isid/ms/2009/12
December 1, 2009
http://www.isid.ac.in/~statmath/eprints

An EM algorithm for estimating the parameters of bivariate Weibull distribution under random censoring

Isha Dewan
Swagata Nandi

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

AN EM ALGORITHM FOR ESTIMATING THE PARAMETERS OF BIVARIATE WEIBULL DISTRIBUTION UNDER RANDOM CENSORING

Isha Dewan and Swagata Nandi*
Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, Delhi Centre
7, S. J. S. Sansanwal Marg
New Delhi - 110016.

Abstract

We consider the problem of estimation of the parameters of the Marshall-Olkin Bivariate Weibull distribution in the presence of random censoring. Since the maximum likelihood estimators of the parameters can not be expressed in a closed form, we suggest an EM algorithm to compute the same. Extensive simulations are done to conclude that the estimators perform efficiently under random censoring.

Keywords and Phrases: Marshall-Olkin Bivariate Distribution; Random Censoring; EM Algorithm; Pseudo Likelihood.

2000 Mathematics Subject Classification: 62N01, 62N02

1 Introduction

Many a times the life/failure data of interest is bivariate in nature. Any study on twins or on failure data recorded twice on the same system naturally leads to bivariate data. For example, Houggard, Harvald and Holm (1992) studied data on lifelength of Danish twins and Lin, Sun and Ying (1999) considered a data on patients of colon cancer where the paired data consists of the time from treatment to recurrence of the cancer and the time from treatment to death. Paired data could consist of blindness in the left/right eye, failure time of the left/right kidney or age at death of parent/child in a genetic study. However, one or both components of the

*Corresponding Author

paired data could be subject to random censoring. It could arise because a parent and/or the child might be alive till the end of the study and hence the failure time would be censored.

Weibull distribution is often used to model reliability/survival data. When one looks at bivariate data it is natural to look at extensions of Weibull distribution to fit such data. One such distribution is the Marshall-Olkin Bivariate Weibull (MOBW) distribution. It is a generalisation of the Marshall-Olkin Bivariate Exponential (MOBE) distribution which was introduced by Marshall and Olkin (1967). The motivation behind MOBE was the common failure of components induced from Poisson shocks. Under MOBW the arrival of shocks is governed by non homogeneous Poisson processes, each having power law intensity. MOBW is a bivariate distribution which, like MOBE, has the absolutely continuous part and a singular part, that is, the pair of random variables can be equal with positive probability. This distribution fits a bivariate data set very well if it has unimodal marginal density function or has non constant hazard function. Besides, it is often used to fit paired data in survival studies where there is a possibility of simultaneous occurrence of both the events.

Meintanis (2007) considered soccer data from UEFA Champions League for the years 2004-05 and 2005-06. Let X_1 denote the time (in minutes) of the first kick goal (penalty kick, foul kick, or other kick) scored by any team, and X_2 denote the time of the first goal of any type scored by the home team. The data consists of all three cases $X_1 < X_2$, $X_1 = X_2$, and $X_1 > X_2$. Kundu and Dey (2009) showed that MOBW fits to the data. Meintanis (2007) also studied the white blood cells (WBC) counts of 60 patients. Let X_1 be the WBC count of a patient before an operation, and X_2 be the corresponding WBC at a specific time during or after the operation. WBC count varies significantly during or after treatment as compared to pre operation value. One could fit MOBW to this data.

Lu (1989,1992) considered Bayes estimators of the parameters of Weibull distribution. Gupta and Kundu (2003) studied how to discriminate between Weibull and Generalised Exponential distributions. Hanagal (2005,2006) studied Bivariate Weibull regression model for uncensored and censored data. Kundu and Dey (2009) studied an EM algorithm for computing the maximum likelihood estimators (M.L.E.'s) of the parameters of the bivariate Weibull distribution in case of complete data, that is, data on all pairs of variables are available.

We study the M.L.E.'s of the parameters of Weibull distribution under random censoring. In section 2, we state expressions for the joint density and survival function of MOBW distribution for completeness and study the likelihood function of data coming from MOBW distribution when it is subjected to random right censoring. In section 3, we suggest the use of an EM (conditional) algorithm for finding the M.L.E.'s of the parameters. In section 4, we carry out a simulation study to see the performance of the proposed estimators. In section 5, we re-study the soccer data analysed by Meintanis (2007) under random censoring. In the final section we summarize our results and indicate directions for future work. The details of the likelihood function and the Fisher information matrix are given in Appendices A and B, respectively.

2 Marshall-Olkin Bivariate Weibull Distribution

Consider a Weibull distribution ($WE(\alpha, \theta)$) with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$. The density function f, distribution function F and survival function S, are given by,

$$f_{WE}(x; \alpha, \theta) \quad = \quad \alpha\theta x^{\alpha-1}e^{-\theta x^{\alpha}}, \;\; S_{WE}(x; \alpha, \theta) = 1 - F_{WE}(x; \alpha, \theta) = e^{-\theta x^{\alpha}}, \; x > 0.$$

Suppose U_0, U_1, U_2, respectively, are independent $WE(\alpha, \lambda_0), WE(\alpha, \lambda_1), WE(\alpha, \lambda_2)$, random variables. Let $X_1 = \min(U_0, U_1)$ and $X_2 = \min(U_0, U_2)$. Then (X_1, X_2) has MOBW distribution with parameters $\alpha, \lambda_0, \lambda_1, \lambda_2$ and is expressed as MOBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$. It should be noted that the three random variables U_0, U_1, U_2 have the common shape parameter. This ensures that the marginal distributions of X_1 and X_2 are $WE(\alpha, \lambda_0 + \lambda_1)$ and $WE(\alpha, \lambda_0 + \lambda_2)$, respectively. Further, the distribution of $T = \min(X_1, X_2)$ is $WE(\alpha, \lambda_0 + \lambda_1 + \lambda_2)$. The ageing properties of the Weibull distribution are characterised by the shape parameter, $\alpha < (=, >)\ 1$, respectively, indicating decreasing (constant and increasing failure) rate. The three random variables U_0, U_1, U_2 which denote lifetimes under three Weibull shock processes would have failure rates all increasing, decreasing or constant. When $\alpha = 1$, it reduces to MOBE distribution and is expressed as MOBE$(\lambda_0, \lambda_1, \lambda_2)$. We use $\lambda = \lambda_0 + \lambda_1 + \lambda_2$ throughout the manuscript.

The joint survival function of (X_1, X_2) is given as follows, where $z = \max(x, y)$,

$$S(x,y) = \begin{cases} S_{WE}(x;\alpha,\lambda_1)S_{WE}(y;\alpha,\lambda_0+\lambda_2) & \text{if} \quad x < y \\ \\ S_{WE}(x;\alpha,\lambda_0+\lambda_1)S_{WE}(y;\alpha,\lambda_2) & \text{if} \quad x > y \\ \\ S_{WE}(x;\alpha,\lambda) & \text{if} \quad x = y. \end{cases}$$

The joint density function of (X_1, X_2) is given as

$$f(x,y) = \begin{cases} f_{WE}(x;\alpha,\lambda_1) f_{WE}(y;\alpha,\lambda_0+\lambda_2) & \text{if} \quad x < y \\ \\ f_{WE}(x;\alpha,\lambda_0+\lambda_1) f_{WE}(y;\alpha,\lambda_2) & \text{if} \quad x > y \\ \\ \frac{\lambda_0}{\lambda} f_{WE}(x;\alpha,\lambda) & \text{if} \quad x = y. \end{cases}$$

Following two expressions are required for writing the likelihood function.

$$\int_y^\infty f(x,u)du = \begin{cases} f_{WE}(x;\alpha,\lambda_1) S_{WE}(y;\alpha,\lambda_0+\lambda_2) & \text{if} \quad x < y \\ \\ f_{WE}(x;\alpha,\lambda_0+\lambda_1)[S_{WE}(y;\alpha,\lambda_2) - S_{WE}(x;\alpha,\lambda_2)] & \text{if} \quad x > y \\ \\ f_{WE}(x;\alpha,\lambda_1) S_{WE}(x;\alpha,\lambda_0+\lambda_2) & \text{if} \quad x = y, \end{cases}$$

$$\int_x^\infty f(u,y)du = \begin{cases} f_{WE}(y;\alpha,\lambda_0+\lambda_2)[S_{WE}(x;\alpha,\lambda_1) - S_{WE}(y;\alpha,\lambda_1)] & \text{if} \quad x < y \\ \\ S_{WE}(x;\alpha,\lambda_0+\lambda_1) f_{WE}(y;\alpha,\lambda_2) & \text{if} \quad x > y \\ \\ f_{WE}(x;\alpha,\lambda_2) S_{WE}(x;\alpha,\lambda_0+\lambda_1) & \text{if} \quad x = y. \end{cases}$$

The pair (X_1, X_2) is subject to random censoring by an independent pair of random variables (Z_1, Z_2). We observe

$$\begin{aligned} T_1 &= \min(X_1, Z_1) \quad \text{and} \quad \delta_1 = I(X_1 < Z_1), \\ T_2 &= \min(X_2, Z_2) \quad \text{and} \quad \delta_2 = I(X_2 < Z_2), \end{aligned}$$

where $I(A)$ denotes the indicator function of set A.

The likelihood function, based on observed pairs $(t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i}),\ i = 1, 2, \ldots, n$ is given by

$$\begin{aligned} L &= L(\alpha, \lambda_0, \lambda_1, \lambda_2, t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i},\ i = 1, 2, \ldots, n) \\ &= \prod_{i=1}^n [f(t_{1i}, t_{2i}]^{\delta_{1i}\delta_{2i}} \left[\int_{t_{2i}}^\infty f(t_{1i}, y)dy\right]^{\delta_{1i}(1-\delta_{2i})} \\ &\qquad \times \left[\int_{t_{1i}}^\infty f(x, t_{2i})dx\right]^{(1-\delta_{1i})\delta_{2i}} [S(t_{1i}, t_{2i}]^{(1-\delta_{1i})(1-\delta_{2i})}. \end{aligned}$$

Note that, when $\delta_1 = \delta_2 = 1$, both failure times are observed and the contribution to the likelihood is $f(t_1, t_2)$. When $\delta_1 = 1 - \delta_2 = 1$, the first component fails at t_1 and the second component is censored (lives beyond t_2) and the contribution to the likelihood is $\int_{t_2}^\infty f(t_1, y)dy$. Similarly, when $1 - \delta_1 = \delta_2 = 1$, the first component is censored and the second component fails and the contribution to the likelihood is $\int_{t_1}^{\infty} f(x, t_2)dx$. Finally, when $1 - \delta_1 = 1 - \delta_2 = 1$, both failure times are censored and the contribution to the likelihood is $S(t_1, t_2)$.

Let I_0, I_1, I_2, denote the following sets

$$I_0 = \{i|t_{1i} = t_{2i} = t_i\},\ I_1 = \{i|t_{1i} < t_{2i}\},\ I_2 = \{i|t_{1i} > t_{2i}\}.$$

Let n_0, n_1, n_2, respectively, denote the number of elements in the sets I_0, I_1, I_2.

Then the likelihood function can be written as

$$L = \prod_{i \in I_0} L_0(t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i}) \prod_{i \in I_1} L_1(t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i}) \prod_{i \in I_2} L_2(t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i}),$$

where $L_k(t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i}) \equiv L_k(\alpha, \lambda_0, \lambda_1, \lambda_2, t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i})$ is the contribution from I_k to the likelihood function, $k = 0, 1, 2$ and they are given in Appendix A explicitly. Define

$$\begin{aligned}
n_{11} &= \text{number of pairs for which } \ \delta_1 = \delta_2 = 1, \\
n_{10} &= \text{number of pairs for which } \ \delta_1 = 1 - \delta_2 = 1, \\
n_{01} &= \text{number of pairs for which } \ 1 - \delta_1 = \delta_2 = 1, \\
n_{00} &= \text{number of pairs for which } \ 1 - \delta_1 = 1 - \delta_2 = 1,
\end{aligned}$$

then

$$\begin{aligned}
n &= n_{11} + n_{10} + n_{01} + n_{00}, \\
n_0 &= n_{11}^0 + n_{10}^0 + n_{01}^0 + n_{00}^0, \\
n_1 &= n_{11}^1 + n_{10}^1 + n_{01}^1 + n_{00}^1, \\
n_2 &= n_{11}^2 + n_{10}^2 + n_{01}^2 + n_{00}^2,
\end{aligned}$$

where n_{ij}^k denotes the number of individuals in I_k with $\delta_1 = i, \delta_2 = j,\ i, j = 0, 1,\ k = 0, 1, 2$; $n_{ij} = \sum_{k=0}^{2} n_{ij}^k$.

3 EM Algorithm under Random Censoring

The problem of finding M.L.E.'s of the unknown parameters of MOBW has been studied earlier, when there is no censoring. Bemis, Bain and Higgins (1972) showed that when $\alpha = 1$, that is, underlying distribution is MOBE, the M.L.E.'s do not exist if one of $n_i = 0$. If each one of $n_0, n_1, n_2 > 0$, then the M.L.E.'s of $\lambda_0, \lambda_1, \lambda_2$ exist and can be obtained by solving three non linear equations. Kundu and Dey (2009) showed that M.L.E.'s of MOBW distribution exist when $n_0, n_1, n_2 > 0$. Maximising the likelihood with respect to $\alpha, \lambda_0, \lambda_1, \lambda_2$ is a non linear optimisation problem. They have looked at the pseudo likelihood with information on ordering of U_0, U_1, U_2 missing. They used the EM (conditional) algorithm to compute the M.L.E.'s . To the best of our knowledge, nobody has considered maximum likelihood estimation of parameters of a bivariate distribution under random censoring.

Under random censoring, on the set I_0, all the four parameters are identifiable. On the set I_1, we can identify $\alpha, \lambda_0 + \lambda_1, \lambda_2$, whereas on the set I_2, we can identify $\alpha, \lambda_1, \lambda_0 + \lambda_2$.

Let γ denote the parameter vector $(\alpha, \lambda_0, \lambda_1, \lambda_2)^T$. It is easy to see that

$$P(U_1 < U_0 < U_2) = \frac{\lambda_0 \lambda_1}{(\lambda_0 + \lambda_2)\lambda}, \quad P(U_1 < U_2 < U_0) = \frac{\lambda_1 \lambda_2}{(\lambda_0 + \lambda_2)\lambda},$$

$$P(U_2 < U_0 < U_1) = \frac{\lambda_0 \lambda_2}{(\lambda_0 + \lambda_1)\lambda}, \quad P(U_2 < U_1 < U_0) = \frac{\lambda_1 \lambda_2}{(\lambda_0 + \lambda_1)\lambda}.$$

Then

$$\begin{aligned} \mu_1(\gamma) &= P(U_1 < U_0 < U_2 | X_1 < X_2) = \frac{\lambda_0}{\lambda_0 + \lambda_2}, \\ \mu_2(\gamma) &= P(U_1 < U_2 < U_0 | X_1 < X_2) = \frac{\lambda_2}{\lambda_0 + \lambda_2}, \\ \nu_1(\gamma) &= P(U_2 < U_0 < U_1 | X_1 > X_2) = \frac{\lambda_0}{\lambda_0 + \lambda_1}, \\ \nu_2(\gamma) &= P(U_2 < U_1 < U_0 | X_1 > X_2) = \frac{\lambda_1}{\lambda_0 + \lambda_1}. \end{aligned}$$

In order to identify all parameters uniquely, we write the 'E' step of the algorithm as follows. We form a pseudo likelihood by replacing the log-likelihood contribution of the observed $(T_1, \delta_1, T_2, \delta_2)$ by its expected value.

The log-likelihood function of the 'pseudo data' has three parts corresponding to contributions from the sets I_0, I_1, I_2.

The contribution to the pseudo log-likelihood from I_0 is

$$\begin{aligned} & \sum_{i \in I_0} \Big\{ \delta_{1i}\delta_{2i}[\log \lambda_0 + \log \alpha + (\alpha - 1)\log t_i - \lambda t_i^{\alpha}] \\ & +\delta_{1i}(1 - \delta_{2i})[\log \lambda_1 + \log \alpha + (\alpha - 1)\log t_i - \lambda t_i^{\alpha}] \\ + \quad & (1 - \delta_{1i})\delta_{2i}[\log \lambda_2 + \log \alpha + (\alpha - 1)\log t_i - \lambda t_i^{\alpha}] + (1 - \delta_{1i})(1 - \delta_{2i})[-\lambda t_i^{\alpha}] \Big\}. \end{aligned}$$

The contribution to the pseudo log-likelihood from I_1 is

$$\sum_{i \in I_1} \Big\{ \delta_{1i}\delta_{2i} \Big(\mu_1 [\log \lambda_0 + \log \lambda_1 + 2\log \alpha + (\alpha - 1)\log t_{1i} + (\alpha - 1)\log t_{2i} -$$

$$
\begin{aligned}
& \lambda_1 t_{1i}^{\alpha} - (\lambda_0+\lambda_2)t_{2i}^{\alpha}] + \mu_2[\log\lambda_1 + \log\lambda_2 + 2\log\alpha + (\alpha-1)\log t_{1i} + \\
& (\alpha-1)\log t_{2i} - \lambda_1 t_{1i}^{\alpha} - (\lambda_0+\lambda_2)t_{2i}^{\alpha}]\Big) \\
+ \quad & \delta_{1i}(1-\delta_{2i})\Big(\log\lambda_1 + \log\alpha + (\alpha-1)\log t_{1i} - \lambda_1 t_{1i}^{\alpha} - (\lambda_0+\lambda_2)t_{2i}^{\alpha}\Big) \\
+ \quad & (1-\delta_{1i})\delta_{2i}\Big(\mu_1[\log\lambda_0 + \log\alpha + (\alpha-1)\log t_{2i} - (\lambda_0+\lambda_2)t_{2i}^{\alpha} + \\
& \log\{e^{-\lambda_1 t_{1i}^{\alpha}} - e^{-\lambda_1 t_{2i}^{\alpha}}\}] + \mu_2[\log\lambda_2 + \log\alpha + (\alpha-1)\log t_{2i} - \\
& (\lambda_0+\lambda_2)t_{2i}^{\alpha} + \log\{e^{-\lambda_1 t_{1i}^{\alpha}} - e^{-\lambda_1 t_{2i}^{\alpha}}\}]\Big) \\
+ \quad & (1-\delta_{1i})(1-\delta_{2i})[-\lambda_1 t_{1i}^{\alpha} - (\lambda_0+\lambda_2)t_{2i}^{\alpha}]\Big\}.
\end{aligned}
$$

Finally, the contribution to the pseudo log-likelihood from I_2 is

$$
\begin{aligned}
& \sum_{i\in I_2}\Big\{\delta_{1i}\delta_{2i}\Big(\nu_1[\log\lambda_0 + \log\lambda_2 + 2\log\alpha + (\alpha-1)\log t_{1i} + (\alpha-1)\log t_{2i} - \\
& (\lambda_0+\lambda_1)t_{1i}^{\alpha} - \lambda_2 t_{2i}^{\alpha}] + \nu_2[\log\lambda_1 + \log\lambda_2 + 2\log\alpha + (\alpha-1)\log t_{1i} + \\
& (\alpha-1)\log t_{2i} - (\lambda_0+\lambda_1)t_{1i}^{\alpha} - \lambda_2 t_{2i}^{\alpha}]\Big) \\
+ \quad & (1-\delta_{1i})\delta_{2i}\Big(\log\lambda_2 + \log\alpha + (\alpha-1)\log t_{2i} - (\lambda_0+\lambda_1)t_{1i}^{\alpha} - \lambda_2 t_{2i}^{\alpha}\Big) \\
+ \quad & \delta_{1i}(1-\delta_{2i})\Big(\nu_1[\log\lambda_0 + \log\alpha + (\alpha-1)\log t_{1i} - (\lambda_0+\lambda_1)t_{1i}^{\alpha} + \\
& \log\{e^{-\lambda_2 t_{2i}^{\alpha}} - e^{-\lambda_2 t_{1i}^{\alpha}}\}] + \nu_2[\log\lambda_1 + \log\alpha + (\alpha-1)\log t_{1i} - \\
& (\lambda_0+\lambda_1)t_{1i}^{\alpha} + \log\{e^{-\lambda_2 t_{2i}^{\alpha}} - e^{-\lambda_2 t_{1i}^{\alpha}}\}]\Big) \\
+ \quad & (1-\delta_{1i})(1-\delta_{2i})[-(\lambda_0+\lambda_1)t_{1i}^{\alpha} - \lambda_2 t_{2i}^{\alpha}]\Big\}.
\end{aligned}
$$

Let

$$
\begin{aligned}
N_0 &= n_{11}^0 + \mu_1(n_{11}^1 + n_{01}^1) + \nu_1(n_{11}^2 + n_{10}^2), \\
N_1 &= n_{10}^0 + n_{11}^1 + n_{10}^1 + \nu_2(n_{11}^2 + n_{10}^2), \\
N_2 &= n_{01}^0 + \mu_2(n_{11}^1 + n_{01}^1) + n_{11}^2 + n_{01}^2, \\
N_3 &= n + n_{11}^1 + n_{11}^2 - n_{00}. \qquad (1)
\end{aligned}
$$

Hence, the pseudo log-likelihood is given by

$$
\begin{aligned}
& N_0\log\lambda_0 + N_1\log\lambda_1 + N_2\log\lambda_2 + N_3\log\alpha \\
& +(\alpha-1)\sum_{i\in I_0}[\delta_{1i}\delta_{2i} + (1-\delta_{1i})\delta_{2i} + \delta_{1i}(1-\delta_{2i})]\log t_i - \lambda\sum_{i\in I_0} t_i^{\alpha} \\
& +(\alpha-1)\sum_{i\in I_1}[\delta_{1i}\log t_{1i} + \delta_{2i}\log t_{2i}] + (\alpha-1)\sum_{i\in I_2}[\delta_{1i}\log t_{1i} + \delta_{2i}\log t_{2i}]
\end{aligned}
$$

$$-\lambda_1 \sum_{i \in I_1} [\delta_{1i}\delta_{2i} + \delta_{1i}(1-\delta_{2i}) + (1-\delta_{1i})(1-\delta_{2i})] t_{1i}^{\alpha} - (\lambda_0 + \lambda_2) \sum_{i \in I_1} t_{2i}^{\alpha}$$
$$-(\lambda_0 + \lambda_1) \sum_{i \in I_2} t_{1i}^{\alpha} - \lambda_2 \sum_{i \in I_2} [\delta_{1i}\delta_{2i} + (1-\delta_{1i})\delta_{2i} + (1-\delta_{1i})(1-\delta_{2i})] t_{2i}^{\alpha}$$
$$+\sum_{i \in I_1} (1-\delta_{1i})\delta_{2i} \log(\exp(-\lambda_1 t_{1i}^{\alpha}) - \exp(-\lambda_1 t_{2i}^{\alpha}))$$
$$+\sum_{i \in I_2} \delta_{1i}(1-\delta_{2i}) \log(\exp(-\lambda_2 t_{2i}^{\alpha}) - \exp(-\lambda_2 t_{1i}^{\alpha})). \quad (2)$$

The 'M' step involves in maximising the pseudo log-likelihood w.r.t. $\alpha, \lambda_0, \lambda_1$, and λ_2.

Let
$$\frac{\partial \log L}{\partial \alpha} = g_1, \qquad \frac{\partial \log L}{\partial \lambda_0} = g_2, \qquad \frac{\partial \log L}{\partial \lambda_1} = g_3, \qquad \frac{\partial \log L}{\partial \lambda_2} = g_4,$$
where g_i, $i = 1, \ldots, 4$, defined in Appendix B, are first order derivatives of the pseudo log-likelihood function.

Then, the M.L. equations are given by

$$g_i = 0, \quad i = 1, \ldots, 4.$$

For fixed α, the maximum w.r.t. λ_0 of the pseudo log-likelihood function can be obtained from the M.L. equation of λ_0 directly, whereas the M.L. equations of α, λ_1 and λ_2 are interrelated and can not be solved explicitly. In order to maximize the pseudo log-likelihood function w.r.t. α, λ_1 and λ_2, we use a method suggested in Kundu and Gupta (2006). We will solve three fixed point type equations iteratively.

$$g_{\alpha}(\alpha, \lambda_0, \lambda_1, \lambda_2) = \alpha, \quad (3)$$
$$g_{\lambda_1}(\alpha, \lambda_1) = \lambda_1, \quad (4)$$
$$g_{\lambda_2}(\alpha, \lambda_2) = \lambda_2, \quad (5)$$

where,

$$g_{\alpha}(\alpha, \lambda_0, \lambda_1, \lambda_2) = \frac{N_3}{h_1(\alpha, \lambda_0, \lambda_1, \lambda_2)},$$
$$g_{\lambda_1}(\alpha, \lambda_1) = \frac{N_1}{h_2(\alpha, \lambda_1)}, \qquad g_{\lambda_2}(\alpha, \lambda_2) = \frac{N_2}{h_3(\alpha, \lambda_2)},$$

where $N_i, i = 1, 2, 3$, are defined in equations (1) and $h_1(\alpha, \lambda_0, \lambda_1, \lambda_2)$, $h_2(\alpha, \lambda_1)$, and $h_3(\alpha, \lambda_2)$ are given in Appendix B.

Hence, in order to solve fixed point equations (3)-(5), we start with an initial guess of the parameter vector $\gamma^{(0)} = (\alpha^{(0)}, \lambda_0^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)})^T$. Suppose at the i^{th} step the estimates of the parameters $\alpha, \lambda_0, \lambda_1$, and λ_2 are $\alpha^{(i)}, \lambda_0^{(i)}, \lambda_1^{(i)}$, and $\lambda_2^{(i)}$, respectively. Then the $(i+1)^{th}$ step of the EM algorithm is obtained as follows

1. Compute $\mu_1, \mu_2, \nu_1, \nu_2$ using $\alpha^{(i)}, \lambda_0^{(i)}, \lambda_1^{(i)}, \lambda_2^{(i)}$.

2. For fixed $\lambda_0^{(i)}, \lambda_1^{(i)}, \lambda_2^{(i)}$, find $\alpha^{(i+1)}$ by solving the fixed point equation $g_\alpha(\alpha, \lambda_0^{(i)}, \lambda_1^{(i)}, \lambda_2^{(i)}) = \alpha$ using the initial estimate as $\alpha^{(i)}$.

3. Given $\alpha^{(i+1)}$, compute $\lambda_0^{(i+1)} = N_0 \left[\sum_{i \in I_0} t_i^{\alpha^{(i+1)}} + \sum_{i \in I_1} t_{2i}^{\alpha^{(i+1)}} + \sum_{i \in I_2} t_{1i}^{\alpha^{(i+1)}} \right]^{-1}$.

4. For fixed $\alpha^{(i+1)}$, starting with $\lambda_1^{(i)}$, find $\lambda_1^{(i+1)}$ by solving $g_{\lambda_1}(\alpha^{(i+1)}, \lambda_1) = \lambda_1$ iteratively.

5. Find $\lambda_2^{(i+1)}$, by solving $g_{\lambda_2}(\alpha^{(i+1)}, \lambda_2) = \lambda_2$ for fixed $\alpha^{(i+1)}$, similarly as step 4.

6. Repeat steps 1-5 using $\alpha^{(i+1)}, \lambda_0^{(i+1)}, \lambda_1^{(i+1)}, \lambda_2^{(i+1)}$.

This version of the EM algorithm is called ECM (expectation-conditional maximization) algorithm. Steps 1-5 describe one iteration of the algorithm and individual steps 2, 4, and 5 corresponds to fixed point iterations of α, λ_1, and λ_2 respectively. A stopping criterion is indicated in the next section.

4 Numerical Experiments

In this section, we present results of numerical experiments to see how the proposed EM algorithm performs for different sample sizes and different parameter values when a certain percentage of data is randomly censored. For conducting the experiment, we assume that the pair of censoring random variables (Z_1, Z_2) is distributed as MOBW with the same shape parameter α as the original pair (X_1, X_2) and different scale parameters, say, λ_0^*, λ_1^*, and λ_2^*. If $(X_1, X_2) \sim$ MOBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$ and $(Z_1, Z_2) \sim$ MOBW$(\alpha, \lambda_0^*, \lambda_1^*, \lambda_2^*)$,

$$P(X_1 > Z_1) = \frac{\lambda_0^* + \lambda_1^*}{\lambda_0 + \lambda_1 + \lambda_0^* + \lambda_1^*}, \qquad \text{and} \qquad P(X_2 > Z_2) = \frac{\lambda_0^* + \lambda_2^*}{\lambda_0 + \lambda_2 + \lambda_0^* + \lambda_2^*}.$$

Since both pairs (X_1, X_2) and (Z_1, Z_2) have the same shape parameter, this ensures that the percentage of censoring does not depend on α. The two probabilities are equal if $\lambda_1 = \lambda_2$ and $\lambda_1^* = \lambda_2^*$.

We replicate the experiment 5000 times. In Appendix B, we have provided the observed Fisher information matrix and we have used it for interval estimation. This matrix has been derived using the procedure given by Louis (1982).

Table 1: Average estimates, mean squared errors, coverage probabilities and average lengths of the confidence interval when $\alpha = .25$

Sample Size →		$N = 50$		$N = 100$	
Parameter	Censoring	AVEST (MSE)	AVLEN (COVP)	AVEST (MSE)	AVLEN (COVP)
α	9%	.256187 (5.25098e-4)	8.50074e-2 (.9386)	.253493 (2.37826e-4)	5.87637e-2 (.9364)
	20%	.251962 (1.13729e-3)	8.60392e-2 (.9014)	.254354 (2.67290e-4)	5.98616e-2 (.9114)
λ_0	9%	1.040218 (6.07387e-2)	.747532 (.8840)	1.040594 (3.01952e-2)	.528609 (.8874)
	20%	.954804 (5.88477e-2)	.746050 (.8740)	.989496 (2.42145e-2)	.545141 (.9132)
λ_1	9%	1.126363 (8.44284e-2)	.907930 (.9218)	1.118709 (4.78498e-2)	.638749 (.8962)
	20%	1.126483 (8.79527e-2)	.939340 (.9316)	1.153017 (5.60403e-2)	.681842 (.8922)
λ_2	9%	1.113779 (7.70360e-2)	.899759 (.9290)	1.105677 (4.23663e-2)	.633200 (.9174)
	20%	1.130571 (9.18659e-2)	.940163 (.9326)	1.150479 (5.49980e-2)	.680465 (.8964)

We have carried out the experiment for various choices of four parameters and the sample size. However, only a few cases are reported for illustration. The average estimate, mean squared error, average length of confidence intervals and coverage probability at 95% nominal level are reported in Tables 1-2 for the case with fixed $\lambda_0 = \lambda_1 = \lambda_2 = 1.0$ and varying shape parameter α =0.25, 1.0 and sample size $n = 50, 100$. We have used $(\lambda_0^*, \lambda_1^*, \lambda_2^*) = (.1, .1, .1)$ and $(.25, .25, .25)$, that is, the parameters of the censoring variables are also equal. In case of $\lambda_0^* = \lambda_1^* = \lambda_2^* = .1$, 9% data are censored, whereas 20% data are censored when $\lambda_0^* = \lambda_1^* = \lambda_2^* = .25$. In order to implement the proposed EM algorithm, we have used the initial estimates of α, λ_0, λ_1 and λ_2 as $.5, .5, .5$, and $.5$ in each case. We observe that the change of initial estimates gives similar results. To solve any fixed point type equation in steps 2, 4 or 5, we stop the iteration if $|\theta_k^i - \theta_{k-1}^i| < 10^{-6}$, where θ_k^i stands for the k^{th} fixed point iteration of α, λ_1 or λ_2 in overall

Table 2: Average estimates, mean squared errors, coverage probabilities and average lengths of the confidence interval when $\alpha = 1.0$

Sample Size $\rightarrow$		$N = 50$		$N = 100$	
Parameter	Censoring	AVEST (MSE)	AVLEN (COVP)	AVEST (MSE)	AVLEN (COVP)
α	9%	1.024758 (8.40164e-3)	.339463 (.9386)	1.013977 (3.80584e-3)	.235242 (.9364)
	20%	1.017553 (9.10435e-3)	.346425 (.9098)	1.017394 (4.28030e-3)	.240189 (.9120)
λ_0	9%	1.040250 (6.08027e-2)	.747102 (.8836)	1.040692 (3.02421e-2)	.528677 (.8872)
	20%	.962852 (4.98865e-2)	.751932 (.8826)	.989177 (2.42485e-2)	.545180 (.9132)
λ_1	9%	1.126531 (8.43788e-2)	.907647 (.9220)	1.11869 (4.78454e-2)	.638820 (.8962)
	20%	1.137626 (7.93969e-2)	.947190 (.9402)	1.153074 (5.62133e-2)	.682229 (.8918)
λ_2	9%	1.113871 (7.71294e-2)	.899494 (.9286)	1.105683 (4.23678e-2)	.633325 (.9174)
	20%	1.141197 (8.37064e-2)	.948062 (.9408)	1.150473 (5.48781e-2)	.680898 (.8970)

i^{th} iteration. The algorithm stops if

$$|\alpha^{(i+1)} - \alpha^{(i)}| + |\lambda_0^{(i+1)} - \lambda_0^{(i)}| + |\lambda_1^{(i+1)} - \lambda_1^{(i)}| + |\lambda_2^{(i+1)} - \lambda_2^{(i)}| \leq 10^{-5}.$$

Some of the salient features of the numerical experiments based on Tables 1-2 are given below.

(i) We observe that the average estimators of all the four parameters $\alpha, \lambda_0, \lambda_1, \lambda_2$ are very close to the true values for both choices of the shape parameter α. The estimators have a positive bias in almost all cases. However, the estimator for λ_0 has a negative bias in case of 20% censoring. The results are similar for sample sizes 50 and 100.

(ii) The mean square error of the estimators decreases with increase in sample size. The value of α and the amount of censoring makes no visible change in its numerical value.

(iii) When $n = 100$ the average lengths of confidence intervals are smaller and the coverage probabilities are slightly higher compared to the case when $n = 50$.

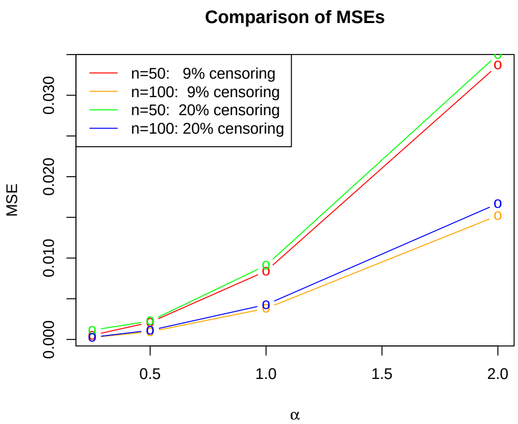

Figure 1: Mean Squared Errors of M.L.E. of α.

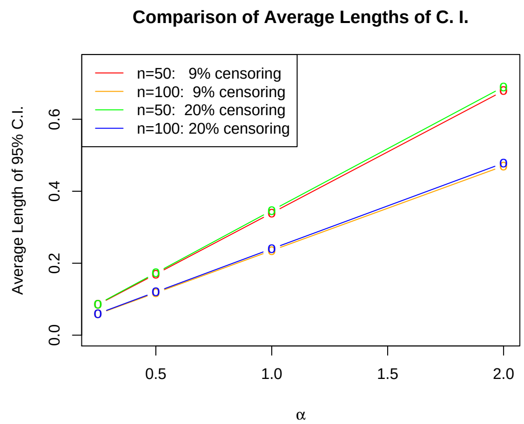

Figure 2: Average Lengths of Confidence intervals of M.L.E. of α.

Similarly when the amount of censoring increases the average length of confidence intervals increases and coverage probability decreases. The coverage probabilities of confidence intervals for all parameters are not influenced by values of α.

(iv) It should be noted that the case $\alpha = 1.0$ refers to MOBE distribution.

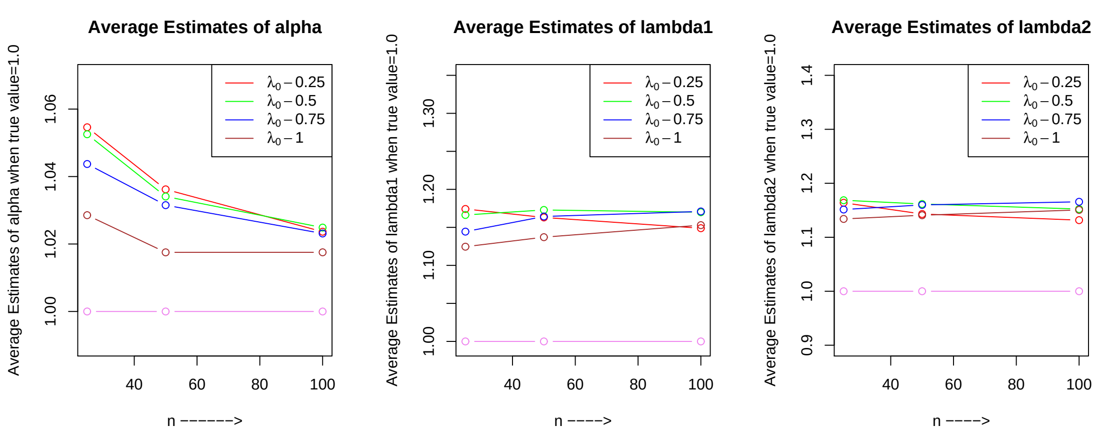

Figure 3: Average Estimators: $\lambda_1 = \lambda_2 = \alpha = 1.0$ and $\lambda_2 = .25, .5, .75, 1.0$

Some of the observations have been presented graphically in Figures 1 and 2.

A few other results of the numerical experiments are given in Figures 3-5. In all the cases reported the values of $\alpha = \lambda_1 = \lambda_2 = 1.0$ and that of λ_0 varies from $.25, .5, .75$, and 1.0.

The censoring distribution is MOBW $(1.0, .25, .25, .25)$. Note that the percentage of censoring changes with change in values of λ_0. Hence the different values of λ_0 indicate different levels of censoring. Rsults for sample sizes $n = 25, 50, 100$ are reported. The results are similar to the ones discussed above for Tables 1-2.

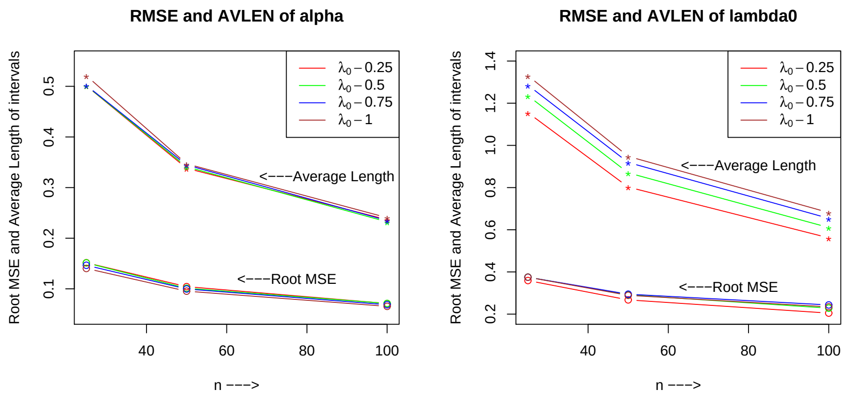

Figure 4: Root MSE and Average Length of confidence Intervals of α and λ_0.

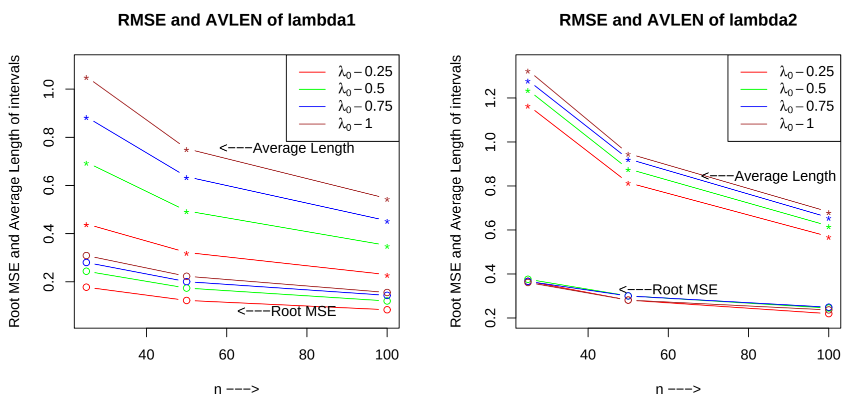

Figure 5: Root MSE and Average Length of confidence Intervals of α and λ_0.

5 Data Analysis

In this section, we have analyzed the soccer data for the years 2004-05 and 2005-06, considered in Meintanis (2007). Kundu and Dey (2009) showed that MOBW distribution fits well to this

Table 3: The points estimates and confidence interval for football data .

	10% Censoring		5% Censoring	
Parameter	Point Est.	Conf. Int.	Point Est.	Conf. Int.
λ_0	2.532	(1.377, 3.686)	2.643	(1.532, 3.753)
λ_1	1.178	(.421, 1.935)	1.299	(.508, 2.089)
λ_2	3.227	(1.583, 4.871)	2.986	(1.540, 4.432)
α	1.707	(1.348, 2.066)	1.709	(1.371, 2.047)

data. We have introduced censoring artificially and then estimated the parameters. This brings out the effect of censoring on the estimates of the parameters.

The data (X_1, X_2) contain 37 data points. We assume that the pair (X_1, X_2) has MOBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$. The pair of censoring random variables (Z_1, Z_2) has MOBW$(\tilde{\alpha}, \lambda_0^*, \lambda_1^*, \lambda_2^*)$. In order to ensure that $P(X_1 > Z_1) = P(X_2 > Z_2) = .1$, we take $(\lambda_0^*, \lambda_1^*, \lambda_2^*) = (.2, .23, .41)$ and $\tilde{\alpha}, \tilde{\lambda}_0, \tilde{\lambda}_1, \tilde{\lambda}_2$ are the estimates of $\alpha, \lambda_0, \lambda_1, \lambda_2$, obtained by Kundu and Dey (2009). Similarly $(\lambda_0^*, \lambda_1^*, \lambda_2^*) = (.1, .12, .19)$ ensures that $P(X_1 > Z_1) = P(X_2 > Z_2) = .05$. We have used the proposed EM algorithm to estimate the unknown parameters and the initial estimates used for $\alpha, \lambda_0, \lambda_1, \lambda_2$, respectively, were 1.67, 2.7, 1.2 and 2.7 in both the cases. The point estimates and the confidence intervals for 10 % and 5 % censoring are reported in Table 3. Higher censoring is not considered because the sample size is just 37.

6 Summary and Future Work

In this paper we have considered the M.L.E.'s of the four parameters of MOBW distribution when both components of the bivariate variable are subject to random censoring. Since the estimators can not be expressed in a closed form, we suggest the use of expectation-conditional maximization algorithm. We have clearly written the steps involved in the iteration procedure.

The simulations were carried out for several choices of the parameters but only a few cases were reported for illustration. The results indicate that the EM algorithm performs very well for sample sizes 25, 50 and 100 and also for various levels of random censoring that we have studied (9% and 20%).

Our program, has been run to include even higher censoring. As stated in section 2, the data is classified into 12 classes - 4 each in I_0, I_1, I_2. The number of members in each is given by $n_{ij}^k,\ i, j = 0, 1,\ k = 0, 1, 2$. In case of high censoring $n_{11}^k, k = 0, 1, 2$ takes very small values and other n_{ij}^k 's take relatively large values. In such cases, the algorithm does not work efficiently.

The asymptotic confidence intervals give accurate results even for moderate sample sizes and hence can be used for testing purposes. For example, one can test whether the underlying bivariate distribution is MOBE or not, that is $\alpha = 1$ or not.

The case $\alpha = 1$ corresponds to MOBE. Hence the proposed algorithm can also be used to estimate the three parameters of MOBE when pairs of random variables are subject to random censoring.

The above procedures can also be extended to other bivariate distributions subjected to random censoring. Work for Bivariate Generalized Exponential and other bivariate distributions, commonly studied in survival analysis, are in process and will be reported elsewhere.

Appendix A - Details of likelihood function

The contribution to the likelihood on the set I_0 is

$$\begin{aligned} & \prod_{i \in I_0} L_0(t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i}) \\ = & \prod_{i \in I_0} [\frac{\lambda_0}{\lambda} f_{WE}(t_i; \alpha, \lambda)]^{\delta_{1i}\delta_{2i}} [f_{WE}(t_i; \alpha, \lambda_1) S_{WE}(t_i; \alpha, \lambda_0 + \lambda_2)]^{\delta_{1i}(1-\delta_{2i})} \\ & [f_{WE}(t_i; \alpha, \lambda_2) S_{WE}(t_i; \alpha, \lambda_0 + \lambda_1)]^{(1-\delta_{1i})\delta_{2i}} [S_{WE}(t_i; \alpha, \lambda)]^{(1-\delta_{1i})(1-\delta_{2i})}. \end{aligned}$$

And the contribution to the likelihood on the set I_1 is

$$\begin{aligned} & \prod_{i \in I_1} L_1(t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i}) \\ = & \prod_{i \in I_1} [f_{WE}(t_{1i}; \alpha, \lambda_1) f_{WE}(t_{2i}; \alpha, \lambda_0 + \lambda_2)]^{\delta_{1i}\delta_{2i}} [f_{WE}(t_{1i}; \alpha, \lambda_1) \\ & S_{WE}(t_{2i}; \alpha, \lambda_0 + \lambda_2)]^{\delta_{1i}(1-\delta_{2i})} [f_{WE}(t_{2i}; \alpha, \lambda_0 + \lambda_2)[S_{WE}(t_{1i}; \alpha, \lambda_1) \\ & -S_{WE}(t_{2i}; \alpha, \lambda_1)]^{(1-\delta_{1i})\delta_{2i}} [S_{WE}(t_{1i}; \alpha, \lambda_1) S_{WE}(t_{2i}; \alpha, \lambda_0 + \lambda_2)]^{(1-\delta_{1i})(1-\delta_{2i})}. \end{aligned}$$

Finally, the contribution to the likelihood on the set I_2 is

$$\begin{aligned} & \prod_{i \in I_2} L_2(t_{1i}, \delta_{1i}; t_{2i}, \delta_{2i}) \\ = & \prod_{i \in I_2} [f_{WE}(t_{1i}; \alpha, \lambda_0 + \lambda_1) f_{WE}(t_{2i}; \alpha, \lambda_2)]^{\delta_{1i}\delta_{2i}} \\ & [f_{WE}(t_{1i}; \alpha, \lambda_0 + \lambda_1)[S_{WE}(t_{2i}; \alpha, \lambda_2) - S_{WE}(t_{1i}, \alpha, \lambda_2)]^{\delta_{1i}(1-\delta_{2i})} \end{aligned}$$

$$[S_{WE}(t_{1i};\alpha,\lambda_0+\lambda_1)f_{WE}(t_{2i};\alpha,\lambda_2)]^{(1-\delta_{1i})\delta_{2i}}$$
$$[S_{WE}(t_{1i};\alpha,\lambda_0+\lambda_1)S_{WE}(t_{2i};\alpha,\lambda_2)]^{(1-\delta_{1i})(1-\delta_{2i})}.$$

Appendix B - Observed Fisher Information Matrix of M.L.E.'s

In this section, the observed Fisher information matrix is provided. We follow the procedure described in Louis (1982), which is used when the EM algorithm is applied to obtain the M.L.E.'s in case of incomplete data problem. The observed Fisher information matrix is used for computation of asymptotic confidence intervals in numerical experiment in section 4. We denote $\mathbf{g} = (g_1, g_2, g_3, g_4)^T$ as the gradient vector and $\mathbf{H} = ((H_{ij}))$ as the Hessian matrix of the pseudo log-likelihood function defined in (2). Then using N_0, N_1 and N_2, the elements of vector $\mathbf{g}$ are as follows;

$$g_1 = \frac{1}{\alpha}N_3 + h_1(\alpha,\lambda_0,\lambda_1,\lambda_2), \quad g_2 = \frac{1}{\lambda_0}N_0 - \left[\sum_{i\in I_0} t_i^\alpha + \sum_{i\in I_1} t_{2i}^\alpha + \sum_{i\in I_2} t_{1i}^\alpha\right],$$
$$g_3 = \frac{1}{\lambda_1}N_1 - h_2(\alpha,\lambda_1), \qquad g_4 = \frac{1}{\lambda_2}N_2 - h_3(\alpha,\lambda_2),$$

$$\begin{aligned}
h_1(\alpha,\lambda_0,\lambda_1,\lambda_2) \quad = \quad & \left[\sum_{i\in I_0}(\delta_{i1}+\delta_{i2}-\delta_{1i}\delta_{2i})\log t_i + \sum_{i\in I_1\cup I_2}(\delta_{1i}\log t_{1i}+\delta_{2i}\log t_{2i})\right] \\
& -\lambda_0\left[\sum_{i\in I_0} t_i^\alpha \log t_i + \sum_{i\in I_1} t_{2i}^\alpha \log t_{2i} + \sum_{i\in I_2} t_{1i}^\alpha \log t_{1i}\right] \\
& -\lambda_1\left[\sum_{i\in I_0} t_i^\alpha \log t_i + \sum_{i\in I_1}(1-\delta_{2i}+\delta_{1i}\delta_{2i}) t_{1i}^\alpha \log t_{1i} + \sum_{i\in I_2} t_{1i}^\alpha \log t_{1i}\right] \\
& -\lambda_2\left[\sum_{i\in I_0} t_i^\alpha \log t_i + \sum_{i\in I_1} t_{2i}^\alpha \log t_{2i} + \sum_{i\in I_2}(1-\delta_{1i}+\delta_{1i}\delta_{2i}) t_{2i}^\alpha \log t_{2i}\right] \\
& +\sum_{i\in I_1}(1-\delta_{1i})\delta_{2i}\frac{\lambda_1(t_{2i}^\alpha e^{-\lambda_1 t_{2i}^\alpha}\log t_{2i} - t_{1i}^\alpha e^{-\lambda_1 t_{1i}^\alpha}\log t_{1i})}{e^{-\lambda_1 t_{1i}^\alpha} - e^{-\lambda_1 t_{2i}^\alpha}} \\
& +\sum_{i\in I_2}\delta_{1i}(1-\delta_{2i})\frac{\lambda_2(t_{1i}^\alpha e^{-\lambda_2 t_{1i}^\alpha}\log t_{1i} - t_{2i}^\alpha e^{-\lambda_2 t_{2i}^\alpha}\log t_{2i})}{e^{-\lambda_2 t_{2i}^\alpha} - e^{-\lambda_2 t_{1i}^\alpha}},
\end{aligned}$$

$$\begin{aligned}
h_2(\alpha,\lambda_1) \quad = \quad & \left[\sum_{i\in I_0} t_i^\alpha + \sum_{i\in I_1}(1-\delta_{2i}+\delta_{1i}\delta_{2i}) t_{1i}^\alpha + \sum_{i\in I_2} t_{1i}^\alpha\right] \\
& +\sum_{i\in I_1}(1-\delta_{1i})\delta_{2i}\frac{t_{2i}^\alpha e^{-\lambda_1 t_{2i}^\alpha} - t_{1i}^\alpha e^{-\lambda_1 t_{1i}^\alpha}}{e^{-\lambda_1 t_{1i}^\alpha} - e^{-\lambda_1 t_{2i}^\alpha}},
\end{aligned}$$

$$
\begin{aligned}
h_3(\alpha,\lambda_2) &= \left[\sum_{i\in I_0} t_i^\alpha + \sum_{i\in I_1} t_{2i}^\alpha + \sum_{i\in I_2}(1-\delta_{1i}+\delta_{1i}\delta_{2i})t_{2i}^\alpha\right] \\
&\quad + \sum_{i\in I_2}\delta_{1i}(1-\delta_{2i})\frac{t_{1i}^\alpha e^{-\lambda_2 t_{1i}^\alpha} - t_{2i}^\alpha e^{-\lambda_2 t_{2i}^\alpha}}{e^{-\lambda_2 t_{2i}^\alpha} - e^{-\lambda_2 t_{1i}^\alpha}}.
\end{aligned}
$$

The Hessian matrix $\mathbf{H}$ is symmetric, so $H_{ij} = H_{ji}, i > j$ and in the following, the elements are given.

$$
\begin{aligned}
H_{11} &= -\frac{1}{\alpha^2}N_3 - \lambda_0\left[\sum_{i\in I_0} t_i^\alpha(\log t_i)^2 + \sum_{i\in I_1} t_{2i}^\alpha(\log t_{2i})^2 + \sum_{i\in I_2} t_{1i}^\alpha(\log t_{1i})^2\right] \\
&\quad -\lambda_1\left[\sum_{i\in I_0} t_i^\alpha(\log t_i)^2 + \sum_{i\in I_1}(1-\delta_{2i}+\delta_{1i}\delta_{2i})t_{1i}^\alpha(\log t_{1i})^2 + \sum_{i\in I_2} t_{1i}^\alpha(\log t_{1i})^2\right] \\
&\quad -\lambda_2\left[\sum_{i\in I_0} t_i^\alpha(\log t_i)^2 + \sum_{i\in I_1} t_{2i}^\alpha(\log t_{2i})^2 + \sum_{i\in I_2}(1-\delta_{1i}+\delta_{1i}\delta_{2i})t_{2i}^\alpha(\log t_{2i})^2\right] \\
&\quad +\sum_{i\in I_1}(1-\delta_{1i})\delta_{2i}\lambda_1\left[\lambda_1 e^{-\lambda_1 t_{1i}^\alpha}(t_{1i}^\alpha\log t_{1i})^2 - e^{-\lambda_1 t_{1i}^\alpha}t_{1i}^\alpha(\log t_{1i})^2 - \lambda_1 e^{-\lambda_1 t_{2i}^\alpha}(t_{2i}^\alpha\log t_{2i})^2\right. \\
&\quad \left.+e^{-\lambda_1 t_{2i}^\alpha}t_{2i}^\alpha(\log t_{2i})^2 - \lambda_1(e^{-\lambda_1 t_{2i}^\alpha}t_{2i}^\alpha\log t_{2i} - e^{-\lambda_1 t_{1i}^\alpha}t_{1i}^\alpha\log t_{1i})^2\right] / \left(e^{-\lambda_1 t_{1i}^\alpha} - e^{-\lambda_1 t_{2i}^\alpha}\right)^2 \\
&\quad +\sum_{i\in I_2}\delta_{1i}(1-\delta_{2i})\lambda_2\left[\lambda_2 e^{-\lambda_2 t_{2i}^\alpha}(t_{2i}^\alpha\log t_{2i})^2 - e^{-\lambda_2 t_{2i}^\alpha}t_{2i}^\alpha(\log t_{2i})^2 - \lambda_2 e^{-\lambda_2 t_{1i}^\alpha}(t_{1i}^\alpha\log t_{1i})^2\right. \\
&\quad \left.+e^{-\lambda_2 t_{1i}^\alpha}t_{1i}^\alpha(\log t_{1i})^2 - \lambda_2(e^{-\lambda_2 t_{1i}^\alpha}t_{1i}^\alpha\log t_{1i} - e^{-\lambda_2 t_{2i}^\alpha}t_{2i}^\alpha\log t_{2i})^2\right] / \left(e^{-\lambda_2 t_{2i}^\alpha} - e^{-\lambda_2 t_{1i}^\alpha}\right)^2,
\end{aligned}
$$

$$
\begin{aligned}
H_{12} &= -\left[\sum_{i\in I_0} t_i^\alpha\log t_i + \sum_{i\in I_1} t_{2i}^\alpha\log t_{2i} + \sum_{i\in I_2} t_{1i}^\alpha\log t_{1i}\right], \\
H_{13} &= -\left[\sum_{i\in I_0} t_i^\alpha\log t_i + \sum_{i\in I_1}(1-\delta_{2i}+\delta_{1i}\delta_{2i})t_{1i}^\alpha\log t_{1i} + \sum_{i\in I_2} t_{1i}^\alpha\log t_{1i}\right] + \sum_{i\in I_1}(1-\delta_{1i})\delta_{i2}\times \\
&\quad \left[\frac{\lambda_1 t_{1i}^{2\alpha}e^{-\lambda_1 t_{1i}^\alpha}\log t_{1i} - t_{1i}^\alpha e^{-\lambda_1 t_{1i}^\alpha}\log t_{1i} - \lambda_1 t_{2i}^{2\alpha}e^{-\lambda_1 t_{2i}^\alpha}\log t_{2i} + t_{2i}^\alpha e^{-\lambda_1 t_{2i}^\alpha}\log t_{2i}}{e^{-\lambda_1 t_{1i}^\alpha} - e^{-\lambda_1 t_{2i}^\alpha}}\right. \\
&\quad \left.-\frac{\lambda_1(t_{2i}^\alpha e^{-\lambda_1 t_{2i}^\alpha} - t_{1i}^\alpha e^{-\lambda_1 t_{1i}^\alpha})(t_{2i}^\alpha e^{-\lambda_1 t_{2i}^\alpha}\log t_{2i} - t_{1i}^\alpha e^{-\lambda_1 t_{1i}^\alpha}\log t_{1i})}{(e^{-\lambda_1 t_{1i}^\alpha} - e^{-\lambda_1 t_{2i}^\alpha})^2}\right], \\
H_{14} &= -\left[\sum_{i\in I_0} t_i^\alpha\log t_i + \sum_{i\in I_1} t_{2i}^\alpha\log t_{2i} + \sum_{i\in I_2}(1-\delta_{1i}+\delta_{1i}\delta_{2i})t_{2i}^\alpha\log t_{2i}\right] + \sum_{i\in I_2}\delta_{1i}(1-\delta_{i2})\times \\
&\quad \left[\frac{\lambda_2 t_{2i}^{2\alpha}e^{-\lambda_2 t_{2i}^\alpha}\log t_{2i} - t_{2i}^\alpha e^{-\lambda_2 t_{2i}^\alpha}\log t_{2i} - \lambda_2 t_{1i}^{2\alpha}e^{-\lambda_2 t_{1i}^\alpha}\log t_{1i} + t_{1i}^\alpha e^{-\lambda_2 t_{1i}^\alpha}\log t_{1i}}{e^{-\lambda_2 t_{2i}^\alpha} - e^{-\lambda_2 t_{1i}^\alpha}}\right.
\end{aligned}
$$

$$
\left. - \frac{\lambda_2 (t_{1i}^{\alpha} e^{-\lambda_2 t_{1i}^{\alpha}} - t_{2i}^{\alpha} e^{-\lambda_2 t_{2i}^{\alpha}})(t_{1i}^{\alpha} e^{-\lambda_2 t_{1i}^{\alpha}} \log t_{1i} - t_{2i}^{\alpha} e^{-\lambda_2 t_{2i}^{\alpha}} \log t_{2i})}{(e^{-\lambda_2 t_{2i}^{\alpha}} - e^{-\lambda_2 t_{1i}^{\alpha}})^2} \right],
$$

$$
\begin{aligned}
H_{22} &= -\frac{1}{\lambda_0^2} N_0, \qquad H_{23} = H_{24} = 0, \\
H_{33} &= -\frac{1}{\lambda_1^2} N_1 + \sum_{i \in I_1} (1 - \delta_{1i}) \delta_{2i} \left[\frac{(t_{1i}^{2\alpha} e^{-\lambda_1 t_{1i}^{\alpha}} - t_{2i}^{2\alpha} e^{-\lambda_1 t_{2i}^{\alpha}})}{(e^{-\lambda_1 t_{1i}^{\alpha}} - e^{-\lambda_1 t_{2i}^{\alpha}})} - \frac{(t_{2i}^{\alpha} e^{-\lambda_1 t_{2i}^{\alpha}} - t_{1i}^{\alpha} e^{-\lambda_1 t_{1i}^{\alpha}})^2}{(e^{-\lambda_1 t_{1i}^{\alpha}} - e^{-\lambda_1 t_{2i}^{\alpha}})^2} \right], \\
H_{34} &= 0, \\
H_{44} &= -\frac{1}{\lambda_2^2} N_2 + \sum_{i \in I_2} \delta_{1i} (1 - \delta_{2i}) \left[\frac{(t_{2i}^{2\alpha} e^{-\lambda_2 t_{2i}^{\alpha}} - t_{1i}^{2\alpha} e^{-\lambda_2 t_{1i}^{\alpha}})}{(e^{-\lambda_2 t_{2i}^{\alpha}} - e^{-\lambda_2 t_{1i}^{\alpha}})} - \frac{(t_{1i}^{\alpha} e^{-\lambda_2 t_{1i}^{\alpha}} - t_{2i}^{\alpha} e^{-\lambda_2 t_{2i}^{\alpha}})^2}{(e^{-\lambda_2 t_{2i}^{\alpha}} - e^{-\lambda_2 t_{1i}^{\alpha}})^2} \right].
\end{aligned}
$$

The observed Fisher information matrix is given by $\mathbf{H} - \mathbf{g}\mathbf{g}^T$.

Acknowledgements

We thank the referees for their careful reading and constructive suggestions.

References

[1] Bemis, B., Bain, L.J. and Higgins, J.J. (1972). Estimation and hypothesis testing for the parameters of a bivariate exponential distribution. *Journal of the American Statistical Association*, **67**, 927-929.

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

[3] Gupta, R.D. and Kundu, D. (2003). Discriminating Between the Weibull and the GE Distributions. *Computational Statistics and Data Analysis*, **43**, 179-196.

[4] Hanagal, D.D. (2005). A bivariate Weibull regression model. *Economic Quality Control*, **20**, 143-150.

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

[6] Houggard, P., Harvald, B. and Holm, N.V. (1992). Measuring the similarities between the lifetimes of adult Danish twins born between 1881-1930. *J. Amer. Statistit. Assoc.*, **87**, 17-24.

[7] 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*, **53**, no. 4, 956-965.

[8] Kundu, D. and Gupta, R.D. (2006). Estimation of $P(Y < X)$ for Weibull distribution. *IEEE Trans. Reliability*, **55**, 270-280.

[9] Lin, D.Y., Sun, W. and Ying, Z. (1999). Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika*, **86**, 59-70.

[10] Louis, T. A. (1982), Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 226-233.

[11] Lu, Jye-Chyi (1989). Weibull extension of the Freund and Marshall-Olkin bivariate exponential model. *IEEE Trans. Reliability*, **38**, 615-619.

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

[13] Marshall, A.W. and Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statistit. Assoc.*, **62**, 30-44.

[14] Meintanis, S.G. (2007). Test of fit for Marshall-Olkin distributions with applications. *Journal of Statistical Planning and Inference*, **137**, 3954-3963.