure data", *Biometrical Journal*, vol. 46, 165 - 179.

[9] Kundu, D. (2008), "Bayesian inference and life testing plan for the Weibull distribution", *Technometrics*, vol. 50, 144 - 154.

[10] Kundu, D. and Pradhan, B. (2011), "Bayesian analysis of progressively censored competing risks data", *Sankhya, Ser. B*, vol.73, 276 - 296.

[11] Pena, E. A. and Gupta, A. K. (1990), "Bayes estimation for the Marshall-Olkin exponential distribution", *Journal of the Royal Statistical Society*, Ser. B, vol. 52, 379 - 389.

[12] Prentice, R. L., Kalbfleish, J. D., Peterson, Jr., A. V., Flurnoy, N., Farewell, V. T. and Breslow, N. E. (1978), "The analysis of failure times in presence of competing risks", *Biometrics*, vol. 34, 541 - 554.