Analyzing Competing Risks Data Using Bivariate Weibull-Geometric Distribution

Debasis Kundu *and Shuvashree Mondal †

Abstract

The motivation of this paper came from a study which was conducted to examine the effect of laser treatment in delaying the onset of blindness in patients with diabetic retinopathy. The data is a competing risks data with two dependent competing causes of failures, and there are ties. Recently the Marshall-Olkin bivariate Weibull (MOBW) distribution has been used to analyze this dependent competing risks data set. The bivariate Weibull-Geometric (BWG) distribution is a more flexible distribution than the MOBW distribution, and the MOBW can be obtained as a special case of the BWG distribution. The BWG has five parameters and it can be used quite effectively when there are ties in the data set. It is well known that the Bayesian inference has certain advantages over the classical inference in certain cases. In the classical inference most of the results are asymptotic in nature, whereas the Bayesian inference is valid even for small sample sizes also. In this paper first we develop the Bayesian inference of the unknown parameters of the BWG model, under a fairly flexible class of priors and analyze one real data set with ties to show the effectiveness of the model. Further, it is observed that the BWG can be used to analyze dependent competing risk data quite effectively when there are ties. The analysis of the above mentioned competing risks data set indicates that the BWG is preferred compared to the MOBW in this case.

Key Words and Phrases: Marshall-Olkin bivariate exponential distribution; Block and Basu bivariate distributions; Maximum likelihood estimators; EM algorithm; Competing risks.

AMS Subject Classifications: 62F10, 62F03, 62H12.

*Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500

†Department of Mathematics and Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad, Pin 826004

1 Introduction

The motivation of this paper came from a data set obtained from the Diabetic Retinopathy Study of the National Eye Institute, and the experiment was conducted to examine the effect of the laser treatment in delaying the onset of blindness in patients with diabetic retinopathy. For each patient, at the beginning of the experiment, one eye was selected for laser treatment and the other eye was not given any treatment. The minimum time to blindness (T) and along with it the indicator denoting whether the treated eye ($\delta = 1$) or the untreated eye ($\delta = 2$) has first failed been recorded. It may happen that both the eyes have failed simultaneously, and in that case it has been recorded as $\delta = 0$. The data set has been presented in Table 3.

The main objective of this study is to examine whether the laser treatment has any effect of delaying the onset of blindness to patients with diabetic retinopathy. The treatment or lack of treatment can be regarded as two causes of blindness (failures), hence the data set can be treated as a competing risks data. Clearly, the two competing causes cannot be taken as independent in this case. Moreover, there is a positive probability of simultaneous occurrence of both the causes. Recently Feizjavadian and Hashemi [5] used the Marshall-Olkin bivariate Weibull (MOBW) distribution to analyze this data set, see also Bai et al. [1], Zhang et al. [18], Shen and Xu [17] in this respect.

The Marshall-Olkin bivariate exponential (MOBE) distribution, see Marshall and Olkin [13] is the most popular bivariate distribution with singular components. It has exponential marginals and there is a positive probability that the marginals can be equal. It has an interesting physical representation in terms of random shocks, and due to this reason it is often called as the shock model also. Due to the presence of the singular component, it can be used quite effectively to analyze a bivariate data set when there are ties. One major problem about the MOBE distribution is that it can have only exponential marginals. Hence, the marginals can have only decreasing probability density functions (PDFs) and constant hazard rates. Because of this restriction, Marshall and Olkin in the same paper [13] suggested a more flexible bivariate distribution by replacing the exponential distribution with a Weibull distribution, and we call it as the Marshall-Olkin bivariate Weibull (MOBW) distribution. Clearly, the MOBW distribution is more flexible than the MOBE distribution, and it can be used quite effectively when there are ties in the data set and when the marginals may not have non-constant hazard functions, see for example Hanagal [7], Kundu and Dey [9], Lu [12] and the references cited therein.

Marshall and Olkin [14] provided a general method to introduce an extra parameter to a class of distributions and it brings more flexibility to the PDFs and hazard functions. They have illustrated their method with the exponential and Weibull class of distributions. From now on we call them as the univariate exponential-geometric (UEG) and the univariate Weibull-geometric (UWG) distribution, respectively. It is observed that the UEG and UWG are more flexible than the exponential and Weibull distributions, respectively. In the same paper, Marshall and Olkin also indicated how their method can be extended to the bivariate case, although the extension may not be unique. Following the method proposed by Marshall and Olkin [14], Kundu and Gupta [10] provided a bivariate extension of the MOBW distribution, and called it as the bivariate Weibull-Geometric (BWG) distribution.

The BWG distribution is more flexible than the MOBW distribution and it has five parameters. The MOBW and MOBE can be obtained as special cases of the BWG distribution. It is also a bivariate distribution with a singular component, and its marginals are UWG distributions. Since the BWG has a singular component, it can be used quite effectively when there are ties in the data set. The maximum likelihood estimators (MLEs) cannot be obtained in explicit forms, and they have to be derived by solving a five dimensional optimization problem. To avoid that computational complexity, Kundu and Gupta [10] proposed an Expectation Maximization (EM) algorithm to compute the MLEs of the unknown parameters, and it involves solving only one non-linear equation at each 'E'-step of the EM algorithm. It is observed by the authors that the proposed EM algorithm works very well in practice.

It is known that Bayesian inference has certain advantages in many cases. In this case it is observed that most of the results for the classical inference are asymptotic in nature, whereas in the Bayesian inference it is possible to obtain exact results for small sample sizes also. This is one of the major motivations to provide Bayesian inference in this case. In this paper first we develop the Bayesian inference of the unknown parameters of the BWG distribution. We have assumed a fairly flexible set of priors on the unknown parameters. It is observed that the Bayes estimates cannot be obtained in closed form. One needs to solve a five dimensional integration to compute the Bayes estimates. To avoid that we have used an EM algorithm to compute the posterior mode and that can be used as Bayes estimates of the unknown parameters. We have further used the importance sampling technique to construct credible intervals. One real data set has been analyzed to check how the proposed method works in practice. We have further used the BWG distribution which is more flexible than the MOBW distribution in analyzing dependent competing risks data when there are ties. We have analyzed the same data set based on the BWG distribution and came to the conclusion that it might be better to use the BWG distribution than the MOBW distribution in this case.

The rest of the paper is organized as follows. In Section 2, we provide the BWG distribution and discuss some basic properties. The prior assumptions are presented in Section 3. In Section 4, the posterior analysis has been performed and the applications have been presented in Section 5. Some simulation results have been provided in Section 6, and finally we conclude the paper in Section 7.

2 Bivariate Weibull-Geometric Distribution

2.1 Preliminaries and Notations

We use the following notations in this paper. A two-parameter Weibull distribution with the shape parameter $\alpha > 0$ and the scale parameter $\lambda > 0$, with the PDF

$$f_{WE}(x;\alpha,\lambda) = \begin{cases} \alpha\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}} & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \le 0, \end{cases} \tag{1}$$

will be denoted by WE(α, λ). If the shape parameter $\alpha = 1$, it becomes an exponential distribution and we will denote it by EXP(λ). A geometric random variable N with the probability mass function (PMF)

$$P(N = n) = \theta(1-\theta)^{n-1}; \qquad n = 1, 2, \ldots, \tag{2}$$

for $0 < \theta \le 1$, will be denoted by GE(θ). Note that when $\theta = 1$, N is a degenerate random variable at 1. If F is any distribution function we will use $\bar{F}$ as its survival function.

Marshall and Olkin [14] used the following method to introduce an extra parameter to a class of distribution functions. Suppose $\{X_n; n = 1, 2, \ldots\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with the common cumulative distribution function (CDF) $F_X(x)$ and the PDF $f_X(x)$, then define a new random variable

$$Y = \min\{X_1, \ldots, X_N\},$$

where N is a GE(p) random variable, and N is independent of $X_1, X_2, \ldots$. The survival function of Y for $y > 0$, becomes

$$\bar{F}_Y(y) = P(Y > y) = \sum_{n=1}^{\infty} P(Y > y|N = n)P(N = n) = \frac{p\bar{F}_X(y)}{1-(1-p)\bar{F}_X(y)}. \tag{3}$$

Marshall and Olkin [14] considered in details two important classes of distribution functions namely (i) exponential and (ii) Weibull. When X_i follows ($\sim$) EXP(λ), then the survival function and the PDF of Y become

$$\bar{F}_{UEG}(y) = \frac{pe^{-\lambda y}}{1-(1-p)e^{-\lambda y}} \quad \text{and} \quad f_{UEG}(y) = \frac{p\lambda e^{-\lambda y}}{(1-(1-p)e^{-\lambda y})^2},$$

respectively. From now on, we call it as the univariate exponential-geometric (UEG), and it is denoted by UEG(λ, p). Note that EXP(λ) is a member of UEG(λ, p), when $p = 1$. Similarly, when $X_i \sim$ WE(α, λ), then the survival function and the PDF of Y become

$$\bar{F}_{UWG}(y) = \frac{pe^{-\lambda y^\alpha}}{1-(1-p)e^{-\lambda y^\alpha}} \quad \text{and} \quad f_{UWG}(y) = \frac{p\alpha\lambda y^{\alpha-1}e^{-\lambda y^\alpha}}{(1-(1-p)e^{-\lambda y^\alpha})^2},$$

respectively. From now on we call it as the univariate Weibull-geometric (UWG), and it is denoted by UWG(α, λ, p). In this case WE(α, λ), EXP(λ) and UEG(λ, p) are members of UWG(α, λ, p), with the appropriate choices of p and α. The PDF of the UWG is either a decreasing or an unimodal function, and the hazard function can be an increasing, decreasing or an unimodal function.

2.2 Bivariate Weibull-Geometric Distribution

A bivariate random vector (Y_1, Y_2) is said to have the Marshall-Olkin bivariate Weibull (MOBW) distribution if the joint survival function of Y_1, Y_2, for $y_1 > 0, y_2 > 0$, is

$$\bar{F}_{MOBW} = \begin{cases} e^{-(\lambda_0+\lambda_1)y_1^\alpha - \lambda_2 y_2^\alpha} & \text{if} \quad y_1 \geq y_2 > 0 \\ e^{-(\lambda_0+\lambda_2)y_2^\alpha - \lambda_1 y_1^\alpha} & \text{if} \quad y_2 > y_1 > 0, \end{cases} \tag{4}$$

and it will be denoted by MOBW($\alpha, \lambda_0, \lambda_1, \lambda_2$). If $(Y_1, Y_2) \sim$ MOBW($\alpha, \lambda_0, \lambda_1, \lambda_2$), then the joint PDF of Y_1, Y_2 becomes

$$f_{MOBW}(y_1, y_2) = \begin{cases} f_{WE}(y_1; \alpha, \lambda_0 + \lambda_1) f_{WE}(y_2; \alpha, \lambda_2) & \text{if} \quad y_1 > y_2 > 0, \\ f_{WE}(y_1; \alpha, \lambda_1) f_{WE}(y_2; \alpha, \lambda_0 + \lambda_2) & \text{if} \quad y_2 > y_1 > 0, \\ \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} f_{WE}(y; \alpha, \lambda_0 + \lambda_1 + \lambda_2) & \text{if} \quad y_1 = y_2 = y. \end{cases} \tag{5}$$

When we write the joint PDF (5) corresponds to the joint survival function (4) it is understood that the first two components of (5) are with respect to the two dimensional

Lebesgue measure and the third component is with respect to the one dimensional Lebesgue measure. Now we define the bivariate Weibull- Geometric (BWG) distribution as introduced by Kundu and Gupta [10]. Suppose $\{(X_{1n}, X_{2n}); n = 1, 2, \ldots\}$ is a sequence of i.i.d. MOBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$ random variables, and $N \sim$ GE(θ), and it is independent of $\{(X_{1n}, X_{2n}); n = 1, 2, \ldots\}$, then the random vector (Y_1, Y_2), where

$$Y_1 = \min\{X_{11}, \ldots, X_{1N}\} \quad \text{and} \quad Y_2 = \min\{X_{21}, \ldots, X_{2N}\},$$

is said to have the BWG distribution with parameters $\alpha, \lambda_0, \lambda_1, \lambda_2, \theta$ and will be denoted by BWG$(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)$. If $(Y_1, Y_2) \sim$ BWG$(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)$, then the joint survival function of (Y_1, Y_2) for $y_1 > 0, y_2 > 0$ is

$$\bar{F}_{BWG}(y_1, y_2) = P(Y_1 > y_1, Y_2 > y_2) = \begin{cases} \frac{\theta e^{-(\lambda_0+\lambda_1)y_1^\alpha - \lambda_2 y_2^\alpha}}{1-(1-\theta)e^{-(\lambda_0+\lambda_1)y_1^\alpha - \lambda_2 y_2^\alpha}} & \text{if} \quad y_1 \geq y_2 > 0 \\ \frac{\theta e^{-\lambda_1 y_1^\alpha - (\lambda_0+\lambda_2) y_2^\alpha}}{1-(1-\theta)e^{-\lambda_1 y_1^\alpha - (\lambda_0+\lambda_2) y_2^\alpha}} & \text{if} \quad y_2 > y_1 > 0, \end{cases} \tag{6}$$

and the joint PDF for $y_1 > 0, y_2 > 0$ becomes

$$f_{BWG}(y_1, y_2) = \begin{cases} f_1(y_1, y_2) & \text{if} \quad y_1 > y_2 > 0 \\ f_2(y_1, y_2) & \text{if} \quad y_2 > y_1 > 0 \\ f_0(y) & \text{if} \quad y_1 = y_2 = y, \end{cases} \tag{7}$$

where

$$\begin{aligned} f_1(y_1, y_2) &= \frac{\alpha^2 y_1^{\alpha-1} y_2^{\alpha-1} \theta(\lambda_0 + \lambda_1)\lambda_2 e^{-(\lambda_0+\lambda_1)y_1^\alpha - \lambda_2 y_2^\alpha}(1 + (1-\theta)e^{-(\lambda_0+\lambda_1)y_1^\alpha - \lambda_2 y_2^\alpha})}{(1-(1-\theta)e^{-(\lambda_0+\lambda_1)y_1^\alpha - \lambda_2 y_2^\alpha})^3} \\ f_2(y_1, y_2) &= \frac{\alpha^2 y_1^{\alpha-1} y_2^{\alpha-1} \theta(\lambda_0 + \lambda_2)\lambda_1 e^{-\lambda_1 y_1^\alpha - (\lambda_0+\lambda_2) y_2^\alpha}(1 + (1-\theta)e^{-\lambda_1 y_1^\alpha - (\lambda_0+\lambda_2) y_2^\alpha})}{(1-(1-\theta)e^{-\lambda_1 y_1^\alpha - (\lambda_0+\lambda_2) y_2^\alpha})^3} \\ f_0(y) &= \frac{\alpha y^{\alpha-1} \theta \lambda_0 e^{-(\lambda_0+\lambda_1+\lambda_2)y^\alpha}}{(\lambda_0 + \lambda_1 + \lambda_2)(1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2)y^\alpha})^2}. \end{aligned}$$

It has been shown by Kundu and Gupta [10] that the joint PDF of a BWG distribution can take variety of shapes.

The BWG distribution is a singular distribution with $P(Y_1 = Y_2) = \dfrac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} > 0$. Moreover, if $(Y_1, Y_2) \sim$ BWG$(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)$, then $Y_1 \sim$ UWG$(\alpha, \lambda_0 + \lambda_1, \theta)$ and $Y_2 \sim$

UWG($\alpha, \lambda_0+\lambda_2, \theta$). The survival copula associate with the BWG distribution can be written as

$$\widehat{C}(u,v) = \begin{cases} \widehat{C}_1(u,v) & \text{if} \quad \left(1-\theta+\frac{\theta}{u}\right)^{\lambda_0+\lambda_1} \geq \left(1-\theta+\frac{\theta}{v}\right)^{\lambda_0+\lambda_2} \\ \widehat{C}_2(u,v) & \text{if} \quad \left(1-\theta+\frac{\theta}{u}\right)^{\lambda_0+\lambda_1} < \left(1-\theta+\frac{\theta}{v}\right)^{\lambda_0+\lambda_2} \end{cases}, \tag{8}$$

where

$$\begin{aligned} \widehat{C}_1(u,v) &= \frac{\theta}{(1-\theta+\frac{\theta}{u})(1-\theta+\frac{\theta}{v})^{\lambda_2/(\lambda_0+\lambda_2)} - (1-\theta)} \\ \widehat{C}_2(u,v) &= \frac{\theta}{(1-\theta+\frac{\theta}{u})^{\lambda_1/(\lambda_0+\lambda_1)}(1-\theta+\frac{\theta}{v}) - (1-\theta)}. \end{aligned}$$

If $\lambda_1 = \lambda_2$ and if we denote $\dfrac{\lambda_1}{\lambda_0+\lambda_1} = \dfrac{\lambda_2}{\lambda_0+\lambda_2} = \gamma$, then (8) becomes

$$\widehat{C}(u,v) = \begin{cases} \widehat{C}_1(u,v) & \text{if} \quad u \leq v \\ \widehat{C}_2(u,v) & \text{if} \quad u > v, \end{cases} \tag{9}$$

where

$$\begin{aligned} \widehat{C}_1(u,v) &= \frac{\theta}{(1-\theta+\frac{\theta}{u})(1-\theta+\frac{\theta}{v})^{\gamma} - (1-\theta)} \\ \widehat{C}_2(u,v) &= \frac{\theta}{(1-\theta+\frac{\theta}{u})^{\gamma}(1-\theta+\frac{\theta}{v}) - (1-\theta)}. \end{aligned}$$

We call this copula (9) as the geometric Marshall-Olkin (GMO) copula. Note that when $\theta = 1$, the GMO copula becomes

$$\widehat{C}(u,v) = \begin{cases} uv^{\gamma} & \text{if} \quad u \leq v \\ u^{\gamma}v & \text{if} \quad u > v. \end{cases} \tag{10}$$

Note that, (10) is the well known Marshall-Olkin copula or Cuadras-Auge family of copulas, see for example Nelsen [16]. Using the copula, different dependency properties and dependency measures can be established.

We need the following results for further development. Suppose $(Y_1, Y_2) \sim$ BWG($\alpha, \lambda_0, \lambda_1, \lambda_2, \theta$) and N is same as defined before where $\mathbf{\Theta} = (\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)^\top$, then the conditional probability mass function of N given $Y_1 = y_1$ and $Y_2 = y_2$ becomes:

$$f_N(n|y_1,y_2,\mathbf{\Theta}) = \begin{cases} c_1(y_1,y_2,\mathbf{\Theta})(1-\theta)^{n-1}n^2 e^{-(n-1)(\lambda_0+\lambda_1)y_1^{\alpha}-(n-1)\lambda_2 y_2^{\alpha}} & \text{if} \quad y_2 < y_1 \\ c_2(y_1,y_2,\mathbf{\Theta})(1-\theta)^{n-1}n^2 e^{-(n-1)(\lambda_0+\lambda_2)y_2^{\alpha}-(n-1)\lambda_1 y_1^{\alpha}} & \text{if} \quad y_1 < y_2 \\ c_0(y,\mathbf{\Theta})(1-\theta)^{n-1} e^{-(n-1)(\lambda_0+\lambda_1+\lambda_2)y^{\alpha}} & \text{if} \quad y_1 = y_2 = y, \end{cases} \tag{11}$$

where

$$\begin{aligned}
c_1(y_1, y_2, \boldsymbol{\Theta}) &= \frac{(1-(1-\theta)e^{-(\lambda_0+\lambda_1)y_1^\alpha-\lambda_2 y_2^\alpha})^3}{(1+(1-\theta)e^{-(\lambda_0+\lambda_1)y_1^\alpha-\lambda_2 y_2^\alpha})} \\
c_2(y_1, y_2, \boldsymbol{\Theta}) &= \frac{(1-(1-\theta)e^{-\lambda_1 y_1^\alpha-(\lambda_0+\lambda_2)y_2^\alpha})^3}{(1+(1-\theta)e^{-\lambda_1 y_1^\alpha-(\lambda_0+\lambda_2)y_2^\alpha})} \\
c_0(y, \boldsymbol{\Theta}) &= (1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2)y^\alpha}).
\end{aligned}$$

Hence,

$$E(N|y_1, y_2, \boldsymbol{\Theta}) = \begin{cases} \frac{(1-\psi_1(y_1,y_2,\boldsymbol{\Theta}))^2-6(1-\psi_1(y_1,y_2,\boldsymbol{\Theta}))+6}{(1-\psi_1(y_1,y_2,\boldsymbol{\Theta}))^3} & \text{if} \quad y_2 < y_1 \\ \frac{(1-\psi_2(y_1,y_2,\boldsymbol{\Theta}))^2-6(1-\psi_2(y_1,y_2,\boldsymbol{\Theta}))+6}{(1-\psi_2(y_1,y_2,\boldsymbol{\Theta}))^3} & \text{if} \quad y_1 < y_2 \\ \frac{1}{c_0(y,\boldsymbol{\Theta})} & \text{if} \quad y_1 = y_2 = y, \end{cases} \tag{12}$$

where

$$\begin{aligned}
\psi_1(y_1, y_2, \boldsymbol{\Theta}) &= (1-\theta)e^{-(\lambda_0+\lambda_1)y_1^\alpha-\lambda_2 y_2^\alpha} \\
\psi_2(y_1, y_2, \boldsymbol{\Theta}) &= (1-\theta)e^{-(\lambda_0+\lambda_2)y_2^\alpha-\lambda_1 y_1^\alpha}.
\end{aligned}$$

3 Prior Assumptions and Available Data

3.1 Prior Assumption

We use the following notations from now on. A beta distribution with parameter $\alpha_0 > 0$ and $\beta_0 > 0$ will be denoted by Beta(α_0, β_0). A gamma distribution with the shape parameter α_0, the scale parameter λ_0 and the mean (α_0/λ_0) will be denoted by Gamma(α_0, λ_0). We make the following prior assumptions on the unknown parameters.

$$\lambda_0 \sim \text{Gamma}(a_0, b_0), \tag{13}$$

$$\lambda_1 \sim \text{Gamma}(a_1, b_1), \tag{14}$$

$$\lambda_2 \sim \text{Gamma}(a_2, b_2), \tag{15}$$

$$\theta \sim \text{Beta}(a_3, b_3), \tag{16}$$

$$\alpha \sim \text{Gamma}(a_4, b_4), \tag{17}$$

and they are independently distributed. It should be mentioned that it is possible to assume a more general prior on α rather than (17). We can assume that it has support on $(0, \infty)$ and its prior PDF is log-concave in nature. This kind of assumption is quite common in the statistical literature, see for example Berger and Sun [2] or Kundu [8]. All our analyses will go through even with this general prior also.

3.2 Available Data

Bivariate Data: It is assumed that the available bivariate data set is of the following form:

$$\mathcal{D}_1 = \{(y_{11}, y_{21}), \ldots, (y_{1n}, y_{2n})\}. \tag{18}$$

In this paper we will be using the following notations:

$$I_0 = \{i : y_{1i} = y_{2i}\}, \quad I_1 = \{i : y_{1i} > y_{2i}\}, \quad I_2 = \{i : y_{1i} < y_{2i}\}, \quad I = \{1, \ldots, n\},$$

$|I_0| = k_0$, $|I_1| = k_1$, $|I_2| = k_2$, here $|I_j|$ denotes the number of elements in set I_j, for $j = 0, 1, 2$. It is assumed that $k_0 > 0$, $k_1 > 0$ and $k_2 > 0$.

Competing Risk Data: It is assumed that the available competing risks data set is of the following form:

$$\mathcal{D}_2 = \{(t_1, \delta_1), \ldots, (t_n, \delta_n)\}. \tag{19}$$

Here t_i denotes the time of failure and δ_i denotes the cause of failure. In this paper, it is assumed that there are only two causes of failures, say Cause 1 and Cause 2. When $\delta_i = 1$, it indicates that the item has failed due to Cause 1, similarly, when $\delta_i = 2$, it means the item has failed due to Cause 2. When $\delta_i = 0$, it means that the item has failed at t_i due to both the causes. In this paper we will be using the following notations:

$$J_0 = \{i : \delta_i = 0\}, \quad J_1 = \{i : \delta_i = 1\}, \quad J_2 = \{i : \delta_i = 2\},$$

$|J_0| = m_0$, $|J_1| = m_1$, $|J_2| = m_2$. It is assumed that $m_0 > 0$, $m_1 > 0$ and $m_2 > 0$.

4 Posterior Analysis

In this section we provide the posterior analyses for the bivariate data set as well as for the competing risks data set.

4.1 Bivariate Data

The joint likelihood function of the observed data can be written as

$$L(\mathcal{D}_1|\mathbf{\Theta}) = \left\{\prod_{i\in I_1} f_1(y_{1i}, y_{2i})\right\}\left\{\prod_{i\in I_2} f_2(y_{1i}, y_{2i})\right\}\left\{\prod_{i\in I_0} f_o(y_i)\right\}, \tag{20}$$

here $\mathbf{\Theta} = (\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)$. Hence, based on the likelihood function (20) and the priors (13) to (17), the joint posterior density function can be written as

$$\pi(\mathbf{\Theta}|\mathcal{D}_1) \propto g_1(\alpha|\mathcal{D}_1)g_2(\theta|\mathcal{D}_1)g_3(\alpha,\lambda_0|\mathcal{D}_1)g_4(\alpha,\lambda_1|\mathcal{D}_1)g_5(\alpha,\lambda_2|\mathcal{D}_1)g_6(\mathbf{\Theta}|\mathcal{D}_1), \tag{21}$$

where

$$g_1(\alpha|\mathcal{D}_1) \propto \alpha^{2k_1+2k_2+k_0+a_4-1}e^{-\alpha(b_4-\sum_{i\in I_1\cup I_2}\ln(y_{1i}y_{2i})-\sum_{i\in I_0}\ln y_i)} \tag{22}$$

$$g_2(\theta|\mathcal{D}_1) \propto \theta^{n+a_3-1}(1-\theta)^{b_3-1} \tag{23}$$

$$g_3(\alpha,\lambda_0|\mathcal{D}_1) \propto \lambda_0^{k_0+a_0-1}e^{-\lambda_0(b_0+\sum_{i\in I_1}y_{1i}^{\alpha}+\sum_{i\in I_2}y_{2i}^{\alpha}+\sum_{i\in I_0}y_i^{\alpha})}, \tag{24}$$

$$g_4(\alpha,\lambda_1|\mathcal{D}_1) \propto \lambda_1^{k_2+a_1-1}e^{-\lambda_1(b_1+\sum_{i\in I_1\cup I_2}y_{1i}^{\alpha}+\sum_{i\in I_0}y_i^{\alpha})}, \tag{25}$$

$$g_5(\alpha,\lambda_2|\mathcal{D}_1) \propto \lambda_2^{k_1+a_2-1}e^{-\lambda_2(b_2+\sum_{i\in I_1\cup i_2}y_{2i}^{\alpha}+\sum_{i\in I_0}y_i^{\alpha})}, \tag{26}$$

$$g_6(\mathbf{\Theta}|\mathcal{D}_1) = \frac{(\lambda_0+\lambda_1)^{k_1}(\lambda_0+\lambda_2)^{k_2}}{(\lambda_0+\lambda_1+\lambda_2)^{k_0}}\prod_{i\in I_1}\frac{(1+(1-\theta)e^{-(\lambda_0+\lambda_1)y_{1i}^{\alpha}-\lambda_2 y_{2i}^{\alpha}})}{(1-(1-\theta)e^{-(\lambda_0+\lambda_1)y_{1i}^{\alpha}-\lambda_2 y_{2i}^{\alpha}})^3}$$
$$\prod_{i\in I_2}\frac{(1+(1-\theta)e^{-\lambda_1 y_{1i}^{\alpha}-(\lambda_0+\lambda_2)y_{2i}^{\alpha}})}{(1-(1-\theta)e^{-\lambda_1 y_{1i}^{\alpha}-(\lambda_0+\lambda_2)y_{2i}^{\alpha}})^3}\prod_{i\in I_0}\frac{1}{(1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2)y_i^{\alpha}})^2}. \tag{27}$$

It is clear that the joint posterior density function is not in a standard form, and the Bayes estimates cannot be obtained in closed form.

It should be mentioned that the posterior mode plays an important role for any posterior analysis. First we will obtain the posterior mode of $\pi(\mathbf{\Theta}|\mathcal{D}_1)$ which can be used to provide Bayes estimates of the unknown parameters with respect to the 0-1 loss function. To compute the posterior mode we have used the EM algorithm technique as suggested by Ghosh, Delampady and Samanta [6]. It should be pointed out that Kundu and Gupta [10] suggested a very efficient EM algorithm to compute the maximum likelihood estimators of the unknown parameters based on the bivariate data from a BWG distribution. We have closely followed the method suggested by Kundu and Gupta [10], but we have used it in case of the posterior distribution.

Here also the basic idea remains same. To compute the posterior mode, we treat this as a missing value problem, and in this case the missing values are the observed value of N at each data point. Therefore, the 'complete' data $\mathcal{D}_c$ will be of the form

$$\mathcal{D}_c = \{(y_{11}, y_{21}, n_1), \ldots, (y_{1n}, y_{2n}, n_n)\}, \tag{28}$$

here n_i is the missing value of N for the data point (y_{1i}, y_{2i}), for $i = 1, \ldots, n$. For brevity, we write $\mathcal{D}_c = (\boldsymbol{y}, \boldsymbol{z})$, where $\boldsymbol{y}$ denotes the observed data $\mathcal{D}_1$ and $\boldsymbol{z} = (n_1, \ldots, n_n)$ denotes the missing data. We use the following procedure. For fixed θ, by using the EM algorithm we obtain $\widehat{\lambda}_0(\theta)$, $\widehat{\lambda}_1(\theta)$, $\widehat{\lambda}_2(\theta)$, $\widehat{\alpha}(\theta)$, which maximizes $\pi(\mathbf{\Theta}|\boldsymbol{y})$, for all $\alpha > 0$, $\lambda_0 > 0$, $\lambda_1 > 0$, $\lambda_2 > 0$, i.e.

$$\pi(\widehat{\alpha}(\theta), \widehat{\lambda}_0(\theta), \widehat{\lambda}_1(\theta), \widehat{\lambda}_2(\theta), \theta|\boldsymbol{y}) \geq \pi(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta).$$

Then find $\widehat{\theta}$, so that

$$\pi(\widehat{\alpha}(\widehat{\theta}), \widehat{\lambda}_0(\widehat{\theta}), \widehat{\lambda}_1(\widehat{\theta}), \widehat{\lambda}_2(\widehat{\theta}), \theta|\boldsymbol{y}) \geq \pi(\widehat{\alpha}(\theta), \widehat{\lambda}_0(\theta), \widehat{\lambda}_1(\theta), \widehat{\lambda}_2(\theta), \theta|\boldsymbol{y}).$$

Note that since $\pi(\mathbf{\Theta}|\boldsymbol{y}) = \pi(\mathbf{\Theta}, \boldsymbol{z}|\boldsymbol{y})/p(\boldsymbol{z}|\boldsymbol{y}, \mathbf{\Theta})$, we have

$$\ln \pi(\mathbf{\Theta}|\boldsymbol{y}) = \ln \pi(\mathbf{\Theta}, \boldsymbol{z}|\boldsymbol{y}) - \ln p(\boldsymbol{z}|\boldsymbol{y}, \mathbf{\Theta}). \tag{29}$$

We use the following notation now $\boldsymbol{\gamma} = (\alpha, \lambda_0, \lambda_1, \lambda_2)^\top$, $\boldsymbol{\gamma}^{(k)} = (\alpha^{(k)}, \lambda_0^{(k)}, \lambda_1^{(k)}, \lambda_2^{(k)})^\top$ is the estimate of the parameter vector $\boldsymbol{\gamma}$ at the k-th stage of the EM algorithm for fixed θ and $\boldsymbol{\Theta}^{(k)} = (\boldsymbol{\gamma}^{(k)}, \theta)$. It may be mentioned that $\alpha^{(k)}, \lambda_0^{(k)}, \lambda_1^{(k)}, \lambda_2^{(k)}$ depend on θ, but we do not explicitly mention it for brevity.

Taking expectation with respect to $\boldsymbol{Z}|\boldsymbol{\Theta}^{(k)}$ on both sides of (29), we get

$$\begin{aligned}
\ln \pi(\boldsymbol{\Theta}|\boldsymbol{y}) &= \sum_{\boldsymbol{z}} \ln \pi(\boldsymbol{\Theta}, \boldsymbol{z}|\boldsymbol{y}) p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\Theta}^{(k)}) - \sum_{\boldsymbol{z}} \ln p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\Theta}) p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\Theta}^{(k)}) \\
&= Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(k)}) - H(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(k)}).
\end{aligned}$$

Therefore, $\boldsymbol{\Theta}^{(k+1)}$ can be obtained by maximizing $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(k)})$ with respect to α, λ_0, λ_1 and λ_2, see for example Ghosh, Delampady and Samanta [6]. It can be easily shown that in that case

$$\ln \pi(\boldsymbol{\Theta}^{(k+1)}|\boldsymbol{y}) \geq \ln \pi(\boldsymbol{\Theta}^{(k)}|\boldsymbol{y}).$$

In the Appendix, we provide the explicit expression of $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(k)})$ and will show how it can be maximized by solving one non-linear equation only. Note that the posterior mode can be used as a Bayes estimate of $\boldsymbol{\Theta}$.

Now we will describe how we can obtain the Bayes estimator of any function say h of $\boldsymbol{\Theta}$ and to obtain the associate credible interval based on the importance sampling technique. The procedure is quite simple. From (22) to (26), we rewrite as follows:

$$\pi_1(\theta|\mathcal{D}_1) \sim \text{Beta}(n + a_3, b_3) \tag{30}$$

$$\pi_2(\lambda_0|\mathcal{D}_1, \alpha) \sim \text{Gamma}\left(k_0 + a_0, b_0 + \sum_{i \in I_1} y_{1i}^\alpha + \sum_{i \in I_2} y_{2i}^\alpha + \sum_{i \in I_0} y_i^\alpha\right) \tag{31}$$

$$\pi_3(\lambda_1|\mathcal{D}_1, \alpha) \sim \text{Gamma}\left(k_2 + a_1, b_1 + \sum_{i \in I_1 \cup I_2} y_{1i}^\alpha + \sum_{i \in I_0} y_i^\alpha\right), \tag{32}$$

$$\pi_4(\lambda_2|\mathcal{D}_1, \alpha) \sim \text{Gamma}\left(k_1 + a_2, b_2 + \sum_{i \in I_1 \cup I_2} y_{2i}^\alpha + \sum_{i \in I_0} y_i^\alpha\right), \tag{33}$$

$$\pi_5(\alpha|\mathcal{D}_1) = kv(\alpha)\alpha^{2k_1+2k_2+k_0+a_4-1} e^{-\alpha(b_4 - \sum_{i \in I_1 \cup I_2} \ln(y_{1i} y_{2i}) - \sum_{i \in I_0} \ln y_i)}, \tag{34}$$

here

$$\begin{aligned} v(\alpha) &= \left(b_0 + \sum_{i\in I_1} y_{1i}^{\alpha} + \sum_{i\in I_2} y_{2i}^{\alpha} + \sum_{i\in I_0} y_{i}^{\alpha}\right)^{-(k_0+a_0)} \times \left(b_1 + \sum_{i\in I_1\cup I_2} y_{1i}^{\alpha} + \sum_{i\in I_0} y_{i}^{\alpha}\right)^{-(k_2+a_1)} \times \\ & \left(b_2 + \sum_{i\in I_1\cup i_2} y_{2i}^{\alpha} + \sum_{i\in I_0} y_{i}^{\alpha}\right)^{-(k_1+a_2)}, \end{aligned}$$

and k is the normalizing constant. Finally we take $g(\mathbf{\Theta}) = g_6(\mathbf{\Theta}|\mathcal{D}_1)$. We need the following result for further development.

Theorem 1: $\pi_5(\alpha|\mathcal{D}_1)$ is log-concave.

Proof: It can be obtained similarly as the proof of Theorem 2 of Kundu [8]. The details are avoided. ∎

Since $\pi_5(\alpha|\mathcal{D}_1)$ is a log-concave function, it is very simple to generate samples from $\pi_5(\alpha|\mathcal{D}_1)$ by using the method of Devroye [4]. Alternatively, we can use the method of Kundu [8] also to generate samples from $\pi_5(\alpha|\mathcal{D}_1)$. Therefore, the following algorithm can be used to compute the Bayes estimate of $h(\mathbf{\Theta})$ and the associated highest posterior density (HPD) credible interval also.

Algorithm 1:

Step 1: Generate $\alpha_1, \ldots, \alpha_M$ from $\pi_5(\alpha|\mathcal{D}_1)$ using the method of Devroye [4] or Kundu [8].

Step 2: For a given α_m, generate

$$\begin{aligned} \lambda_{0m} &\sim \text{Gamma}\left(k_0 + a_0, b_0 + \sum_{i\in I_1} y_{1i}^{\alpha_m} + \sum_{i\in I_2} y_{2i}^{\alpha_m} + \sum_{i\in I_0} y_{i}^{\alpha_m}\right) \\ \lambda_{1m} &\sim \text{Gamma}\left(k_2 + a_1, b_1 + \sum_{i\in I_1\cup I_2} y_{1i}^{\alpha_m} + \sum_{i\in I_0} y_{i}^{\alpha_m}\right) \\ \lambda_{2m} &\sim \text{Gamma}\left(k_1 + a_2, b_2 + \sum_{i\in I_1\cup I_2} y_{2i}^{\alpha_m} + \sum_{i\in I_0} y_{i}^{\alpha_m}\right), \end{aligned}$$

for $m = 1, \ldots, M$.

Step 3: Generate $\theta_1, \ldots, \theta_M$ from Beta$(n + a_3, b_3)$.

Step 4: Compute $h_m = h(\alpha_m, \lambda_{0m}, \lambda_{1m}, \lambda_{2m}, \theta_m)$, for $i = 1, \ldots, M$.

Step 5: Calculate

$$w_m = \frac{g(\alpha_m, \lambda_{0m}, \lambda_{1m}, \lambda_{2m}, \theta_m)}{\sum_{i=1}^{M} g(\alpha_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \theta_i)}.$$

Step 6: Compute the Bayes estimate of $h(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)$ under the squared error loss function as

$$\widehat{h}(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta) = \sum_{m=1}^{M} w_m h_m.$$

Step 7: To construct a $100(1-\gamma)\%$ $(0 < \gamma < 1)$ credible interval of $h(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)$, first order h_m's for $m = 1, \ldots, M$, say $h_{(1)} < \ldots < h_{(M)}$, and arrange w_m's accordingly to get $w_{[1]}, \ldots, w_{[M]}$. Note that $w_{[1]}, \ldots, w_{[M]}$ may not be ordered.

Step 8: A $100(1-\gamma)\%$ credible interval (CRI) can be obtained as (h_{j_1}, h_{j_2}), where j_1 and j_2 satisfy, $j_1, j_2 \in \{1, \ldots, M\}$, $j_1 < j_2$, and

$$\sum_{i=j_1}^{j_2} w_{[i]} \leq 1 - \gamma < \sum_{i=j_1}^{j_2+1} w_{[i]}. \tag{35}$$

The $100(1-\gamma)\%$ highest posterior density (HPD) CRI can be obtained as $(h_{j_1^*}, h_{j_2^*})$, where j_1^* and j_2^* satisfy, $j_1^*, j_2^* \in \{1, \ldots, M\}$, $j_1^* < j_2^*$,

$$\sum_{i=j_1^*}^{j_2^*} w_{[i]} \leq 1 - \gamma < \sum_{i=j_1^*}^{j_2^*+1} w_{[i]} \quad \text{and} \quad h_{j_2^*} - h_{j_1^*} \leq h_{j_2} - h_{j_1}.$$

where j_1, j_2 satisfy (35).

4.2 Competing Risk Data

In this section we develop the posterior analysis of the BWG model for competing risks data as provided in (19). There are several ways to model competing risks data, and we have used the latent failure type model assumption of Cox [3] and it can defined as given below for two competing causes of failures. Suppose Y_1 and Y_2 are the lifetimes of two competing causes of failures, then the observed failure time (T) and the cause of failure (Δ) is defined as follows

$$T = \min\{Y_1, Y_2\}, \qquad \Delta = \begin{cases} 1 & \text{if} \quad Y_1 < Y_2 \\ 2 & \text{if} \quad Y_1 > Y_2 \\ 0 & \text{if} \quad Y_1 = Y_2. \end{cases}$$

In most of the existing literature it has been assumed that lifetimes of the competing causes are independently distributed. Here it is assumed that Y_1 and Y_2 are dependent and $(Y_1, Y_2) \sim \text{BWG}(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)$. We need the following results for further development.

THEOREM 2: Suppose $(Y_1, Y_2) \sim \text{BWG}(\alpha, \lambda_0, \lambda_1, \lambda_2, \theta)$, then

$$(a) \quad P(y_1 \le Y_1 < y_1 + dy_1, Y_2 > y_1) = \frac{\theta\alpha\lambda_1 y_1^{\alpha-1} e^{-(\lambda_0+\lambda_1+\lambda_2)y_1^\alpha}}{(1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2)y_1^\alpha})^2} dy_1.$$

$$(b) \quad P(Y_1 > y_2, y_2 \le Y_2 < y_2 + dy_2) = \frac{\theta\alpha\lambda_2 y_1^{\alpha-1} e^{-(\lambda_0+\lambda_1+\lambda_2)y_2^\alpha}}{(1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2)y_2^\alpha})^2} dy_2.$$

PROOF: See in the Appendix. ∎

Based on Theorem 2, the joint likelihood function becomes:

$$L(\mathcal{D}_2|\mathbf{\Theta}) = \left(\frac{\alpha^n\theta^n\lambda_0^{m_0}\lambda_1^{m_1}\lambda_2^{m_2}}{(\lambda_0+\lambda_1+\lambda_2)^{m_0}}\right)\left\{\prod_{i=1}^{n} \frac{t_i^{\alpha-1} e^{-(\lambda_0+\lambda_1+\lambda_2)t_i^\alpha}}{(1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2)t_i^\alpha})^2}\right\}.$$

It is assumed that we have the competing risks data as defined in (19). Therefore, based on the priors (13) to (17), the joint posterior of $\mathbf{\Theta}|\mathcal{D}_2$ becomes

$$\pi(\mathbf{\Theta}|\mathcal{D}_2) \propto \pi_1(\theta|\mathcal{D}_1)\pi_2(\lambda_0|\mathcal{D}_1,\alpha)\pi_3(\lambda_1|\mathcal{D}_2,\alpha)\pi_4(\lambda_2|\mathcal{D}_1,\alpha)\pi_5(\alpha|\mathcal{D}_1)g(\mathbf{\Theta}|\mathcal{D}_2), \qquad (36)$$

here

$$\begin{aligned} \pi_1(\theta|\mathcal{D}_2) &\sim \text{Beta}(n+a_3, b_3), \\ \pi_2(\lambda_0|\mathcal{D}_2,\alpha) &\sim \text{Gamma}\left(m_0+a_0, b_0+\sum_{i=1}^{n} t_i^\alpha\right), \end{aligned}$$

$$\begin{aligned}
\pi_3(\lambda_1|\mathcal{D}_2,\alpha) &\sim \text{Gamma}\left(m_1+a_1, b_1+\sum_{i=1}^n t_i^{\alpha}\right),\\
\pi_4(\lambda_2|\mathcal{D}_2,\alpha) &\sim \text{Gamma}\left(m_2+a_2, b_2+\sum_{i=1}^n t_i^{\alpha}\right),\\
\pi_5(\alpha|\mathcal{D}_2) &\propto u(\alpha)\alpha^{n+a_4-1}e^{-\alpha(b_4-\sum_{i=1}^n \ln t_i)},\\
g(\boldsymbol{\Theta}|\mathcal{D}_2) &= \frac{1}{(\lambda_0+\lambda_1+\lambda_2)^{m_0}} \times \frac{1}{\prod_{i=1}^n (1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2)t_i^{\alpha}})^2},
\end{aligned}$$

and

$$u(\alpha) = \left(b_0+\sum_{i=1}^n t_i^{\alpha}\right)^{-(m_0+a_0)} \times \left(b_1+\sum_{i=1}^n t_i^{\alpha}\right)^{-(m_1+a_1)} \times \left(b_2+\sum_{i=1}^n t_i^{\alpha}\right)^{-(m_2+a_2)}.$$

We need the following result for further development.

Theorem 3: $\pi_5(\alpha|\mathcal{D}_2)$ is log-concave.

Proof: It follows along the same line as the proof of Theorem 1. ∎

Therefore, using the method of Devroye [4] or by Kundu [8] it is possible to generate samples from $\pi_5(\alpha|\mathcal{D}_2)$. Hence, it is possible to use a similar algorithm as Algorithm 1 to compute Bayes estimate of any function of the parameters and to construct credible and HPD credible intervals. The following algorithm can be used for that purposes.

Algorithm 2:

Step 1: Generate $\alpha_1, \ldots, \alpha_M$ from $\pi_5(\alpha|\mathcal{D}_2)$ using the method of Devroye [4] or Kundu [8].

Step 2: For a given α_m, generate

$$\begin{aligned}
\lambda_{0m} &\sim \text{Gamma}\left(m_0+a_0, b_0+\sum_{i=1}^n t_i^{\alpha_m}\right),\\
\lambda_{1m} &\sim \text{Gamma}\left(m_1+a_1, b_1+\sum_{i=1}^n t_i^{\alpha_m}\right),\\
\lambda_{2m} &\sim \text{Gamma}\left(m_2+a_2, b_2+\sum_{i=1}^n t_i^{\alpha_m}\right),
\end{aligned}$$

for $m = 1, \ldots, M$.

Step 3 to Step 8: Exactly same as Algorithm 1.

5 Applications

In this section we provide the analyses of two data sets based on the BWG model

5.1 Soccer Data

This data set represents the soccer score data, where at least one goal has been scored by the home team and at least one goal has been scored directly from a penalty kick, foul kick or any other direct kicks (together they will be called as 'kick goal') by one of the teams. The data set has been originally reported in Meintanis [15]. Here Y_1 represents the time in minutes of the first 'kick' goal scored by any team and Y_2 represents the time in minutes the first goal of any type scored by the home team. Therefore, $Y_1 = 26$ and $Y_2 = 20$ means that the home team scored its first goal at 20th minutes, and it was not a kick goal, where the first kick goal by one of the teams took place at the 26th minutes. Similarly, $Y_1 = Y_2 = 8$ means that the home team scored its first goal at the 8th minute and it was the first kick goal also. Clearly, it is a bivariate data set with ties and it occurs with positive probability. Due to computational issue, we divide all the data values by 100, and it is not going to make any difference in the inference. We have used the BWG distribution to analyze this data set.

Kundu and Gupta [10] analyzed this same data set in the classical framework and obtained the maximum likelihood estimates of the unknown parameters as: $\widehat{\theta} = 0.630$, $\widehat{\alpha} = 1.722$, $\widehat{\lambda}_0 = 1.301$, $\widehat{\lambda}_1 = 0.875$ and $\widehat{\lambda}_2 = 2.044$. We have obtained the Bayes estimates of the parameters based on 0-1 loss function through EM algorithm with the initial guesses of $\alpha, \lambda_0, \lambda_1, \lambda_2$ as $1.722, 1.301, 0.875, 2.044$, respectively. In the prior assumptions, the hyper-parameters have been chosen in such a way that the prior expectations are proportional to the maximum likelihood estimates of the parameters and set $a_0 = 10$, $b_0 = 7.69$, $a_1 = 10$, $b_1 = 12.50$, $a_2 = 10$, $b_2 = 5.00$, $a_3 = 10$, $b_3 = 6.66$, $a_4 = 10$, $b_4 = 5.88$. In Table 2, we have computed the Bayes estimates based on 0-1 loss function, i.e. the posterior modes, the Bayes estimates based on squared error loss function and 90% HPD credible intervals.

2005-2006	Y_1	Y_2	2004-2005	Y_1	Y_2
Lyon-RealMadrid	26	20	Internazionale-Bremen	34	34
Milan-Fenerbahce	63	18	RealMadrid-Roma	53	39
Chelsea-Anderlecht	19	19	Man.United-Fenerbahce	54	7
ClubBrugge-Juventus	66	85	Bayern-Ajax	51	28
Fenerbahce-PSV	40	40	Moscow-PSG7	66	4
Internazionale-Rangers	49	49	Barcelona-Shakhtar	64	15
Panathinaikos-Bremen	8	8	Leverkusen-Roma	26	48
Ajax-Arsenal	69	71	Arsenal-Panathinaikos	16	16
Man.United-BenØca	39	39	DynamoKyiv-RealMadrid	44	13
RealMadrid-Rosenborg	82	48	Man.United-Sparta	25	14
Villarreal-BenØca	72	72	Bayern-M.TelAviv	55	11
Juventus-Bayern	66	62	Bremen-Internazionale	49	49
ClubBrugge-Rapid	25	9	Anderlecht-Valencia	24	24
Olympiacos-Lyon	41	3	Panathinaikos-PSV	44	30
Internazionale-Porto	16	75	Arsenal-Rosenborg	42	3
Schalke-PSV	18	18	Liverpool-Olympiacos	27	47
Barcelona-Bremen	22	14	M. Tel-Aviv-Juventus	28	28
Milan-Schalke	42	42	Bremen-Panathinaikos	2	2
Rapid-Juventus	36	52			

Table 1: UEFA Champion's League data

5.2 Diabetic Retinopathy Data

A real data set in presence to two competing risks has been analyzed to see the effectiveness of the proposed model. The data set as it has been mentioned before, came from a Diabetic Retinopathy Study conducted by National Eye Institute to estimate the effect of laser treatment in delaying the onset of blindness with diabetic Retinopathy. In the onset of

parameter	Posterior mode	Bayes estimator (based on squared error loss)	HPD CRI (90%)
θ	0.699	0.720	(0.661, 0.804)
α	1.701	1.263	(1.162, 1.367)
λ_0	1.260	1.235	(0.926, 1.448)
λ_1	0.885	0.633	(0.385, 0.938)
λ_2	2.928	1.563	(1.243, 2.062)

Table 2: Bayes estimators based on UEFA Champion's League data

the experiment, whether the effect of laser on delaying the time to blindness of the patients suffering from diabetes has any effect or not is not evident, so treatment or lack of treatment can be regarded as the causes of blindness. Hence, this can be treated as a competing risks data. Clearly, time to blindness of two eyes cannot be treated independent, and there are some ties also, hence we try to use the BWG model in this case.

For computational convenience, we have divided the failure times by 10^3. It is not going to make any difference in the inference procedure. In a BWG distribution, minimum of the two marginals follows a UWG distribution. To check if the failure time in presence of two competing risks follows a UWG distribution (say UWG$(\alpha, \lambda, \theta)$), we have conducted Kolmogorov-Smirnov (K-S) test. The maximum likelihood estimates of α, λ, θ are $1.949, 0.311, 1.286$, respectively. The K-S distance between the empirical and the fitted distribution is 0.049 and the associated p value is 0.995. Therefore, it is not irrational to assume that the failure time due to two competing causes follows a BWG distribution.

The hyper-parameters have been chosen in such a way that the prior mean of α is proportional to 1.9, prior mean of θ is proportional to 0.3 and prior means of λ_is are proportional to $1.2/3 = 0.4$ for $i = 0, 1, 2$. The values of the hyper-parameters have been set as $a_0 = 1, b_0 = 2.50, a_1 = 1, b_1 = 2.50, a_2 = 1, b_2 = 2.50, a_3 = 1, b_3 = 2.33, a_4 = 1, b_4 = 0.52$. The Bayes estimates based on the squared error loss function and 90% HPD credible intervals have been recored in Table 4. Based on the credible interval of θ, we conclude that the BWG is preferred to MOBW in this case.

T	Δ	T	Δ	T	Δ
266	1	272	0	203	0
91	2	1137	0	84	1
154	2	1484	1	392	1
285	0	315	1	1140	2
583	1	287	2	901	1
547	2	1252	1	1247	0
79	1	717	2	448	2
622	0	642	1	904	2
707	2	141	2	276	1
469	2	407	1	520	1
93	1	356	1	485	2
1313	2	1653	0	248	2
805	1	427	2	503	1
344	1	699	1	423	2
790	2	36	2	285	2
125	2	667	1	315	2
777	2	588	2	727	2
306	1	471	0	210	2
415	1	126	1	409	2
307	2	350	2	584	1
637	2	350	1	355	1
577	2	663	0	1302	1
178	1	567	2	227	2
517	2	966	0		

Table 3: Minimum time to blindness in days and its causes

6 Simulation Results

In this section we have presented some simulation results mainly to show how the proposed algorithms work in practice. The main purpose of these simulation results is to see how the hyper-parameters affect the parameters estimates in both the cases considered. Since the Bayesian inferences are valid for any finite sample sizes, we have kept the sample size to be fixed. We have changed the hyper-parameters in such a manner so that prior means are

parameter	Bayes estimator (based on squared error loss)	HPD CRI (90%)
θ	0.882	(0.847, 0.925)
α	1.557	(1.355, 1.684)
λ_0	0.260	(0.171, 0.345)
λ_1	0.709	(0.552, 0.889)
λ_2	0.824	(0.659, 1.067)

Table 4: Bayes estimators based on competing risks data

equal to the true parameter values, but the prior variances decrease. We have obtained the Bayes estimates based on the posterior mode and posterior mean in case of the bivariate data. We have obtained the average estimates and associated mean squared errors (MSEs) in each case. All the results are based on 1000 replications. The results are reported in Tables 5 and 6. We have further computed the Bayes estimates based on the posterior mean in case of the competing risks data and the results are reported in Table 7. In this case also the results are based on 1000 replications. In all these cases it is clear that the proposed algorithms are working quite well. In all the cases, as the prior variances decrease the biases and MSEs decrease, as expected.

7 Conclusions

In this paper, we have developed the Bayesian inference of the BGW distribution. Since, BGW distribution has singular component, it can be used quite effectively in data analysis, where there are ties in data. The posterior mode has been derived through applying a EM algorithm and this can be interpreted as the Bayes estimator based on 0-1 loss function. Also we have obtained the Bayes estimator based on the squared error loss function and the associated credible intervals. As an application, the BWG model has been employed under dependent competing risk frame-work, and we have studied the Bayesian inference also. Though in this paper, we have worked under bivariate set-up, it can be extended to

Hyper-parameters	Parameter	Av. Estimate	MSE
$a_0 = 5,\ b_0 = 3.968$	θ	0.782	0.029
$a_1 = 5,\ b_1 = 5.649$	α	1.686	0.036
$a_2 = 5,\ b_2 = 1.707$	λ_0	1.105	0.057
$a_3 = 5,\ b_3 = 1.505$	λ_1	0.965	0.036
$a_4 = 5,\ b_4 = 2.939$	λ_2	4.015	0.966
$a_0 = 10,\ b_0 = 7.936$	θ	0.755	0.019
$a_1 = 10,\ b_1 = 11.299$	α	1.667	0.028
$a_2 = 10,\ b_2 = 3.415$	λ_0	1.117	0.038
$a_3 = 10,\ b_3 = 3.010$	λ_1	0.937	0.015
$a_4 = 10,\ b_4 = 5.878$	λ_2	3.670	0.662
$a_0 = 20,\ b_0 = 15.873$	θ	0.707	0.013
$a_1 = 20,\ b_1 = 22.598$	α	1.668	0.022
$a_2 = 20,\ b_2 = 6.830$	λ_0	1.137	0.025
$a_3 = 20,\ b_3 = 6.020$	λ_1	0.913	0.006
$a_4 = 20,\ b_4 = 11.757$	λ_2	3.383	0.315

Table 5: Average estimates and mean squared errors of the posterior mode of the parameters with $\theta = 0.699, \alpha = 1.701, \lambda_0 = 1.260, \lambda_1 = 0.885, \lambda_2 = 2.928$ and $n = 25$

multivariate set-up. Also under different censoring schemes, we can develop the analysis of BWG model. More works are needed on those directions.

Acknowledgements:

The authors would like to thank two unknown reviewers and the associate editor for their constructive comments which have helped to improve the earlier draft of the paper significantly.

Appendix:

In this section we provide the explicit expression of $Q(\mathbf{\Theta}, \mathbf{\Theta}^{(k)})$ and will show how it can be maximized by solving one non-linear equation only. In this section we use the following

Hyper-parameters	Parameter	Av. Estimate	MSE
$a_0 = 5,\ b_0 = 3.968$	θ	0.771	0.020
$a_1 = 5,\ b_1 = 5.649$	α	1.503	0.071
$a_2 = 5,\ b_2 = 1.707$	λ_0	1.140	0.057
$a_3 = 5,\ b_3 = 1.505$	λ_1	0.824	0.047
$a_4 = 5,\ b_4 = 2.939$	λ_2	2.437	0.559
$a_0 = 10,\ b_0 = 7.936$	θ	0.747	0.011
$a_1 = 10,\ b_1 = 11.299$	α	1.539	0.057
$a_2 = 10,\ b_2 = 3.415$	λ_0	1.167	0.033
$a_3 = 10,\ b_3 = 3.010$	λ_1	0.833	0.023
$a_4 = 10,\ b_4 = 5.878$	λ_2	2.507	0.385
$a_0 = 20,\ b_0 = 15.873$	θ	0.709	0.006
$a_1 = 20,\ b_1 = 22.598$	α	1.590	0.038
$a_2 = 20,\ b_2 = 6.830$	λ_0	1.205	0.014
$a_3 = 20,\ b_3 = 6.020$	λ_1	0.857	0.008
$a_4 = 20,\ b_4 = 11.757$	λ_2	2.630	0.197

Table 6: Average estimates and mean squared errors of the Bayes estimator based on squared error loss function of the parameters with $\theta = 0.699, \alpha = 1.701, \lambda_0 = 1.260, \lambda_1 = 0.885, \lambda_2 = 2.928$ and $n = 25$

notation

$$E(N|y_{1i}, y_{2i}, \theta, \boldsymbol{\gamma}) = a_i(\theta, \boldsymbol{\gamma}), \quad E(N|y_{1i}, y_{2i}, \theta, \boldsymbol{\gamma}^{(k)}) = a_i^{(k)}.$$

Note that $E(N|y_{1i}, y_{2i}, \theta, \boldsymbol{\gamma})$ is obtained from (12). We further use

$$u_1^{(k)} = \frac{\lambda_0^{(k)}}{\lambda_1^{(k)} + \lambda_0^{(k)}}, \quad u_2^{(k)} = \frac{\lambda_1^{(k)}}{\lambda_1^{(k)} + \lambda_0^{(k)}}, \quad v_1^{(k)} = \frac{\lambda_0^{(k)}}{\lambda_2^{(k)} + \lambda_0^{(k)}}, \quad v_2^{(k)} = \frac{\lambda_2^{(k)}}{\lambda_2^{(k)} + \lambda_0^{(k)}}.$$

Now following exactly the same procedure as in Kundu and Gupta [10], we can write $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(k)})$ without the additive constant (which involves θ also) as follows:

$$\begin{aligned} Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(k)}) &= (2k_1 + 2k_2 + k_0 + a_4 - 1)\ln\alpha + \alpha\left(\sum_{i\in I_0}\ln y_i + \sum_{i\in i_1\cup I_2}(\ln y_{1i} + \ln y_{2i}) - b_4\right) \\ &+ (k_1 u_2^{(k)} + k_2 v_1^{(k)} + k_0 + a_0 - 1)\ln\lambda_0 - \lambda_0\left(\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1} a_i^{(k)} y_{1i}^{\alpha} + \sum_{i\in I_2} a_i^{(k)} y_{2i}^{\alpha} + b_0\right) \\ &+ (k_1 u_1^{(k)} + k_2 + a_1 - 1)\ln\lambda_1 - \lambda_1\left(\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1\cup I_2} a_i^{(k)} y_{1i}^{\alpha} + b_1\right) \end{aligned}$$

Hyper-parameters	Parameter	Av. Estimate	MSE
$a_0 = 5,\ b_0 = 3.968$	θ	0.771	0.022
$a_1 = 5,\ b_1 = 5.649$	α	1.366	0.139
$a_2 = 5,\ b_2 = 1.707$	λ_0	0.967	0.111
$a_3 = 5,\ b_3 = 1.505$	λ_1	0.719	0.057
$a_4 = 5,\ b_4 = 2.939$	λ_2	2.149	0.820
$a_0 = 10,\ b_0 = 7.936$	θ	0.743	0.013
$a_1 = 10,\ b_1 = 11.299$	α	1.431	0.099
$a_2 = 10,\ b_2 = 3.415$	λ_0	1.059	0.055
$a_3 = 10,\ b_3 = 3.010$	λ_1	0.782	0.026
$a_4 = 10,\ b_4 = 5.878$	λ_2	2.325	0.523
$a_0 = 20,\ b_0 = 15.873$	θ	0.711	0.008
$a_1 = 20,\ b_1 = 22.598$	α	1.516	0.057
$a_2 = 20,\ b_2 = 6.830$	λ_0	1.150	0.019
$a_3 = 20,\ b_3 = 6.020$	λ_1	0.831	0.009
$a_4 = 20,\ b_4 = 11.757$	λ_2	2.545	0.239

Table 7: Average estimates and mean squared errors of the Bayes estimator based on squared error loss function of the parameters in presence of competing risk factors with $\theta = 0.699, \alpha = 1.701, \lambda_0 = 1.260, \lambda_1 = 0.885, \lambda_2 = 2.928$ and $n = 25$

$$+ \quad (k_1 + k_2 v_2^{(k)} + a_2 - 1)\ln\lambda_2 - \lambda_2\left(\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1} a_i^{(k)} y_{2i}^{\alpha} + b_2\right).$$

The maximization of $Q(\mathbf{\Theta}, \mathbf{\Theta}^{(k)})$ with respect to λ_0, λ_1, λ_2 and α can be done very easily similarly as in Kundu and Gupta [10]. Obtain

$$\lambda_0^{(k+1)}(\alpha) = \frac{k_1 u_2^{(k)} + k_2 v_1^{(k)} + k_0 + a_0 - 1}{\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1} a_i^{(k)} y_{1i}^{\alpha} + \sum_{i\in I_2} a_i^{(k)} y_{2i}^{\alpha} + b_0}$$

$$\lambda_1^{(k+1)}(\alpha) = \frac{k_1 u_1^{(k)} + k_2 + a_1 - 1}{\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1\cup I_2} a_i^{(k)} y_{1i}^{\alpha} + b_1}$$

$$\lambda_2^{(k+1)}(\alpha) = \frac{k_1 + k_2 v_2^{(k)} + a_2 - 1}{\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1} a_i^{(k)} y_{2i}^{\alpha} + b_2}.$$

Now we will mention how to obtain $\alpha^{(k+1)}$. Let us define the function $h(\alpha)$ as follows:

$$h(\alpha) = \left[\lambda_0^{(k+1)}(\alpha)\left(\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1} a_i^{(k)} y_{1i}^{\alpha} + \sum_{i\in I_2} a_i^{(k)} y_{2i}^{\alpha} + b_0\right)\right.$$

$$
\begin{aligned}
&+\lambda_1^{(k+1)}(\alpha)\left(\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1\cup I_2} a_i^{(k)} y_{1i}^{\alpha} + b_1\right)\\
&+\lambda_2^{(k+1)}(\alpha)\left(\sum_{i\in I_0} a_i^{(k)} y_i^{\alpha} + \sum_{i\in I_1} a_i^{(k)} y_{2i}^{\alpha} + b_2\right)\\
&-\left(\sum_{i\in I_0} \ln y_i + \sum_{i\in i_1\cup I_2} (\ln y_{1i} + \ln y_{2i}) - b_4\right)\Bigg].
\end{aligned}
$$

Then $\alpha^{(k+1)}$ can be obtained as a fixed point solution of

$$g(\alpha) = \alpha,$$

where

$$g(\alpha) = \frac{2k_1 + 2k_2 + k_0 + a_4 - 1}{h(\alpha)}.$$

Therefore, first we can obtain $\alpha^{(k+1)}$, and once $\alpha^{(k+1)}$ is obtained, then $\lambda_0^{(k+1)}$, $\lambda_1^{(k+1)}$, $\lambda_2^{k+1)}$ can be obtained as

$$\lambda_0^{(k+1)} = \lambda_0^{(k+1)}(\alpha^{(k+1)}), \qquad \lambda_1^{(k+1)} = \lambda_1^{(k+1)}(\alpha^{(k+1)}), \qquad \lambda_2^{k+1)} = \lambda_2^{(k+1)}(\alpha^{(k+1)}).$$

∎

PROOF OF THEOREM 2: The proof goes the same way as the conditioning argument. Note that

$$
\begin{aligned}
P(y_1 \le Y_1 < y_1 + dy_1, Y_2 > y_1) &= \sum_{n=1}^{\infty} P(y_1 \le Y_1 < y_1 + dy_1, Y_2 > y_1 | N = n) P(N = n)\\
&= \sum_{n=1}^{\infty} f_{WE}(y_1, \alpha, n\lambda_1) S_{WE}(y_1, \alpha, n(\lambda_0 + \lambda_2)) \theta (1-\theta)^{n-1} dy_1\\
&= \alpha\lambda_1\theta y_1^{\alpha-1} \sum_{n=1}^{\infty} n e^{-n(\lambda_0+\lambda_1+\lambda_2) y_1^{\alpha}} (1-\theta)^{n-1} dy_1\\
&= \frac{\theta\alpha\lambda_1 y_1^{\alpha-1} e^{-(\lambda_0+\lambda_1+\lambda_2) y_1^{\alpha}}}{(1-(1-\theta)e^{-(\lambda_0+\lambda_1+\lambda_2) y_1^{\alpha}})^2} dy_1.
\end{aligned}
$$

∎

References

[1] Bai, X., Shi, Y., Ng, H.K.T. and Liu, Y. (2020), "Inference of accelerated dependent competing risks model for Marshall–Olkin bivariate Weibull distribution with non constant parameters", *Journal of Computational and Applied Mathematics*, vol. 366, Article ID 112398.

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

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

[4] Devroye, L. (1984), "A simple algorithm for generating random variates with a log-concave density", *Computing*, vol. 33, 247-257.

[5] Feizjavadian, S.H. and Hashemi, R. (2015), "Analysis of dependent competing risks in the presence of progressive hybrid censoring using Marshall-Olkin bivariate Weibull distribution", *Computational Statistics and Data Analysis*, vol. 82, 19-34.

[6] Ghosh, J.K., Delampady, M. and Samanta, T. (2006), *An introduction to Bayesian analysis*, Springer, New Delhi.

[7] Hanagal, D. (2006), "Bivariate Weibull regression model based on censored samples", *Statistical Papers*, vol. 47, 137-147.

[8] Kundu, D. (2008), "Bayesian inference and life testing plan for Weibull distribution in presence of progressive censoring", *Technometrics*, vol. 50, 144-154.

[9] Kundu, D. and Dey, A.K. (2009), "Estimating the Parameters of the Marshall Olkin Bivariate Weibull distribution by EM Algorithm", *Computational Statistics and Data Analysis*, vol. 53, 956-965.

[10] Kundu, D. and Gupta, A.K. (2014), "On bivariate Weibull Geometric distribution", *Journal of Multivariate Analysis*, vol. 123, 19-29.

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

[12] Lu, Jye-Chyi (1992), "Bayes parameter estimation for the bivariate Weibull model of Marshall-Olkin for censored data", *IEEE Transactions on Reliability*, vol. 41, 608-615.

[13] Marshall, A.W. and Olkin, I. (1967), "A multivariate exponential distribution", *Journal of the American Statistical Association*, vol. 62, 30-44.

[14] Marshall, A.W. and Olkin, I. (1997), "A new method of adding a parameter to a family of distributions with application to the exponential and Weibull families", *Biometrika*, vol. 84, 641-652.

[15] Meintanis, S.G.(2007), "Marshall-Olkin distributions with applications", *Journal of Statistical Planning and Inference*, vol. 137, 3954-3963.

[16] Nelsen, R.B. (2016), *Introduction to copulas*, 2nd edition, Springer, U.S.A.

[17] Shen, Y. and Xu, A. (2018), "On the dependent competing risks using Marshall-Olkin bivariate Weibull model: Parameter estimation with different methods", *Communications in Statistics - Theory and Methods*, vol. 47, 5558-5572.

[18] Zhang, X., Shang, J., Chen, X., Zhang, C. and Wang, Y. (2014), "Statistical inference of accelerated life testing with dependent competing failures based on copula theory", *IEEE Transactions on Reliability*, vol. 63, 764-780.