Estimating the Parameters of the Marshall Olkin Bivariate Weibull Distribution by EM Algorithm

Debasis Kundu[†] & Arabin Kumar Dey[†]

Abstract

In this paper we consider the Marshall-Olkin bivariate Weibull distribution. The Marshall-Olkin bivariate Weibull distribution is a singular distribution, whose both the marginals are univariate Weibull distributions. This is a generalization of the Marshall-Olkin bivariate exponential distribution. The cumulative joint distribution of the Marshall-Olkin bivariate Weibull distribution is a mixture of an absolute continuous distribution function and a singular distribution function. This distribution has four unknown parameters and it is observed that the maximum likelihood estimators of the unknown parameters can not be obtained in explicit forms. In this paper we discuss about the computation of the maximum likelihood estimators of the unknown parameters using EM algorithm. We perform some simulations to see the performances of the EM algorithm and re-analyze one data set for illustrative purpose.

Keywords: Bivariate model; Joint probability density function; Maximum likelihood estimators; Fisher information matrix; EM algorithm; Pseudo likelihood function.

[†] Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, Pin 208016, INDIA.

Corresponding Author: Debasis Kundu, Phone no. 91-512-2597141, Fax No. 91-512-2597500, e-mail: kundu@iitk.ac.in.

1 Introduction

The Marshall-Olkin bivariate exponential distribution, see [14], is a singular distribution. In this case, both the marginals have exponential distributions, but they can be equal with a positive probability. Because of that reason, if in a bivariate data set, for some cases two components take equal values, the Marshal-Olkin bivariate exponential (MOBE) distribution can be used quite effectively to analyze such data set. Since, the MOBE distribution has exponential marginals, if the bivariate data indicate unimodal marginal probability density function or non-constant hazard function, then MOBE distribution may not be appropriate. Because of this restriction, Marshall and Olkin [14] suggested a more flexible bivariate Weibull (MOBW) distribution, where the marginals are Weibull distributions and it can be obtained along the same line as the MOBE model. This model can be observed as a shock model when the shocks are arriving as a non-homogeneous Poisson process. In fact most of the other interpretations which are valid for MOBE distribution can be easily generalized for MOBW distribution also. Clearly, MOBW model is more flexible than the MOBE model, because of the presence of the shape parameter. Moreover, this model can be used quite effectively if the bivariate data indicate unimodal marginal probability density function or non-constant hazard function.

Lu [13] considered the MOBW model and proposed the Bayes estimates of the unknown parameters. For some of the related work in this connection the readers are referred to Patra and Dey [16], Lu [12], Hanagal [6, 7] and the references cited there. It may be mentioned that although extensive work has been done on MOBE model but not that much of work has been done for the MOBW model. One of the reason might be due to the computational complexity involved in finding the estimates of the MOBW model parameters.

The main aim of this paper is to consider the efficient computation of the maximum likelihood estimators (MLEs) of the MOBW parameters. This model has four unknown parameters, and the MLEs of the unknown parameters do not always exist, see for example Bemis *et al.* [2], and even if they exist they can not be obtained in explicit forms as expected. It is observed that the problems can be treated as a missing value problem and the EM algorithm can be used quite effectively to compute the MLEs by solving a one dimensional optimization problem at each iteration. We have provided a simple iterative procedure to solve the one dimensional optimization procedure also. Moreover, using the idea of Louis [11], from the EM algorithm the observed Fisher information matrix can be easily obtained and they can be used for constructing asymptotic confidence intervals of the unknown parameters and for testing purposes also. Since MOBE is a special case of the MOBW distribution, our method can be applied for the MOBE model also. It may be mentioned at this point that Karlis [8] also developed an EM algorithm for the computation of the MLEs of the MOBE distribution, but the two ideas are quite different. Moreover, it is not immediate how Karlis's method can be extended for the MOBW model.

We have performed some simulations to see the performances of the EM algorithm and the performances are quite satisfactory. Recently, Meintanis [15] analyzed one data set using MOBE distribution. We have re-analyzed the same data set using MOBW distribution. It is observed that the proposed EM algorithm is working very well in this case and MOBW provides a better fit than the MOBE distribution. We have a provided a justification for that also. Finally we have provided the multivariate generalization of our proposed method.

The rest of the paper is organized as follows. In section 2, we describe the models and provide some basic properties. The EM algorithm for MOBW is discussed in section 3. Simulation results and data analysis are presented in section 4 and section 5 respectively. We discuss the multivariate generalization in section 6 and finally conclude the paper in section 7.

2 Marshall-Olkin Bivariate Weibull Distribution

It is assumed that the univariate Weibull distribution with the shape parameter $\alpha > 0$ and the scale parameter $\theta > 0$ has the following probability density function (PDF), cumulative distribution function (CDF) and survival function (SE) for $x > 0$;

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

respectively. From now on a Weibull distribution with the PDF (1) will be denoted by WE(α,θ). Suppose U_0 follows ($\sim$) WE(α,λ_0), $U_1 \sim$ WE(α,λ_1), $U_2 \sim$ WE(α,λ_2) and they are independent. Define $X_1 = \min\{U_0,U_1\}$, and $X_2 = \min\{U_0,U_2\}$, then the bivariate vector (X_1,X_2) has the MOBW distribution with the parameters $\alpha,\lambda_0,\lambda_1,\lambda_2$ and it will be denoted from now on as MOBW$(\alpha,\lambda_0,\lambda_1,\lambda_2)$. When $\alpha = 1$, it coincides with the MOBE model with parameters $\lambda_0,\lambda_1,\lambda_2$, which will be denoted by MOBE$(\lambda_0,\lambda_1,\lambda_2)$.

If $(X_1,X_2) \sim$ MOBW$(\alpha,\lambda_0,\lambda_1,\lambda_2)$ then their joint survival function takes the following form for $z = \max\{x_1,x_2\}$;

$$\begin{aligned} S_{X_1,X_2}(x_1,x_2) &= P(X_1 > x_1, X_2 > x_2) = P(U_0 > z, U_1 > x_1, U_2 > x_2) \\ &= S_{WE}(x_1;\alpha,\lambda_1)S_{WE}(x_2;\alpha,\lambda_2)S_{WE}(z;\alpha,\lambda_0) \\ &= \begin{cases} S_{WE}(x_1;\alpha,\lambda_1)S_{WE}(x_2;\alpha,\lambda_0+\lambda_2) & \text{if} \quad x_1 < x_2 \\ S_{WE}(x_1;\alpha,\lambda_0+\lambda_1)S_{WE}(x_2;\alpha,\lambda_2) & \text{if} \quad x_1 > x_2 \\ S_{WE}(x;\alpha,\lambda_0+\lambda_1+\lambda_2) & \text{if} \quad x_1 = x_2 = x. \end{cases} \end{aligned} \tag{2}$$

Therefore, the joint PDF of X_1 and X_2 can be written as

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} f_1(x_1,x_2) & \text{if} \quad 0 < x_1 < x_2 < \infty \\ f_2(x_1,x_2) & \text{if} \quad 0 < x_2 < x_1 < \infty \\ f_0(x) & \text{if} \quad 0 < x_1 = x_2 = x < \infty, \end{cases} \tag{3}$$

where

$$
\begin{aligned}
f_1(x_1,x_2) &= f_{WE}(x_1;\alpha,\lambda_1) f_{WE}(x_2;\lambda_0+\lambda_2) \\
f_2(x_1,x_2) &= f_{WE}(x_1;\alpha,\lambda_0+\lambda_1) f_{WE}(x_2;\lambda_2) \\
f_0(x) &= \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}\, f_{WE}(x;\alpha,\lambda_0+\lambda_1+\lambda_2). \qquad (4)
\end{aligned}
$$

Note that the function $f_{X_1,X_2}(\cdot,\cdot)$ may be considered to be a density function for MOBW distribution if it is understood that the first two terms are the densities with respect to the two dimensional Lebesgue measure and the third term is a density function with respect to one dimensional Lebesgue measure, see for example Bemis *et al.* [2]. It is clear that the joint cumulative distribution function (CDF) of X_1 and X_2 can be written as a mixture of an absolute continuous part and a singular part as follows;

$$
F_{X_1,X_2}(x_1,x_2) = \frac{\lambda_1+\lambda_2}{\lambda_0+\lambda_1+\lambda_2} F_1(x_1,x_2) + \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} F_0(x_1,x_2). \qquad (5)
$$

Here for $z = \max\{x_1,x_2\}$,

$$
F_0(x_1,x_2) = F_{WE}(z;\alpha,\lambda_0+\lambda_1+\lambda_2).
$$

$F_0(\cdot,\cdot)$ is the singular part and $F_1(\cdot,\cdot)$ is the absolute continuous part and it can be obtained by subtraction.

3 EM Algorithm for MOBW Distribution

In this section we address the problem of computing the MLEs of the unknown parameters of MOBW$(\alpha,\lambda_0,\lambda_1,\lambda_2)$, based on a random bivariate sample $\{(x_{11},x_{21}),\ldots,(x_{1n},x_{2n})\}$. We use the following notation;

$$
I_0 = \{i; x_{1i} = x_{2i} = x_i\}, \quad I_1 = \{i; x_{1i} < x_{2i}\}, \quad I_2 = \{i; x_{1i} > x_{2i}\}, \quad I = I_0 \cup I_1 \cup I_2
$$

$$|I_0| = n_0, \quad |I_1| = n_1, \quad |I_2| = n_2,$$

here $|I_j|$ for $j = 0, 1, 2$ denotes the number of elements in the set I_j.

The log-likelihood function can be written as

$$\begin{aligned}
l(\alpha, \lambda_0, \lambda_1, \lambda_2) &= \sum_{i \in I_1} \ln f_{WE}(x_{1i}; \alpha, \lambda_1) + \sum_{i \in I_1} \ln f_{WE}(x_{2i}; \alpha, \lambda_0 + \lambda_2) \\
&\quad + \sum_{i \in I_2} \ln f_{WE}(x_{1i}; \alpha, \lambda_0 + \lambda_1) + \sum_{i \in I_2} \ln f_{WE}(x_{2i}; \alpha, \lambda_2) \\
&\quad + n_0(\ln \lambda_0 - \ln(\lambda_0 + \lambda_1 + \lambda_2)) + \sum_{i \in I_0} \ln f_{WE}(x_i; \alpha, \lambda_0 + \lambda_1 + \lambda_2) \\
&= (n_0 + 2n_1 + 2n_2) \ln \alpha + n_1 \ln \lambda_1 + n_2 \ln \lambda_2 + n_0 \ln \lambda_0 + n_1 \ln(\lambda_0 + \lambda_2) \\
&\quad + n_2 \ln(\lambda_0 + \lambda_1) + (\alpha - 1) \left[\sum_{i \in I_1 \cup I_2} \ln x_{1i} + \sum_{i \in I_1 \cup I_2} \ln x_{2i} + \sum_{i \in I_0} \ln x_i \right] \\
&\quad - \lambda_1 \left[\sum_{i \in I_1 \cup I_2} x_{1i}^{\alpha} + \sum_{i \in I_0} x_i^{\alpha} \right] - \lambda_2 \left[\sum_{i \in I_1 \cup I_2} x_{2i}^{\alpha} + \sum_{i \in I_0} x_i^{\alpha} \right] \\
&\quad - \lambda_0 \left[\sum_{i \in I_2} x_{1i}^{\alpha} + \sum_{i \in I_1} x_{2i}^{\alpha} + \sum_{i \in I_0} x_i^{\alpha} \right]. \qquad (6)
\end{aligned}$$

It is known, see Bemis *et al.* [2], that even when $\alpha = 1$, the MLEs do not exist if one of the $n_i = 0$. If $n_0 > 0, n_1 > 0, n_2 > 0$, the MLEs of $\lambda_0, \lambda_1, \lambda_2$, exist, but no explicit expressions are available. They have to be obtained by solving three non-linear equations. Since this is a non-trivial problem, several suggestions, approximations, alternative estimators, have been suggested in the literature, see for example Bhattacharyya and Johnson [3], Proschan and Sullo [17], Arnold [1]. In case of MOBW distribution, the MLEs exist when $n_0 > 0, n_1 > 0, n_2 > 0$ and they can be obtained by maximizing (6) with respect to α, λ_0, λ_1 and λ_2. It becomes a non-linear optimization problem and clearly it is a non-trivial issue. No where in the literature the computational issues of the MLEs of MOBW distribution have been discussed.

We suggest to use EM algorithm to compute the MLEs of the unknown parameters of the MOBW model. It is easy to show that estimation of MOBW can be seen as a missing data problem. It is assumed that for the bivariate random vector (X_1, X_2), there is an associated random vector (Δ_1, Δ_2), where (Δ_1, Δ_2) is defined as follows.

$$\Delta_1 = \begin{cases} 0 & \text{if} \quad X_1 = U_0 \\ \\ 1 & \text{if} \quad X_1 = U_1 \end{cases} \qquad \text{and} \qquad \Delta_2 = \begin{cases} 0 & \text{if} \quad X_2 = U_0 \\ \\ 2 & \text{if} \quad X_2 = U_2. \end{cases} \tag{7}$$

Here U_i's are same as defined at the beginning of section 2. It is clear that in this case even if we know (X_1, X_2), but the corresponding (Δ_1, Δ_2) may not be known always. For example, if $X_1 = X_2$, then $\Delta_1 = \Delta_2 = 0$, is known. But, if $X_1 \neq X_2$, then (Δ_1, Δ_2) is not known. If $(x_1, x_2) \in I_1$, then the possible values of (Δ_1, Δ_2) are (1,0) or (1,2) and similarly, if $(x_1, x_2) \in I_2$, then the possible values of (Δ_1, Δ_2) are (0,2) or (1,2), with non-zero probabilities.

Note that if (X_1, X_2) and the associated (Δ_1, Δ_2) are known for all the observations, then the MLEs of the unknown parameters can be obtained very easily, by solving a one dimensional optimization problem. But unfortunately (Δ_1, Δ_2) are not known for all the observations. To implement the EM algorithm, first we obtain the 'E' step similarly as in Dinse [5]. In this case the 'pseudo log-likelihood' function ('E' step) is formed from the log-likelihood function (6) by replacing the log-likelihood contribution of (X_1, X_2) by its expected value, if the corresponding (Δ_1, Δ_2) is missing. The 'M' step is obtained by maximizing the 'pseudo log-likelihood' function with respect to the unknown parameters. It has been implemented as follows.

In the 'E' step we keep the log-likelihood contribution of all the observations belonging to I_0 intact, as in this case the corresponding (Δ_1, Δ_2)'s are known completely. If the observations belong to either I_1 or I_2, we treat them as missing observations. If $(x_1, x_2) \in I_1$, we form 'pseudo observation', similarly as in Dinse [5] or Kundu [9], by fractioning (x_1, x_2) to two partially complete 'pseudo observations' of the form $(x_1, x_2, u_1(\gamma))$ and $(x_1, x_2, u_2(\gamma))$. Here $\gamma = (\alpha, \lambda_0, \lambda_1, \lambda_2)$ and the fractional mass $u_1(\gamma)$ and $u_2(\gamma)$ assigned to the 'pseudo observation' are the conditional probabilities that (Δ_1, Δ_2) takes values (1,0) or (1,2) respectively, given that $X_1 < X_2$. Similarly, if $(x_1, x_2) \in I_2$, then the 'pseudo observation's are formed as $(x_1, x_2, v_1(\gamma))$ and $(x_1, x_2, v_2(\gamma))$, where $v_1(\gamma)$ and $v_2(\gamma)$ are the conditional probabilities that (Δ_1, Δ_2) takes values (0,2) or (1,2) respectively, given that $X_1 > X_2$. Since

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

therefore

$$u_1(\gamma) = \frac{\lambda_0}{\lambda_0+\lambda_2} \quad \text{and} \quad u_2(\gamma) = \frac{\lambda_2}{\lambda_0+\lambda_2}.$$

Similarly,

$$v_1(\gamma) = \frac{\lambda_0}{\lambda_0+\lambda_1} \quad \text{and} \quad v_2(\gamma) = \frac{\lambda_1}{\lambda_0+\lambda_1}.$$

From now on for brevity, we write u_1, u_2, v_1, v_2, instead of $u_1(\gamma), u_2(\gamma), v_1(\gamma), v_2(\gamma)$. The log-likelihood function of the 'pseudo data' can be written as

$$\begin{aligned}
l_{pseudo}(\alpha,\lambda_0,\lambda_1,\lambda_2) \;=\;& n_0\ln\alpha + n_0\ln\lambda_0 - \lambda_0\sum_{i\in I_0} x_i^{\alpha} + (\alpha-1)\sum_{i\in I_0}\ln x_i - \lambda_1\sum_{i\in I_1} x_{1i}^{\alpha} - \lambda_2\sum_{i\in I_2} x_{2i}^{\alpha} \\
&+u_1\Bigg(n_1\ln\alpha + n_1\ln\lambda_1 - \lambda_1\sum_{i\in I_1} x_{1i}^{\alpha} + (\alpha-1)\sum_{i\in I_1}\ln x_{1i} + n_1\ln\alpha \\
&+n_1\ln\lambda_0 - \lambda_0\sum_{i\in I_1} x_{2i}^{\alpha} + (\alpha-1)\sum_{i\in I_1}\ln x_{2i} - \lambda_2\sum_{i\in I_1} x_{2i}^{\alpha}\Bigg) \\
&+u_2\Bigg(n_1\ln\alpha + n_1\ln\lambda_1 - \lambda_1\sum_{i\in I_1} x_{1i}^{\alpha} + (\alpha-1)\sum_{i\in I_1}\ln x_{1i} + n_1\ln\alpha \\
&+n_1\ln\lambda_2 - \lambda_2\sum_{i\in I_1} x_{2i}^{\alpha} + (\alpha-1)\sum_{i\in I_1}\ln x_{2i} - \lambda_0\sum_{i\in I_1} x_{2i}^{\alpha}\Bigg) \\
&+v_1\Bigg(n_2\ln\alpha + n_2\ln\lambda_2 - \lambda_2\sum_{i\in I_2} x_{2i}^{\alpha} + (\alpha-1)\sum_{i\in I_2}\ln x_{2i} + n_2\ln\alpha \\
&+n_2\ln\lambda_0 - \lambda_0\sum_{i\in I_2} x_{1i}^{\alpha} + (\alpha-1)\sum_{i\in I_2}\ln x_{1i} - \lambda_1\sum_{i\in I_2} x_{1i}^{\alpha}\Bigg) \\
&+v_2\Bigg(n_2\ln\alpha + n_2\ln\lambda_2 - \lambda_2\sum_{i\in I_2} x_{2i}^{\alpha} + (\alpha-1)\sum_{i\in I_2}\ln x_{2i} + n_2\ln\alpha \\
&+n_2\ln\lambda_1 - \lambda_1\sum_{i\in I_2} x_{1i}^{\alpha} + (\alpha-1)\sum_{i\in I_2}\ln x_{1i} - \lambda_0\sum_{i\in I_2} x_{1i}^{\alpha}\Bigg).
\end{aligned}$$

It can be simplified as

$$
\begin{aligned}
l_{pseudo}(\alpha,\lambda_0,\lambda_1,\lambda_2) &= (n_0+2n_1+2n_2)\ln\alpha+(\alpha-1)\left(\sum_{i\in I_0}\ln x_i+\sum_{i\in I_1\cup I_2}[\ln x_{1i}+\ln x_{2i}]\right)\\
&\quad -\lambda_0\left(\sum_{i\in I_0}x_i^{\alpha}+\sum_{i\in I_2}x_{1i}^{\alpha}+\sum_{i\in I_1}x_{2i}^{\alpha}\right)+(n_0+u_1n_1+v_1n_2)\ln\lambda_0\\
&\quad -\lambda_1\left(\sum_{i\in I_0}x_i^{\alpha}+\sum_{i\in I_1\cup I_2}x_{1i}^{\alpha}\right)+(n_1+v_2n_2)\ln\lambda_1\\
&\quad -\lambda_2\left(\sum_{i\in I_0}x_i^{\alpha}+\sum_{i\in I_1\cup I_2}x_{2i}^{\alpha}\right)+(n_2+u_2n_1)\ln\lambda_2. \qquad (8)
\end{aligned}
$$

Therefore, 'M' step involves maximizing (8) with respect to (w.r.t.) α, λ_0, λ_1 and λ_2. Note that for fixed α, the maximization of (8) w.r.t λ_0, λ_1 and λ_2 can be obtained at

$$\widehat{\lambda}_0(\alpha)=\frac{n_0+u_1n_1+v_1n_2}{\sum_{i\in I_0}x_i^{\alpha}+\sum_{i\in I_2}x_{1i}^{\alpha}+\sum_{i\in I_1}x_{2i}^{\alpha}}, \qquad (9)$$

$$\widehat{\lambda}_1(\alpha)=\frac{n_1+v_2n_2}{\sum_{i\in I_0}x_i^{\alpha}+\sum_{i\in I_1\cup I_2}x_{1i}^{\alpha}},\quad \widehat{\lambda}_2(\alpha)=\frac{n_2+u_2n_1}{\sum_{i\in I_0}x_i^{\alpha}+\sum_{i\in I_1\cup I_2}x_{2i}^{\alpha}}. \qquad (10)$$

Substituting $\widehat{\lambda}_i(\alpha)$'s for λ_i's in (8), it can be easily observed by taking the second derivative that the profile pseudo log-likelihood function $l_{pseudo}(\alpha,\widehat{\lambda}_0(\alpha),\widehat{\lambda}_1(\alpha),\widehat{\lambda}_2(\alpha))$ is a unimodal function of α. Therefore, it has a unique maximum. The maximization of the profile pseudo log-likelihood function w.r.t. α, can be performed by the standard Newton-Raphson algorithm, bisection method or similar to Kundu and Gupta [10] by solving a fixed point type equation

$$g(\alpha)=\alpha, \qquad (11)$$

where if

$$
\begin{aligned}
h(\alpha) &= \Bigg[\widehat{\lambda}_0(\alpha)\left(\sum_{i\in I_0}x_i^{\alpha}\ln x_i+\sum_{i\in I_2}x_{1i}^{\alpha}\ln x_{1i}+\sum_{i\in I_1}x_{2i}^{\alpha}\ln x_{2i}\right)\\
&\quad +\widehat{\lambda}_1(\alpha)\left(\sum_{i\in I_0}x_i^{\alpha}\ln x_i+\sum_{i\in I_1\cup I_2}x_{1i}^{\alpha}\ln x_{1i}\right)+\widehat{\lambda}_2(\alpha)\left(\sum_{i\in I_0}x_i^{\alpha}\ln x_i\sum_{i\in I_1\cup I_2}x_{2i}^{\alpha}\ln x_{2i}\right)\\
&\quad -\left(\sum_{i\in I_0}\ln x_i+\sum_{i\in I_1\cup I_2}(\ln x_{1i}+\ln x_{2i})\right)\Bigg]
\end{aligned}
$$

then

$$g(\alpha) = \frac{(n_0 + 2n_1 + 2n_2)}{h(\alpha)}.$$

Note that solving (11) is very simple. We can start the initial guess as $\alpha^{(0)}$, then $\alpha^{(1)}$ can be obtained as $g(\alpha^{(0)})$ and the process continues until it converges, see Kundu and Gupta [10]. Now we describe how to obtain the $(i+1)$-th step from the i-th step of the EM algorithm. Suppose at the i-th step the estimates of $\alpha, \lambda_0, \lambda_1, \lambda_2$ are $\alpha^{(i)}, \lambda_0^{(i)}, \lambda_1^{(i)}, \lambda_2^{(i)}$ respectively.

- Step 1: Compute u_1, u_2, v_1, v_2 using $\alpha^{(i)}, \lambda_0^{(i)}, \lambda_1^{(i)}, \lambda_2^{(i)}$.
- Step 2: Find $\alpha^{(i+1)}$ by solving (11) similarly as in Kundu and Gupta [10].
- Step 3: Once $\alpha^{(i+1)}$ is obtained compute $\lambda_0^{(i+1)}, \lambda_1^{(i+1)}, \lambda_2^{(i+1)}$ from (9) and (10).

The process should be continued until the convergence criterion is met. It should be mentioned that this version of EM algorithm is popularly known as ECM (expectation-conditional maximization) algorithm.

Comment 1: As one referee has correctly mentioned that if α is very close to zero, then it may happen at a particular stage that the updated α obtained by using (11) may be negative, which will stop the process. In our extensive simulations, we have not faced that problem. This problem may occur even for Newton-Raphson algorithm also. In this case we suggest to use the bisection method, although in general it takes more number of iterations to converge than the proposed algorithm or the Newton-Raphson method.

Comment 2: Since MOBE model can be obtained from the MOBW model by putting $\alpha = 1$, in MOBW model, therefore the proposed EM algorithm can be used for the MOBE model also. In this case we do not need to solve the fixed point equation (11) and in each EM step the estimates of $\widehat{\lambda}_0$, $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ can be obtained as $\widehat{\lambda}_0(1)$, $\widehat{\lambda}_1(1)$ and $\widehat{\lambda}_2(1)$ from (9) and (10).

COMMENTS 3: Although the computation of the MLEs of the MOBW parameters is a non-trivial problem, but the standard asymptotic properties of the MLEs hold here. The asymptotic dispersion matrix can be easily obtained from the expected Fisher information matrix.

4 Data Analysis

For illustrative purpose, in this section we have analyzed one data set from Meintanis [15] and it is presented in Table 1. It represents the football (soccer) data where at least one goal scored by the home team and at least one goal scored directly from a penalty kick, foul kick or any other direct kick (all of them together will be called as *kick* goal) by any team have been considered. Here X_1 represents the time in minutes of the first *kick* goal scored by any team and X_2 represents the first goal of any type scored by the home team. In this case all possibilities are open, for example $X_1 < X_2$, or $X_1 > X_2$ or $X_1 = X_2 = X$ (say). Meintanis [15] analyzed this data by using MOBE distribution. We would like to analyze the data using MOBW model also. All the data points have been divided by 100 so that the shape and scale parameters are of the same order. This is not going to make any difference in any statistical inference.

Before going to analyze the data using MOBW model, we fit Weibull distributions to X_1, X_2 and $\min\{X_1, X_2\}$ separately. Apart from model checking, it will help us to guess the initial values also. The MLEs of the shape and scale parameters of the respective Weibull distribution for X_1, X_2 and $\min\{X_1, X_2\}$ are (2.121, 5.207), (1.421, 4.263) and (1.478, 5.340) respectively. The corresponding Kolmogorov-Smirnov distances between the fitted distribution and the empirical distribution function and the associated p values (in brackets) for X_1, X_2 and $\min\{X_1, X_2\}$ are 0.083 (0.961), 0.105(0.805) and 0.069 (0.995)respectively. Based on the p values Weibull distribution can not be rejected for the marginals and for the

2005-2006	X_1	X_2	2004-2005	X1	X2
Lyon-Real Madrid	26	20	Internazionale-Bremen	34	34
Milan-Fenerbahce	63	18	Real Madrid-Roma	53	39
Chelsea-Anderlecht	19	19	Man. United-Fenerbahce	54	7
Club Brugge-Juventus	66	85	Bayern-Ajax	51	28
Fenerbahce-PSV	40	40	Moscow-PSG	76	64
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-Benfica	39	39	Dynamo Kyiv-Real Madrid	44	13
Real Madrid-Rosenborg	82	48	Man. United-Sparta	25	14
Villarreal-Benfica	72	72	Bayern-M. TelAviv	55	11
Juventus-Bayern	66	62	Bremen-Internazionale	49	49
Club Brugge-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

minimum also.

Now we will fit the MOBW model. To start the EM algorithm we need some initial guesses of the unknown parameters. For α, we suggest to take the average values of 2.121, 1.421 and 1.478, *i.e.* 1.67. Assuming the initial guess of α as 1.67, solving three linear equations in three unknowns for λ's, we get the initial guess values of λ_0, λ_1 and λ_2 as 2.7, 1.2 and 2.4 respectively. Using these initial guesses we have started the EM algorithm and the iteration stops when the ratio $|(l(k) - l(k-1))/l(k-1)| < 10^{-8}$, where $l(k)$ denotes the value of the log-likelihood function at the $k-th$ iterate. In each iteration we need to solve $g(\alpha) = \alpha$, and we use the stopping criterion as $|\alpha^{(j)} - \alpha^{(j+1)}| < 10^{-6}$. The iteration stops after 18 steps and we obtain the following estimates of $\alpha, \lambda_0, \lambda_1, \lambda_2$ as 1.6954, 2.6927, 1.2192, 2.8052 respectively. The corresponding log-likelihood value is -13.118047. It is interesting to

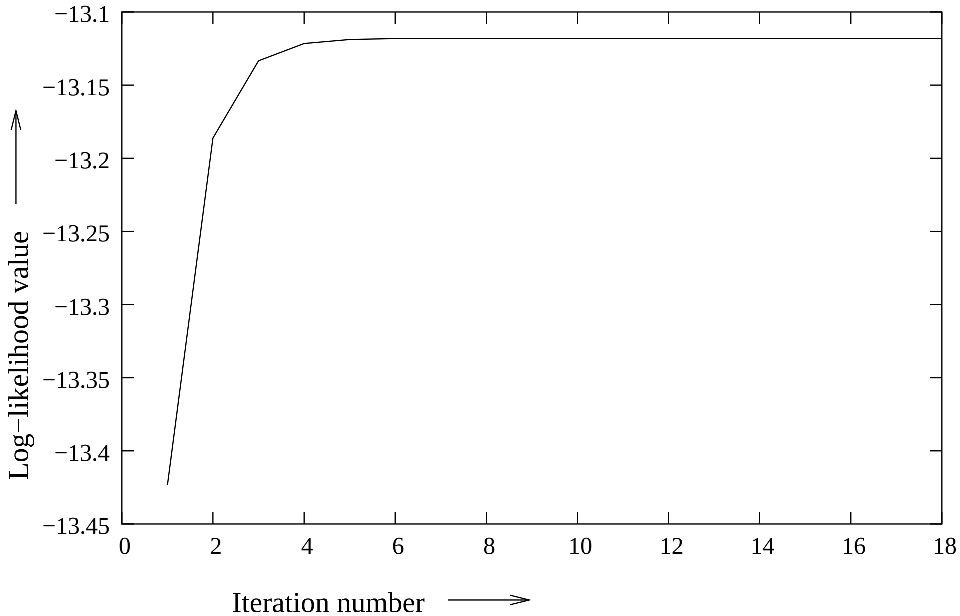

Figure 1: Log-likelihood value at different iteration.

see that at each step the log-likelihood function is gradually increasing and it is presented in Figure 1. From the Figure 1 it is clear that the log-likelihood value almost stabilizes after 4-th iteration. We have tried some other initial guesses also, for example with the initial guess of 1.0, 1.0, 1.0, 1.0 for $\alpha, \lambda_0, \lambda_1, \lambda_2$ respectively, the EM algorithm converges to the same point after 20 steps when we use the same stopping criterion.

We have also computed 95% confidence intervals of $\alpha, \lambda_0, \lambda_1, \lambda_2$ using the observed Fisher information matrix obtained from the EM algorithm (presented in the appendix) as suggested by Louis [11] and they are as follows; (1.3284, 2.0623), (1.5001, 3.8852), 0.2708, 1.1415), (1.2023, 2.4488) respectively. Now the natural question is whether MOBW model fits the data well or not. The Kolmogorov-Smirnov distances and the corresponding p values (reported within brackets) between X_1, X_2 and $\min\{X_1, X_2\}$ with WE(1.6954,3.9119), WE(1.6954,5.4979) and WE(1.6954,6.7171) are 0.1149 (0.713), 0.1307 (0.552) and 0.1043 (0.815) respectively. It indicates that the Weibull distribution can be used for analyzing X_1, X_2 and $\min\{X_1, X_2\}$. As one referee correctly pointed out that although it does not guarantee that (X_1, X_2) will have MOBW distribution, but at least it gives an indication that the MOBW model may be used to analyze this bivariate data set. Currently there are no tests available to test this hypothesis. One may work on extending the idea presented on Meintanis [15] or create some bootstrap based goodness of fit tests, it has not attempted here.

We have fitted the MOBE model also to the data set as suggested by Meintanis [15]. Using the EM algorithm and the same initial guesses for $\lambda_0, \lambda_1, \lambda_2$, we obtain the MLEs as 1.7676, 0.7226 and 1.6352 respectively and the corresponding log-likelihood value is -22.756946. Note that Meintanis [15] did not use the MLEs, he used the corresponding estimates as 1.73, 0.73 and 1.66 respectively, which are quite close to the MLEs. We also obtained the 95% confidence intervals of $\lambda_0, \lambda_1, \lambda_2$ and they are (1.0378, 2.4975), (0.1844, 1.2608) and (0.8877, 2.3826) respectively. The Kolmogorov-Smirnov distances and the corresponding p values (reported within brackets) between X_1, X_2 and $\min\{X_1, X_2\}$ with Exp(2.4902), Exp(3.4028) and Exp(4.1254) are 0.2736 (0.008), 0.1683 (0.245) and 0.2259 (0.046) respectively.

Now from the confidence interval of α, from the log-likelihood values and also from the Kolmogorov-Smirnov distances, it is clear that although Meintanis [15] suggested to use MOBE, MOBW is preferable in this case.

5 Simulation Results

In this section we present some simulation results to verify how the proposed EM algorithm performs for different sample sizes and for different parameter values for MOBW model. We have kept $\lambda_0 = 1.0$, $\lambda_1 = 1.0$ and $\lambda_2 = 1.0$ fixed and used different α and n, namely $\alpha =$ 0.25, 0.50, 1.0, 2.0, 5.0 and n = 25, 50, 100, 500. In each case we have started the EM algorithm with the initial guesses of $\alpha, \lambda_0, \lambda_1, \lambda_2$ as 0.5, 0.5, 0.5 and 0.5 respectively. We have tried other initial guesses also, but the average estimates and the corresponding MSEs are same. We have used the same stopping criterion as in the previous section. The average estimates (AE), mean squared error (MSE), average number of iterations (AI) required and also the coverage percentages (CP) based on 95% confidence intervals obtained from the EM alogorithm are obtained based on 1000 replications. It may be observed that the estimates of λ_0, λ_1 and λ_2 do not depend on α. Therefore, we report the results of α in Table 2 and the results of λ_0, λ_1 and λ_2 for all the cases in Table 3.

For comparison purposes we have also performed the experiment for the MOBE model using our method and the method proposed by Karlis [8] for the same λ_0, λ_1 and λ_2, the same stopping criterion and the same initial guesses. Since both of them compute the MLEs, their average estimates and the mean squared errors are same. The only difference is the number of iterations required in each case. We report the average estimates, the corresponding mean squared errors and the average number of iterations required for the two cases in Table 4.

Some of the points are quite clear from the experimental results. In all the cases the estimates are slightly positively biased, mainly for small sizes, but the average biases and the average MSEs decrease as the sample size increases. If the sample size is not very small the asymptotic normality results can be used for constructing confidence intervals and hence for testing purposes also. Comparing the results of Tables 3 and 4, it is clear that if α is known the estimates of λ_0, λ_1 and λ_2 are better in terms of biases and MSEs. The number of (EM) iterations required for MOBW and MOBE are more or less same. From the experimental results it is clear that the proposed EM algorithm is working quite well for both MOBW and MOBE models.

Now comparing the results between the two methods in Table 4 for MOBE model, it is clear that when we use the same stopping criterion, the average number of iterations required by the method of Karlis [8] is significantly more than the proposed method, although both

Table 2: The average estimates (AE), the mean squared errors (MSE), average number of iterations (AI) and the coverage percentages (CI) of α for MOBW model.

n	$\alpha \longrightarrow$	0.25	0.50	1.0	2.0	5.0
25	AE	0.2637	0.5274	1.0547	2.1095	5.2737
	MSE	(0.0014)	(0.0054)	(0.0216)	(0.0865)	(0.5407)
	CP	0.93	0.93	0.93	0.94	0.95
	AI	16.10	17.22	16.14	16.81	18.06
50	AE	0.2565	0.5131	1.0261	2.0522	5.1305
	MSE	(0.0006)	(0.0023)	(0.0091)	(0.0365)	(0.2283)
	CP	0.93	0.93	0.94	0.94	0.94
	AI	16.37	15.98	14.90	15.30	16.42
100	AE	0.2527	0.5055	1.0110	2.0221	5.0552
	MSE	(0.0003)	(0.0011)	(0.0042)	(0.0168)	(0.1054)
	CP	0.95	0.95	0.94	0.95	0.95
	AI	13.96	15.02	13.97	14.12	15.48
500	AE	0.2505	0.5010	1.0020	2.0039	5.0098
	MSE	(0.00005)	(0.0002)	(0.0008)	(0.0031)	(0.0191)
	CP	0.95	0.95	0.95	0.95	0.95
	AI	11.15	11.84	10.99	10.95	12.93

Table 3: The average estimates (AE), the mean squared errors (MSE), and the coverage percentages (CI) of λ_0, λ_1 and λ_2 for MOBW model.

n		λ_0	λ_1	λ_2
25	AE	1.0385	1.0837	1.0732
	MSE	(0.1130)	(0.1363)	(0.1379)
	CP	0.95	0.93	0.95
50	AE	1.0194	1.0409	1.0469
	MSE	(0.0505)	(0.0580)	(0.0616)
	CP	0.95	0.93	0.95
100	AE	1.0116	1.0225	1.0190
	MSE	(0.0234)	(0.0251)	(0.0287)
	CP	0.95	0.95	0.95
500	AE	1.0036	1.0045	1.0057
	MSE	(0.0044)	(0.0048)	(0.0053)
	CP	0.95	0.95	0.96

Table 4: The average estimates (AE), the mean squared errors (MSE) and average number of iterations (AI) required for two methods, namely the proposed method (PM) and the method proposed by Karlis [8] (KM) are presented. Model MOBE(1.0,1.0,1.0) is used in this case.

n		λ_0	λ_1	λ_2	AI (PM)	AI (KM)
25	AE MSE	1.0335 (0.1115)	1.0393 (0.1251)	1.0570 (0.1055)	20.79	52.67
50	AE MSE	1.0291 (0.0524)	1.0254 (0.0567)	1.0304 (0.0475)	19.36	43.00
100	AE MSE	1.0129 (0.0248)	1.0169 (0.0245)	1.0098 (0.0224)	18.36	38.53
500	AE MSE	1.0046 (0.0052)	1.0016 (0.0051)	1.0051 (0.0045)	17.02	34.08

the methods provide the same solutions. An interesting point is that as the sample size increases the difference becomes smaller and smaller, but it is observed (not reported here) that even with sample size $n = 5000$, the difference in the average number of iterations does not vanish.

6 Multivariate Marshall-Olkin Weibull Distribution

In this section we mainly indicate how the EM algorithm described in section 3 can be extended for the multivariate Marshall-Olkin Weibull model. For notational simplicity we restrict ourselves to the trivariate model only but the idea can be easily used for any dimensions. Marshall-Olkin trivariate Weibull (MOTW) model can be described as follows. Suppose $U_i \sim$ WE(α, λ_i); for $i = 0, 1, 2, 3$ and they are independently distributed. If $X_i = \min\{U_i, U_0\}$ for

$i = 1, 2, 3$, then $(X_1, X_2, X_3) \sim$ MOTW model with parameters $(\alpha, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$ and it will be denoted as MOTW$(\alpha, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$.

We are interested in estimating the unknown parameters $\alpha, \lambda_0, \lambda_1, \lambda_2, \lambda_3$ from a sample $\{(x_{1i}, x_{2i}, x_{3i}), i = 1, \ldots, n\}$. Note that here for each random vector (X_1, X_2, X_3), there is an associated random vector $(\Delta_1, \Delta_2, \Delta_3)$, such that Δ_i takes the value 0 or i, as follows;

$$\Delta_1 = \begin{cases} 0 & \text{if} \quad X_1 = U_0 \\ \\ 1 & \text{if} \quad X_1 = U_1 \end{cases} \qquad \Delta_2 = \begin{cases} 0 & \text{if} \quad X_2 = U_0 \\ \\ 2 & \text{if} \quad X_2 = U_2 \end{cases} \qquad \Delta_3 = \begin{cases} 0 & \text{if} \quad X_3 = U_0 \\ \\ 3 & \text{if} \quad X_3 = U_3. \end{cases} \tag{12}$$

We use the following notation: $I_0 = \{i; x_{1i} = x_{2i} = x_{3i} = x_i\}$, $I_{10} = \{i; x_{1i} < x_{2i} = x_{3i} = x_{10i}\}$, $I_{20} = \{i; x_{2i} < x_{1i} = x_{3i} = x_{20i}\}$, $I_{30} = \{i; x_{3i} < x_{1i} = x_{2i} = x_{30i}\}$, $I_{123} = \{i; x_{1i} > x_{2i} > x_{3i}\}$, $I_{132} = \{i; x_{1i} > x_{3i} > x_{2i}\}$, $I_{213} = \{i; x_{2i} > x_{1i} > x_{3i}\}$, $I_{231} = \{i; x_{2i} > x_{3i} > x_{1i}\}$, $I_{312} = \{i; x_{3i} > x_{1i} > x_{2i}\}$, $I_{321} = \{i; x_{3i} > x_{2i} > x_{1i}\}$. We further use $I_1 = I_{123} \cup I_{132}$, $I_2 = I_{231} \cup I_{213}$, $I_3 = I_{312} \cup I_{321}$, and the number of elements in each set will be denoted as follows; $|I_0| = n_0$, $|I_{10}| = n_{10}$ etc.

Note that for $I_0, I_{10}, I_{20}, I_{30}$, $(\Delta_1, \Delta_2, \Delta_3)$ are known and they are (0,0,0), (1,0,0), (0,2,0), (0,0,3) respectively. For other cases one of the Δ_i is not known.

- If $(X_1, X_2, X_3) \in I_1$, $\Delta_2 = 2$, $\Delta_3 = 3$, but Δ_1 can be 0 or 1 and

$$P(\Delta_1 = 0|(X_1, X_2, X_3) \in I_1) = \frac{\lambda_0}{\lambda_0 + \lambda_1}, \qquad P(\Delta_1 = 1|(X_1, X_2, X_3) \in I_1) = \frac{\lambda_1}{\lambda_0 + \lambda_1}.$$

- If $(X_1, X_2, X_3) \in I_2$, $\Delta_1 = 1$, $\Delta_3 = 3$, but Δ_2 can be 0 or 2 and

$$P(\Delta_2 = 0|(X_1, X_2, X_3) \in I_2) = \frac{\lambda_0}{\lambda_0 + \lambda_2}, \qquad P(\Delta_2 = 2|(X_1, X_2, X_3) \in I_2) = \frac{\lambda_2}{\lambda_0 + \lambda_2}.$$

- If $(X_1, X_2, X_3) \in I_3$, $\Delta_1 = 1$, $\Delta_2 = 2$, but Δ_3 can be 0 or 3 and

$$P(\Delta_3 = 0|(X_1, X_2, X_3) \in I_3) = \frac{\lambda_0}{\lambda_0 + \lambda_3}, \qquad P(\Delta_3 = 3|(X_1, X_2, X_3) \in I_3) = \frac{\lambda_3}{\lambda_0 + \lambda_3}.$$

Then proceeding exactly as before, for a given $u_1, u_2, v_1, v_2, w_1, w_2$, where

$$u_1 = \frac{\lambda_0}{\lambda_0 + \lambda_1}, u_2 = \frac{\lambda_1}{\lambda_0 + \lambda_1}, v_1 = \frac{\lambda_0}{\lambda_0 + \lambda_2}, v_2 = \frac{\lambda_2}{\lambda_0 + \lambda_2}, w_1 = \frac{\lambda_0}{\lambda_0 + \lambda_3}, w_2 = \frac{\lambda_3}{\lambda_0 + \lambda_3},$$

we have

$$\begin{aligned}
\widehat{\lambda}_0(\alpha) &= \frac{n_0 + u_1 n_1 + v_1 n_2 + w_1 n_3}{\sum_{i\in I_0} x_i^\alpha + \sum_{i\in I_{10}} x_{10i}^\alpha + \sum_{i\in I_{20}} x_{20i}^\alpha + \sum_{i\in I_{30}} x_{30i}^\alpha + \sum_{i\in I_1} x_{1i}^\alpha + \sum_{i\in I_2} x_{2i}^\alpha + \sum_{i\in I_3} x_{3i}^\alpha}, \\
\widehat{\lambda}_1(\alpha) &= \frac{n_{10} + u_2 n_1 + n_2 + n_3}{\sum_{i\in I_0} x_i^\alpha + \sum_{i\in I_{10}\cup I_1\cup I_2\cup I_3} x_{1i}^\alpha}, \\
\widehat{\lambda}_2(\alpha) &= \frac{n_{20} + n_1 + v_2 n_2 + n_3}{\sum_{i\in I_0} x_i^\alpha + \sum_{i\in I_{20}\cup I_1\cup I_2\cup I_3} x_{2i}^\alpha}, \\
\widehat{\lambda}_3(\alpha) &= \frac{n_{30} + n_1 + n_2 + w_2 n_3}{\sum_{i\in I_0} x_i^\alpha + \sum_{i\in I_{30}\cup I_1\cup I_2\cup I_3} x_{3i}^\alpha}. \qquad (13)
\end{aligned}$$

We define $g(\cdot)$ and $h(\cdot)$ as in section 3.

$$h(\alpha) = \left[\widehat{\lambda}_0(\alpha)g_0'(\alpha) + \widehat{\lambda}_1(\alpha)g_1'(\alpha) + \widehat{\lambda}_2(\alpha)g_2'(\alpha) + \widehat{\lambda}_3(\alpha)g_3'(\alpha)\right],$$

and

$$g(\alpha) = \frac{n_0 + 2(n_{10} + n_{20} + n_{30}) + 3(n_1 + n_2 + n_3)}{h(\alpha)},$$

here $g_i'(\alpha)$ is the derivative of $g_i(\alpha)$ with respect to α and $g_i(\alpha)$ is the denominator of $\widehat{\lambda}_i(\alpha)$ for $i = 0, 1, 2, 3$.

Now we can define the EM algorithm similar to section 3 how to obtain the $(i+1)$-th step from the i-th step. Suppose at the i-th step the estimates of $\alpha, \lambda_0, \lambda_1, \lambda_2, \lambda_3$ are $\alpha^{(i)}, \lambda_0^{(i)}, \lambda_1^{(i)}, \lambda_2^{(i)}, \lambda_3^{(i)}$ respectively.

- Step 1: Compute $u_1, u_2, v_1, v_2, w_1, w_2$ using $\alpha^{(i)}, \lambda_0^{(i)}, \lambda_1^{(i)}, \lambda_2^{(i)}, \lambda_3^{(i)}$.
- Step 2: Find $\alpha^{(i+1)}$ by solving $g(\alpha) = \alpha$ similarly as in Kundu and Gupta [10].
- Step 3: Once $\alpha^{(i+1)}$ is obtained compute $\lambda_0^{(i+1)}, \lambda_1^{(i+1)}, \lambda_2^{(i+1)}, \lambda_3^{(i+1)}$ from (13).

7 Conclusions

In this paper we have considered the MOBW distribution and discuss the EM algorithm for computing the maximum likelihood estimators of the four unknown parameters. It is observed that the implementation of the EM algorithm is not very difficult and it involves only a one dimensional optimization problem. We have provided a simple iteration technique to perform the one dimensional optimization procedure. The simulation results indicate that the performance of the EM algorithm is quite satisfactory. We have also constructed the asymptotic confidence intervals using the idea of Louis [11] and it is observed that even for moderate sample sizes the asymptotic results can be used for constructing confidence intervals and hence for testing purposes also. We have re-analyzed one data set, which was originally analyzed by Meintanis [15] using MOBE distribution, and we observed that for that data set MOBW model provides a better fit than the bivariate exponential model.

Although we have provided the EM algorithm for MOBW distribution, but it can be extended for the Marshall-Olkin multivariate Weibull model also. We have indicated briefly how it can be done for the trivariate model. It should be mentioned that the above procedure can be extended for other models also. For example in case of mixtures of MOBW or MOMW, the EM algorithm can be used. Moreover, in case of Block and Basu [4] type bivariate Weibull distribution the proposed EM algorithm can be easily extended. The work is in progress and it will reported later.

Acknowledgements

The author would like to thank the associate editor and two referees for their valuable comments. The authors are particularly thankful to one referee for Table 4.

Appendix

In the Appendix we provide the observed Fisher information matrix for the MOBW model. It has been used to compute the asymptotic confidence intervals of the unknown parameters.

Observed Fisher Information Matrix

To compute the observed information matrix, we use the same notation as of Louis [11]. If the matrix $\mathbf{S} = ((S_{ij}))$ denotes the Hessian matrix and the vector $\mathbf{U} = (U_i)$ denotes the gradient vector of the pseudo log-likelihood function, then the observed Fisher information matrix can be obtained as $\mathbf{S} - \mathbf{U}\mathbf{U}^T$. Below we provide the elements of the matrix $\mathbf{S}$ and the vector $\mathbf{U}$.

$$
\begin{aligned}
S_{11} &= \frac{n_0 + 2n_1 + 2n_2}{\widehat{\alpha}^2} + \widehat{\lambda}_0 \left[\sum_{i \in I_2} x_{1i}^{\widehat{\alpha}} (\ln x_{1i})^2 + \sum_{i \in I_1} x_{2i}^{\widehat{\alpha}} (\ln x_{2i})^2 + \sum_{i \in I_0} x_i^{\widehat{\alpha}} (\ln x_i)^2 \right] \\
&\quad + \widehat{\lambda}_1 \left[\sum_{i \in I_1 \cup I_2} x_{1i}^{\widehat{\alpha}} (\ln x_{1i})^2 + \sum_{i \in I_0} x_i^{\widehat{\alpha}} (\ln x_i)^2 \right] + \widehat{\lambda}_2 \left[\sum_{i \in I_1 \cup I_2} x_{2i}^{\widehat{\alpha}} (\ln x_{2i})^2 + \sum_{i \in I_0} x_i^{\widehat{\alpha}} (\ln x_i)^2 \right] \\
S_{12} &= S_{21} = \sum_{i \in I_2} x_{1i}^{\widehat{\alpha}} \ln x_{1i} + \sum_{i \in I_1} x_{2i}^{\widehat{\alpha}} \ln x_{2i} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \ln x_i \\
S_{13} &= S_{31} = \sum_{i \in I_1 \cup I_2} x_{1i}^{\widehat{\alpha}} \ln x_{1i} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \ln x_i \\
S_{14} &= S_{41} = \sum_{i \in I_1 \cup I_2} x_{2i}^{\widehat{\alpha}} \ln x_{2i} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \ln x_i \\
S_{22} &= \frac{n_0}{\widehat{\lambda}_0^2} + \frac{n_1}{(\widehat{\lambda}_0 + \widehat{\lambda}_2)^2} + \frac{n_2}{(\widehat{\lambda}_0 + \widehat{\lambda}_1)^2}, \quad S_{23} = S_{32} = \frac{n_2}{(\widehat{\lambda}_0 + \widehat{\lambda}_1)^2} \\
S_{33} &= \frac{n_1}{\widehat{\lambda}_1^2} + \frac{n_2}{(\widehat{\lambda}_0 + \widehat{\lambda}_1)^2}, \quad S_{34} = S_{43} = 0, \quad S_{44} = \frac{n_2}{\widehat{\lambda}_2^2} + \frac{n_1}{(\widehat{\lambda}_0 + \widehat{\lambda}_2)^2}
\end{aligned}
$$

$$
\begin{aligned}
U_1 &= \frac{n_0 + 2n_1 + 2n_2}{\widehat{\alpha}} + \left[\sum_{i \in I_1 \cup I_2} \ln x_{1i} + \sum_{i \in I_1 \cup I_2} \ln x_{2i} + \sum_{i \in I_0} \ln x_i \right] \\
&\quad - \widehat{\lambda}_1 \left[\sum_{i \in I_1 \cup I_2} x_{1i}^{\widehat{\alpha}} \ln x_{1i} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \ln x_i \right] - \widehat{\lambda}_2 \left[\sum_{i \in I_1 \cup I_2} x_{2i}^{\widehat{\alpha}} \ln x_{2i} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \ln x_i \right] \\
&\quad - \widehat{\lambda}_0 \left[\sum_{i \in I_2} x_{1i}^{\widehat{\alpha}} \ln x_{1i} + \sum_{i \in I_1} x_{2i}^{\widehat{\alpha}} \ln x_{2i} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \ln x_i \right]
\end{aligned}
$$

$$U_2 = \frac{n_0}{\widehat{\lambda}_0} + \frac{n_1}{\widehat{\lambda}_0 + \widehat{\lambda}_2} + \frac{n_2}{\widehat{\lambda}_0 + \widehat{\lambda}_1} - \left[\sum_{i \in I_2} x_{1i}^{\widehat{\alpha}} + \sum_{i \in I_1} x_{2i}^{\widehat{\alpha}} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \right]$$

$$U_3 = \frac{n_1}{\widehat{\lambda}_1} + \frac{n_2}{\widehat{\lambda}_0 + \widehat{\lambda}_1} - \left[\sum_{i \in I_1 \cup I_2} x_{1i}^{\widehat{\alpha}} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \right]$$

$$U_4 = \frac{n_2}{\widehat{\lambda}_2} + \frac{n_1}{\widehat{\lambda}_0 + \widehat{\lambda}_2} - \left[\sum_{i \in I_1 \cup I_2} x_{2i}^{\widehat{\alpha}} + \sum_{i \in I_0} x_i^{\widehat{\alpha}} \right]$$

References

[1] Arnold, B. (1968), "Parameter estimation for a multivarite exponential distribution", *Journal of the American Statistical Association*, vol. 63, 848–852.

[2] 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*, vol. 67, 927–929.

[3] Bhattacharyya, G. K. and Johnson, R. A., (1971), "Maximum Likelihood Estimation and Hypothesis Testing in the Bivariate Exponential Model of Marshall and Olkin", http://handle.dtic.mil/100.2/AD737530.

[4] Block, H. and Basu, A. P. (1974), "A continuous bivariate exponential extension", *Journal of the American Statistical Association*, vol. 69, 1031–1037.

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

[6] Hanagal, D.D. (2005), "A bivariate Weibull regression model", *Economic Quality Control*, vol. 20, 143–150.

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

[8] Karlis, D. (2003), "ML estimation for multivarite shock models via an EM algorithm", *Annals of the Institute of Statistical Mathematics*, 55, 817–830.

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

[10] Kundu, D. and Gupta, R.D. (2006), "Estimation of P(Y < X) for Weibull distribution", *IEEE Transactions on Reliability*, vol. 55, 270–280.

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

[12] Lu, Jye-Chyi (1989), "Weibull extension of the Freund and Marshall-Olkin bivariate exponential model", *IEEE Transactions on Reliability*, vol. 38, 615–619.

[13] 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.

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

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

[16] Patra, K. and Dey, D.K. (1999), "A multivarite mixture of Weibull distributions in reliability modeling", *Statistics and Probability Letters*, vol. 45, 225–235.

[17] Proschan, F. and Sullo, P. (1976), "Estimating the parameters of a multivarite exponential distribution", *Journal of the American Statistical Association*, vol. 71, 465–472.