Estimation of Five Parameter Bivariate Modified Weibull Singular Distribution

Sanjay Kumar[1], Debasis Kundu[2] & Sharmishtha Mitra[3]

Department of Mathematics & Statistics, Indian Institute of Technology Kanpur,

Kanpur, Pin 208016, India

Abstract

Recently, El-Bassiouny et al. [5] proposed the flexible Marshall-Olkin type bivariate modified Weibull distribution (BMWD) and obtained the maximum likelihood estimates (MLEs) of the five unknown parameters of the BMWD by solving five non-linear equations. The MLEs of parameters cannot be obtained in closed form. In this paper, we discuss about the computation of the MLEs using the efficient EM algorithm, which reduces the problem to a two-dimensional (2-D) optimization problem. The proposed EM algorithm performs better than the standard existing optimization algorithms, e.g., optim and nlm. The bivariate goodness-of-fit (GoF) test is discussed based on statistically equivalence blocks (SEBs). Some nonstandard likelihood ratio (LR) tests are considered for testing the relative GoF to the BMWD. An extensive simulation study is carried out. The EM algorithm performs quite well. A real-life data set is analyzed for illustrative purposes.

Keywords: Modified Weibull distribution, Bathtub-shaped hazard rate distribution, The EM algorithm, Maximum likelihood estimation, Likelihood ratio test, Bivariate GoF test.

1 Introduction

The Marshall-Olkin bivariate Weibull (MOBW) distribution, see [10], plays a significant role in life-testing experiments, reliability and survival analysis. Its marginals are Weibull distributions having increasing, decreasing and constant hazard rate functions (HRFs). The major weakness of the MOBW distribution is that its marginals cannot handle non-monotone hazard rates

✉[1] sanjaymk@iitk.ac.in, ✉[2] kundu@iitk.ac.in, ✉[3] smitra@iitk.ac.in

(in particular, bathtub-shaped hazard rates). The BMWD, see [5], is a generalization of the MOBW distribution, whose marginals are modified Weibull distributions (MWDs) that have bathtub-shaped, increasing, decreasing and constant HRFs, see [8]. The BMWD is of the Marshall-Olkin type, having both absolutely continuous and singular parts. Because of this reason, the BMWD can be used quite effectively to analyze a bivariate data set with ties. The BMWD can also be used effectively to analyze competing risks data. However, if the Weibull probability paper (WPP) plot for each marginal of a bivariate data set is a convex curve, see [8], the BMWD fits the data more appropriately than other bivariate distributions whose marginals have bathtub-shaped, decreasing, increasing, and constant HRFs, e.g., bivariate exponentiated Weibull distribution, see [17], bivariate Chen distribution, see [15], etc. In the UEFA data, see [12], the WPP plot for marginals is approximated by the convex curve and likewise for the Cholesterol level data, see [18], we have analyzed in Section 6.

The Marshall-Olkin bivariate exponential (MOBE) distribution, see [11], is a particular case of the MOBW distribution. Some properties and statistical inferences of bivariate exponentiated MWD have been studied by Shahen et al. [17]. El-Bassiouny et al. [5] obtained the MLEs of the unknown parameters of the BMWD by solving five non-linear equations, but did not discuss any numerical optimization technique to find the MLEs, model selection criteria, model validation criteria, asymptotic optimal properties, and hypothesis testing problems. Due to the computational complexity involved in the parameter estimation of the BMWD, not much work has been done for this distribution. The BMWD has the same physical interpretation as the MOBE distribution, see [11].

The main motivation of this paper is two-fold. The first is to show how the proposed EM algorithm can be implemented in a five parameter bivariate singular distribution. The major advantage of the EM algorithm is that it reduces the five dimensional optimization problem to a two dimensional one. Also we have seen that the EM algorithm with a particular convergence criterion converges to the same result irrespective of the initial guesses we take for the five unknown parameters. Secondly, we implement the bivariate Goodness-of-fit (GoF) test for a bivariate distribution with singularity. The main aim of this article is to highlight the efficient computation of the MLEs of five unknown parameters of the BMWD. The MLEs cannot be obtained in closed form see Bemis et al. [4]. Though the parameter estimation of the BMWD is a 5-D optimization problem, it is observed that the EM algorithm can be used quite effectively to compute the MLEs by solving a 2-D optimization problem at each iteration. Using Louis [9], the observed Fisher information matrix (FIM) can be obtained and can be used for constructing asymptotic confidence intervals (CIs) for the unknown parameters and for testing purposes. We

have performed some simulation studies and observed that the EM algorithm performs better than the existing optim and nlm algorithms when the initial guesses are far from the true parameter values. The EM algorithm converges to the true parameter values regardless of the initial guesses. For some initial guesses, the standard numerical methods give a negative value for the asymptotic variance of the MLEs. The EM algorithm is " better" in terms of usage of computational resources and convergence to the desired value at a faster rate regardless of initial choices of the parameters. Recently, Meintanis [12], and Kundu and Dey [7] analyzed the UEFA data set using the MOBE and MOBW distributions, respectively. We analyzed the same data set and the Cholesterol level data using BMWD, and we have seen, based on some model selection and validation criteria, that BMWD fits the data better than MOBE and MOBW models. We have also considered the nonstandard likelihood ratio (LR) tests, i.e., the testing problem at the boundary point of the parameter space.

Testing the GoF of a given data set with respect to a given probabilistic model is an essential aspect of any data analysis. Most GoF tests have been developed for absolutely continuous multivariate distributions. The two most important tests of GoF based on the empirical distribution function (edf), the Kolmogorov-Smirnov (KS) and the Cramer-von Mises statistics, have been extended to the multivariate case. The problem is that the probability distribution of these test statistics is not distribution-free as in the univariate case, if the multivariate distribution is not absolutely continuous. Recently, many multivariate GoF tests have been developed, based on data depth, see Zhang et al. [21], Singh et al. [19], based on Wasserstein distance, see Marc Hallin et al. [6], etc. These can be used to test only for absolutely continuous multivariate distributions. Since the BMWD is not absolutely continuous, we consider a bivariate GoF test based on SEBs using the idea of Alam et al. [1, 2]. The critical value and the corresponding p-value of the test is obtained by the parametric bootstrap method.

The rest of the paper is organized as follows. In Section 2, we introduce the models. We discusse the ML estimation method with non-standard LR tests in Section 3. In Section 4, we discusse the bivariate GoF test. Simulation results and data analysis are presented in Section 5 and Section 6, respectively. Finally, we conclude the paper in Section 7.

2 Model Description

The probability density function (PDF) and the cumulative distribution function (CDF) of the MWD, proposed by Lai et al. [8], with scale parameter $\alpha > 0$, shape parameter $\beta > 0$ and

accelerate parameter $\lambda \geq 0$, denoted by $MW(\alpha, \beta, \lambda)$, are given for $t > 0$ as;

$$f_{MW}(t; \alpha, \beta, \lambda) = \alpha t^{(\beta-1)}(\beta + \lambda t)e^{\lambda t}e^{-\alpha t^{\beta}e^{\lambda t}} \quad \text{and} \quad F_{MW}(t; \alpha, \beta, \lambda) = 1 - e^{-\alpha t^{\beta}e^{\lambda t}} \tag{1}$$

respectively. The HRF of the MWD, given as $h(t) = \alpha t^{(\beta-1)}(\beta + \lambda t)e^{\lambda t}$, is increasing when $\beta \geq 1$, constant when $\beta = 1, \lambda = 0$, decreasing when $0 < \beta < 1, \lambda = 0$, and for $0 < \beta < 1, \lambda > 0$ it is bathtub-shaped.

Suppose $U_i \sim MW(\alpha_i, \beta, \lambda);\ i = 1, 2, 3$, be three independent modified Weibull random variables. Define $X_1 = \min\{U_1, U_3\}$ and $X_2 = \min\{U_2, U_3\}$, then the bivariate random vector (X_1, X_2) has the BMWD with parameters $\alpha_1, \alpha_2, \alpha_3, \beta$ and λ, and we denote it by $(X_1, X_2) \sim BMW(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$. The $BMW(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$ reduces to $MOBW(\alpha_1, \alpha_2, \alpha_3, \beta)$ when $\lambda = 0$, $MOBE(\alpha_1, \alpha_2, \alpha_3)$ when $\beta = 1,\ \lambda = 0$ and bivariate Rayleigh distribution when if $\beta = 2$, $\lambda = 0$. The BMWD is Marshall-Olkin type and has both absolutely continuous and singular parts. The joint survival function (SF) of $(X_1, X_2) \sim BMW(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$ for $x_1, x_2 > 0$, is given by, $S_{(X_1,X_2)}(x_1, x_2) = e^{-\alpha_1 x_1^{\beta}e^{\lambda x_1} - \alpha_2 x_2^{\beta}e^{\lambda x_2} - \alpha_3 z^{\beta}e^{\lambda z}}$, where $z = \max\{x_1, x_2\}$, or equivalently

$$S_{(X_1,X_2)}(x_1, x_2) = \begin{cases} e^{-\alpha_1 x_1^{\beta}e^{\lambda x_1}}e^{-(\alpha_2+\alpha_3)x_2^{\beta}e^{\lambda x_2}} & \text{if} \quad x_1 < x_2 \\ e^{-(\alpha_1+\alpha_3)x_1^{\beta}e^{\lambda x_1}}e^{-\alpha_2 x_2^{\beta}e^{\lambda x_2}} & \text{if} \quad x_1 > x_2 \\ e^{-(\alpha_1+\alpha_2+\alpha_3)x^{\beta}e^{\lambda x}} & \text{if} \quad x_1 = x_2 = x, \end{cases} \tag{2}$$

and the corresponding joint PDF of (X_1, X_2) is given by

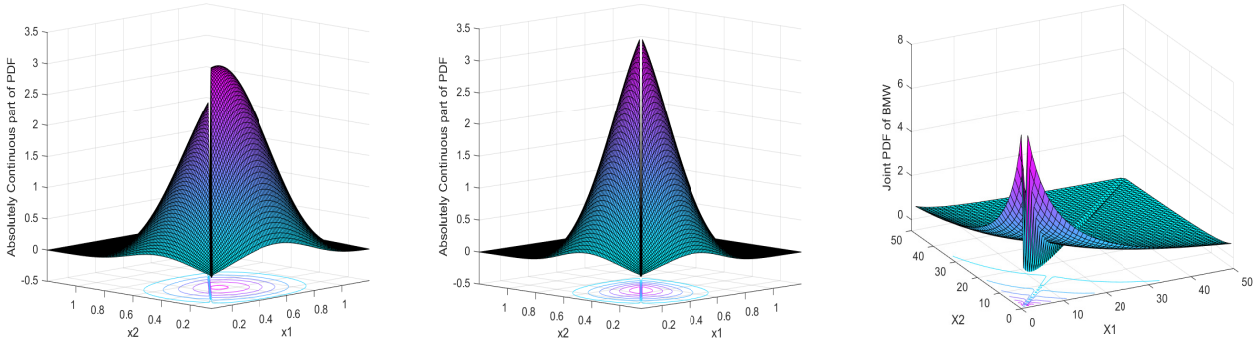
$$f_{(X_1,X_2)}(x_1, x_2) = \begin{cases} f_1(x_1, x_2) & \text{if} \quad 0 < x_1 < x_2 < \infty \\ f_2(x_1, x_2) & \text{if} \quad 0 < x_1 > x_2 < \infty \\ f_3(x) & \text{if} \quad 0 < x_1 = x_2 = x < \infty, \end{cases} \tag{3}$$

where

$$\begin{aligned} f_1(x_1, x_2) &= f_{MW}(x_1; \alpha_1, \beta, \lambda) f_{MW}(x_2; \alpha_2 + \alpha_3, \beta, \lambda) \\ &= \alpha_1(\alpha_2 + \alpha_3)x_1^{\beta-1}(\beta + \lambda x_1)e^{\lambda x_1 - \alpha_1 x_1^{\beta}e^{\lambda x_1}} x_2^{\beta-1}(\beta + \lambda x_2)e^{\lambda x_2 - (\alpha_2+\alpha_3)x_2^{\beta}e^{\lambda x_2}} \\ f_2(x_1, x_2) &= f_{MW}(x_1; \alpha_1 + \alpha_3, \beta, \lambda) f_{MW}(x_2; \alpha_2, \beta, \lambda) \\ &= (\alpha_1 + \alpha_3)\alpha_2 x_1^{\beta-1}(\beta + \lambda x_1)e^{\lambda x_1 - (\alpha_1+\alpha_3)x_1^{\beta}e^{\lambda x_1}} x_2^{\beta-1}(\beta + \lambda x_2)e^{\lambda x_2 - \alpha_2 x_2^{\beta}e^{\lambda x_2}} \\ f_3(x) &= \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} f_{MW}(x; \alpha_1 + \alpha_2 + \alpha_3, \beta, \lambda) = \alpha_3 x^{\beta-1}(\beta + \lambda x)e^{\lambda x}e^{-(\alpha_1+\alpha_2+\alpha_3)x^{\beta}e^{\lambda x}} \end{aligned}$$

Note that the function $f_{(X_1,X_2)}(x_1, x_2)$ may be considered to be a joint PDF of the BMWD, where the first two components $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$ are the absolutely continuous part which is the PDF with respect to the 2-D Lebesgue measure on the first quadrant and the third component $f_3(x)$ is the singular part which is a PDF with respect to the 1-D Lebesgue measure on the positive real line, see, e.g., Bemis et al. [4]. The joint PDF of the BMWD has a variety of shapes, see Figure 1. The following lemma helps us to give an idea for the choice of initial guesses in the parameter estimation.

Figure 1: Plots of absolutely continuous part of PDF of BMWD



(a) BMW(2.0, 1.0, 0.5, 2.0, 1.0) (b) BMW(1.0, 1.0, 1.0, 2.0, 1.5) (c) BMW(1.5, 1.5, 1.5, 0.5, 0.1)

Lemma 2.1. *If* $(X_1, X_2) \sim BMW(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$*, then*

(a) $\min(X_1, X_2) \sim MW(\alpha_1 + \alpha_2 + \alpha_3, \beta, \lambda)$

(b) $P(X_1 < X_2) = \frac{\alpha_1}{\alpha_1+\alpha_2+\alpha_3}$

(c) $P(X_1 > X_2) = \frac{\alpha_2}{\alpha_1+\alpha_2+\alpha_3}$

(d) $P(X_1 = X_2) = \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}$

(e) The marginal distribution of X_i *is* $MW(\alpha_i + \alpha_3, \beta, \lambda)$*, i=1,2.*

Proof. The proof of (a) - (d) is trivial. For the proof of (e), from the joint SF of (X_1, X_2), the marginal SF of X_i, $i = 1, 2$, can be obtained, by setting $x_j \to 0$, $i \neq j$, i.e., $\lim_{x_j \to 0} S_{(X_1,X_2)}(x_1, x_2) = S_{X_i}(x_i) = e^{-(\alpha_i+\alpha_3)x_i^{\beta}e^{\lambda x_i}}$ this implies that $X_i \sim MW(\alpha_i + \alpha_3, \beta, \lambda)$, $i = 1, 2$. ■

3 Estimation of Parameters of the BMWD

The MLEs of the five unknown parameters of the BMWD do not exist in explicit form. El-Bassiouny et al. [5] obtained the MLEs of parameters of the BMWD with the numerical routines to solve the five non-linear likelihood equations. The estimation of parameters is a 5-D optimization problem. In optimization, we need the initial guesses for the parameters. When the dimension of optimization problem increases, the common standard optimization techniques, e.g., optim and nlm methods, pick up the local optima and may fail to find the covariance matrix of MLEs for some initial guesses. Thus the usual standard optimization techniques are more sensitive with initial guesses when the dimension of the problem increases. To avoid this problem, we propose to use the EM algorithm. It is easily implemented in the 5-D problem. The EM algorithm is used to find the MLEs of parameters of the BMWD by solving only a 2-D optimization problem. The EM algorithm converges to the true parameter values, whatever be the initial guesses we have taken.

3.1 Maximum Likelihood Estimation using the EM Algorithm

In this subsection, we address the problem of computing the MLEs of parameters of the BMWD. Suppose the n observed data $\mathcal{D} = \{(x_{11}, x_{21}), (x_{12}, x_{22}), ...(x_{1n}, x_{2n})\}$ are coming from $BMW(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$. Let us define the following notations for further development,
$I_1 = \{(x_{1i}, x_{2i}) : x_{1i} < x_{2i}\}$, $I_2 = \{(x_{1i}, x_{2i}) : x_{1i} > x_{2i}\}$, $I_3 = \{(x_{1i}, x_{2i}) : x_{1i} = x_{2i} = x_i\}$
and $n_1 = |I_1|$, $n_2 = |I_2|$, $n_3 = |I_3|$ with $n = n_1 + n_2 + n_3$. Here $|I_j|$ for $j = 1, 2, 3$ denotes the number of elements in the set I_j. The log-likelihood function is given by

$$
\begin{aligned}
l(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda) = {} & n_1 \ln \alpha_1 + n_1 \ln(\alpha_2 + \alpha_3) + (\beta - 1) \sum_{i \epsilon I_1} \ln x_{1i} + \sum_{i \epsilon I_1} \ln(\beta + \lambda x_{1i}) + \lambda \sum_{i \epsilon I_1} x_{1i} \\
& - \alpha_1 \sum_{i \epsilon I_1} x_{1i}^{\beta} e^{\lambda x_{1i}} + (\beta - 1) \sum_{i \epsilon I_1} \ln x_{2i} + \sum_{i \epsilon I_1} \ln(\beta + \lambda x_{2i}) + \lambda \sum_{i \epsilon I_1} x_{2i} - (\alpha_2 + \alpha_3) \sum_{i \epsilon I_1} x_{2i}^{\beta} e^{\lambda x_{2i}} \\
& + n_2 \ln \alpha_2 + n_2 \ln(\alpha_1 + \alpha_3) + (\beta - 1) \sum_{i \epsilon I_2} \ln x_{1i} + \sum_{i \epsilon I_2} \ln(\beta + \lambda x_{1i}) + \lambda \sum_{i \epsilon I_2} x_{1i} \\
& - (\alpha_1 + \alpha_3) \sum_{i \epsilon I_2} x_{1i}^{\beta} e^{\lambda x_{1i}} + (\beta - 1) \sum_{i \epsilon I_2} \ln x_{2i} + \sum_{i \epsilon I_2} \ln(\beta + \lambda x_{2i}) + \lambda \sum_{i \epsilon I_2} x_{2i} - \alpha_2 \sum_{i \epsilon I_2} x_{2i}^{\beta} e^{\lambda x_{2i}} \\
& + n_3 \ln \alpha_3 + (\beta - 1) \sum_{i \epsilon I_3} \ln x_i + \sum_{i \epsilon I_3} \ln(\beta + \lambda x_i) + \lambda \sum_{i \epsilon I_3} x_i - (\alpha_1 + \alpha_2 + \alpha_3) \sum_{i \epsilon I_3} x_i^{\beta} e^{\lambda x_i}.
\end{aligned}
$$

$$
\begin{aligned}
&= n_1 \ln \alpha_1 + n_2 \ln \alpha_2 + n_3 \ln \alpha_3 + n_1 \ln(\alpha_2 + \alpha_3) + n_2 \ln(\alpha_1 + \alpha_3) \\
&\quad + (\beta - 1)\Bigg[\sum_{i=1}^{n} \ln x_{1i} + \sum_{i \epsilon I_1 \cup I_2} \ln x_{2i}\Bigg] + \sum_{i=1}^{n} \ln(\beta + \lambda x_{1i}) + \sum_{i \epsilon I_1 \cup I_2} \ln(\beta + \lambda x_{2i}) \\
&\quad + \lambda\Bigg[\sum_{i=1}^{n} x_{1i} + \sum_{i \epsilon I_1 \cup I_2} x_{2i}\Bigg] - \alpha_1 \sum_{i=1}^{n} x_{1i}^{\beta} e^{\lambda x_{1i}} - \alpha_2 \sum_{i=1}^{n} x_{2i}^{\beta} e^{\lambda x_{2i}} \\
&\quad - \alpha_3\Bigg[\sum_{i \epsilon I_1} x_{2i}^{\beta} e^{\lambda x_{2i}} + \sum_{i \epsilon I_2} x_{1i}^{\beta} e^{\lambda x_{1i}} + \sum_{i \epsilon I_3} x_i^{\beta} e^{\lambda x_i}\Bigg]. \qquad (4)
\end{aligned}
$$

In the BMWD we have seen from the first partial derivatives of the log-likelihood, the MLEs do not exist if $n_1 n_2 n_3 = 0$, e.g. when $\beta = 1, \lambda = 0$ see Bemis et al. [4]. The MLEs exist when all $n_i > 0$ and they can be obtained by maximizing the log-likelihood function (4) with respect to $\alpha_1, \alpha_2, \alpha_3, \beta,$ and λ. That is, MLEs are obtained by solving five non-linear likelihood equations, but they cannot be obtained in closed forms as expected. By using the EM algorithm we can easily obtain the closed form expression of MLEs for $\alpha_1, \alpha_2,$ and α_3 and the problem reduces to a 2-D optimization problem to find the MLEs of β and λ. Kundu and Dey [7] proposed an EM algorithm based on fractioning the observed data into two partially complete pseudo observations, to find the MLEs of the parameters of MOBW distribution. Here we use the idea of Kundu and Dey to develop the EM algorithm for the BMWD. It is easy to show that estimation of parameters of the BMWD can be seen as a missing data problem. To treat this as missing data problem we consider a new bivariate random vector (Z_1, Z_2) associated with the given bivariate random vector (X_1, X_2) by

$$
Z_1 = \begin{cases} 1 & \text{if } X_1 = U_1 \\ 3 & \text{if } X_1 = U_3 \end{cases} \qquad \text{and} \qquad Z_2 = \begin{cases} 2 & \text{if } X_2 = U_2 \\ 3 & \text{if } X_2 = U_3 \end{cases} \qquad (5)
$$

It is clear that in this setup even if we know (X_1, X_2), the corresponding (Z_1, Z_2) may not be known always. For example, if $X_1 = X_2$, then $Z_1 = Z_2 = 3$, but if $X_1 \neq X_2$, then (Z_1, Z_2) is not known, i.e., we treat them as missing observations. For $(x_1, x_2) \in I_1$ the possible values of (Z_1, Z_2) are (1,3) or (1,2) and similarly for $(x_1, x_2) \in I_2$ the possible values of (Z_1, Z_2) are (3,2) or (1,2), with non-zero probabilities. For $(x_1, x_2) \in I_3$ the possible value of (Z_1, Z_2) is (3,3) with probability 1. If $(x_1, x_2) \in I_1$ the transformed observations can be obtained by (x_1, x_2, r_1) and (x_1, x_2, r_2), where

$$r_1 = P\{(Z_1, Z_2) = (1,3) | X_1 < X_2\} = P(U_1 < U_3 < U_2 | X_1 < X_2) = \frac{\alpha_3}{\alpha_2 + \alpha_3}$$
$$\text{and} \quad r_2 = P\{(Z_1, Z_2) = (1,2) | X_1 < X_2\} = P(U_1 < U_2 < U_3 | X_1 < X_2) = \frac{\alpha_2}{\alpha_2 + \alpha_3}. \quad (6)$$

If $(x_1, x_2) \epsilon I_2$ the transformed observations can be obtained by (x_1, x_2, s_1) and (x_1, x_2, s_2), where

$$s_1 = P\{(Z_1, Z_2) = (3,2) | X_2 < X_1\} = P(U_2 < U_3 < U_1 | X_2 < X_1) = \frac{\alpha_3}{\alpha_1 + \alpha_3}$$
$$\text{and} \quad s_2 = P\{(Z_1, Z_2) = (1,2) | X_2 < X_1\} = P(U_2 < U_1 < U_3 | X_1 < X_2) = \frac{\alpha_1}{\alpha_1 + \alpha_3}. \quad (7)$$

Let $\vartheta = (\beta, \lambda)$, then the pseudo log-likelihood function is given by

$$\begin{aligned} l_p(\alpha_1, \alpha_2, \alpha_3, \vartheta) &= (n_1 + s_2 n_2) \ln \alpha_1 - \alpha_1 \Big[\sum_{i \epsilon I_1 \cup I_2} x_{1i}^{\beta} e^{\lambda x_{1i}} + \sum_{i \epsilon I_3} x_i^{\beta} e^{\lambda x_i} \Big] \\ &+ (r_2 n_1 + n_2) \ln \alpha_2 - \alpha_2 \Big[\sum_{i \epsilon I_1 \cup I_2} x_{2i}^{\beta} e^{\lambda x_{2i}} + \sum_{i \epsilon I_3} x_i^{\beta} e^{\lambda x_i} \Big] \\ &+ (r_1 n_1 + s_1 n_2 + n_3) \ln \alpha_3 - \alpha_3 \Big[\sum_{i \epsilon I_1} x_{2i}^{\beta} e^{\lambda x_{2i}} + \sum_{i \epsilon I_2} x_{1i}^{\beta} e^{\lambda x_{1i}} + \sum_{i \epsilon I_3} x_i^{\beta} e^{\lambda x_i} \Big] \\ &+ (\beta - 1) \Big[\sum_{i \epsilon I_1 \cup I_2} \ln x_{1i} + \sum_{i \epsilon I_1 \cup I_2} \ln x_{2i} + \sum_{i \epsilon I_3} \ln x_i \Big] + \lambda \Big[\sum_{i \epsilon I_1 \cup I_2} x_{1i} + \sum_{i \epsilon I_1 \cup I_2} x_{2i} + \sum_{i \epsilon I_3} x_i \Big] \\ &+ \sum_{i \epsilon I_1 \cup I_2} \ln(\beta + \lambda x_{1i}) + \sum_{i \epsilon I_1 \cup I_2} \ln(\beta + \lambda x_{2i}) + \sum_{i \epsilon I_3} \ln(\beta + \lambda x_i). \end{aligned} \quad (8)$$

For fixed $\vartheta = (\beta, \lambda)$, MLEs of the parameters $\alpha_1, \alpha_2, \alpha_3$ can be obtained by maximizing the log-likelihood (8) with respect to $\alpha_1, \alpha_2, \alpha_3$. The MLEs are obtained as

$$\hat{\alpha}_1(\vartheta) = \frac{(n_1 + s_2 n_2)}{\sum_{i \epsilon I_1 \cup I_2} x_{1i}^{\beta} e^{\lambda x_{1i}} + \sum_{i \epsilon I_3} x_i^{\beta} e^{\lambda x_i}}, \; \hat{\alpha}_2(\vartheta) = \frac{(r_2 n_1 + n_2)}{\sum_{i \epsilon I_1 \cup I_2} x_{2i}^{\beta} e^{\lambda x_{2i}} + \sum_{i \epsilon I_3} x_i^{\beta} e^{\lambda x_i}}$$
$$\hat{\alpha}_3(\vartheta) = \frac{r_1 n_1 + s_1 n_2 + n_3}{\sum_{i \epsilon I_1} x_{2i}^{\beta} e^{\lambda x_{2i}} + \sum_{i \epsilon I_2} x_{1i}^{\beta} e^{\lambda x_{1i}} + \sum_{i \epsilon I_3} x_i^{\beta} e^{\lambda x_i}}, \quad (9)$$

and the MLE of $\vartheta = (\beta, \lambda)$ is obtained by maximizing the profile log-likelihood function using standard 2-D optimization techniques, e.g., optim and nlm algorithms, i.e.,

$$\widehat{\vartheta} = \arg\max_{\vartheta} \; l_p\big(\widehat{\alpha}_1(\vartheta), \widehat{\alpha}_2(\vartheta), \widehat{\alpha}_3(\vartheta), \vartheta\big). \quad (10)$$

Only for the 1-D optimization problems, the optim and nlm algorithms converge to the same point regardless of how large or small initial guess values we took. In a 2-D optimization problem, the optim and nlm are less sensitive to initial guesses than in 3-D and higher dimensional optimization problems. But in the EM algorithm, at each iteration, we can use optim and nlm to solve the 2-D optimization problem. For the optim and nlm methods, at each iteration, the initial guess values change and become closer and closer to the true parameter values. Hence in the EM algorithm, the optim and nlm methods eventually (at the last iteration of the EM algorithm) converge to the same point for any initial guess values at which the methods do not fail to give parameter estimates.

Steps for the EM algorithm

Here we describe how to obtain the $(k+1)^{th}$ step from the k^{th} step of the EM algorithm. Suppose at the k^{th} step the estimates of $\alpha_1, \alpha_2, \alpha_3, \vartheta$ are $\alpha_1^{(k)}, \alpha_2^{(k)}, \alpha_3^{(k)}, \vartheta^{(k)}$, respectively, $k = 0, 1, 2, ...$

- **Step 1. Expectation**
- Compute $r_1^{(k+1)}, r_2^{(k+1)}, s_1^{(k+1)}, s_2^{(k+1)}$ using (6) & (7)
- **Step 2. Maximization**
- Find the MLE $\vartheta^{(k+1)} = \left(\beta^{(k+1)}, \lambda^{(k+1)}\right)$ from (10) using optim or nlm.
- Once $\vartheta^{(k+1)}$ is obtained, compute $\alpha_1^{(k+1)}, \alpha_2^{(k+1)}, \alpha_3^{(k+1)}$ using (9).

repeat the steps 1 and 2 until a convergence criterion is met. For example, the iteration stops when the ratio $|(l_p(k) - l_p(k-1))/l_p(k-1)| < 10^{-8}$, where $l_p(k)$ represents the value of pseudo log-likelihood function at the k^{th} iteration. The pseudo code of the EM algorithm is given in Appendix II. We provide the calculation of the observed FIM for the BMWD at the MLEs and is given in Appendix I. It has been used to estimate asymptotic covariance matrix of MLEs and to compute the asymptotic CIs of the unknown parameters.

3.2 Hypothesis Testing

Suppose a random sample is drawn from $(X_1, X_2) \sim BMW(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$. Let $\hat{\theta}$ be the MLE of $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$. Let $LL(\hat{\theta}_0)$ be the restricted log-likelihood value under the null hypothesis and $LL(\hat{\theta})$ be the unrestricted log-likelihood value. Now we discuss the following problems of testing:

(1) We want to test whether the marginals X_1 and X_2 have the same distribution or not, i.e., $H_0 : \alpha_1 = \alpha_2$ *versus* $H_a : \alpha_1 \neq \alpha_2$. Then the limiting distribution of the test statistics $T = 2\big[LL(\hat{\theta}) - LL(\hat{\theta}_0)\big]$ under the null hypothesis is χ^2_1.

(2) If $\lambda = 0$, the BMW model becomes $MOBW(\alpha_1, \alpha_2, \alpha_3, \beta)$. We test whether the addition of parameter λ is significant or not, i.e., $H_0 : \lambda = 0$ $(MOBW)$ *versus* $H_a : \lambda > 0$ (BMW). This is the one sided LR test and under the null hypothesis the true parameter of interest λ is at the boundary of the parameter space $\Theta_5 = [0, \infty)$. One can still use the LR test. However, the limiting distribution under the null hypothesis is not χ^2_1, for more details see Self & Liang [16]. The limiting distribution of test statistic $T = 2\big[LL(\hat{\theta}) - LL(\hat{\theta}_0)\big]$ is a mixture of a degenerate distribution with point mass at zero and a chi-square distribution with one degree of freedom, i.e., $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$, where $P\big(\chi^2_0 = 0\big) = 1$. The p-value of the asymptotic LR test is obtained as $0.5P(\chi^2_0 > T_{cal}) + 0.5P(\chi^2_1 > T_{cal})$.

(3) The problem is to test whether the parameter β is significant or not, i.e. $H_0 : \beta = 0$ *versus* $H_a : \beta > 0$. This is also a one sided and nonstandard LR test where the true parameter of interest is at the boundary of the parameter space. The test statistic $T = 2\big[LL(\hat{\theta}) - LL(\hat{\theta}_0)\big]$ follows 50:50 mixture of χ^2_0 and χ^2_1, as mentioned in the part (2).

(4) If $\lambda = 0$, $\beta = b_0$, where b_0 is either 1 or 2, the BMW model reduces to $MOBE(\alpha_1, \alpha_2, \alpha_3)$ when $b_0 = 1$ and bivariate Rayleigh distribution when $b_0 = 2$. We test the hypothesis $H_0 : \lambda = 0,\ \beta = b_0$ *versus* $H_a : \lambda > 0,\ \beta \neq b_0$. This is another nonstandard LR test where there are two parameters of interest in which one parameter λ is a boundary point and other is an interior point, see Self & Liang [16]. In this testing problem, the limiting distribution of LR test statistic $T = 2\big[LL(\hat{\theta}) - LL(\hat{\theta}_0)\big]$ follows 50:50 mixture of χ^2_1 and χ^2_2, i.e. $\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_2$. The p-value of the asymptotic LR test is obtained as $0.5P(\chi^2_1 > T_{cal}) + 0.5P(\chi^2_2 > T_{cal})$.

4 Bivariate GoF for the BMWD

In this section, using the same idea of Alam et al. [1, 2] we consider a bivariate GoF based on SEB. For more details see, Anderson [3], Pyke [14]. The SEB yield spacings which are distributed as sample spacings in the univariate case. Let $\mathcal{D} = \{(x_{11}, x_{21}), ..., (x_{1n}, x_{2n})\}$ be n observations from a population with unkown CDF $F(\cdot)$. We want to test that the given sample

is drawn from a family of $BMW(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$, i.e.,

$$H_0 : F = F_0 \ \forall \ (x_1, x_2) \ \ versus \quad H_0 : F \neq F_0 \ \text{ for at least one } (x_1, x_2),$$

where $F_0 = F_0(x; \alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$ is the CDF of the BMWD. Since the parameters $(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$ are unknown, the null hypothesis H_0 is composite and the test is known as composite GoF test.

Let us consider the alternating components of the bivariare observations in the sample, denoted as $z_{(1),1}, z_{(2),2}, z_{(3),1}, z_{(4),2}, ...$, where $z_{(r),j}$ denotes the smallest value among the j^{th} components, $j = 1, 2$, in the sample, excluding the observations associated with $z_{(1),1}, z_{(2),2}, z_{(3),1}, ..., z_{(r-1),j'}$, where r=1,2,...,n and

$$j' = \begin{cases} j+1 & \text{for } j = 1 \\ j-1 & \text{for } j = 2. \end{cases}$$

That is, $z_{(1),1}$ denotes the smallest value among the first components, $z_{(2),2}$ denotes the smallest value among the second components, discarding the observation associated with $z_{(1),1}$, etc. The partial sums of coverages are given by

$$v_1 = 1 - S_{X_1}(z_{(1),1}), \qquad v_2 = 1 - S_{(X_1,X_2)}(z_{(1),1}, z_{(2),2})$$

and for $1 \leq t \leq N$, $N = \frac{n-1}{2}$ or $N = \frac{n-2}{2}$ according as n is odd or even

$$v_{2t+1} = 1 - S_{(X_1,X_2)}(z_{(2t+1),1}, z_{(2t),2}) \qquad v_{2t+2} = 1 - S_{(X_1,X_2)}(z_{(2t+1),1}, z_{(2t+2),2})$$

where, $S_{(X_1,X_2)}(x_1, x_2)$ is the joint SF (2) of BMWD and $S_{X_1}(x_1)$ is the marginal SF (1) of X_1. The KS test statistic (two-sided) is

$$D_n = \max\left\{0, \max_{1\leq i\leq n}\left(\frac{i}{n} - v_i\right), \max_{1\leq i\leq n}\left(v_i - \frac{i-1}{n}\right)\right\}. \tag{11}$$

In our GoF test problem the null hypothesis is composite so the asymptotic distribution of D_n is obtained by the parametric bootstrap method. Thus for real data case the critical value and p-value of the GoF test are obtained by the parametric bootstrap method briefly discussed in section 5.

4.0.1 Simulation of GoF for BMWD

In this section, we perform the simulation using 10^4 replications to check the performance of the bivariate KS test. We report the average KS test statistic and the corresponding p-value for different sample sizes and different parameter combinations in Table 1. From Table 1 we see that, when the sample sizes increase the KS distance decreases.

Table 1: KS distance and p-value for BMWD

True parameters $(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda) = (0.5, 0.4, 0.6, 1.5, 1.0)$				
$n \rightarrow$	35	50	100	500
KS distance	0.1081	0.0691	0.0269	0.0169
p-value	0.8084	0.9709	0.9999	0.9988
True parameters $(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda) = (1.0, 1.0, 1.0, 1.5, 1.0)$				
$n \rightarrow$	35	50	100	500
KS distance	0.1980	0.1326	0.0934	0.0486
p-value	0.1287	0.3431	0.3477	0.9884
True parameters $(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda) = (1.5, 0.5, 1.2, 1.5, 1.0)$				
$n \rightarrow$	35	50	100	500
KS distance	0.1166	0.0756	0.0304	0.0254
p-value	0.7277	0.9378	0.9999	0.9029

5 Simulation Study

In this section, we give some simulation results to check the performance of the EM algorithm for different parameter combinations and sample sizes using the same stopping criterion for each case. The iteration stops when the ratio $|(l(k) - l(k-1))/l(k-1)| < 10^{-8}$, where $l(k)$ represents the value of the log-likelihood at the k^{th} iteration. In each iteration of the EM algorithm, we use the optim (e.g., Nelder-Mead method, see [13]) available in the R software for the 2-D optimization problem to find the MLEs of the parameters β and λ. The average estimates (AE) and mean square error (MSE) give us an idea about the bias and consistency of the estimators. The average length of confidence interval (ALCI) and the average standard error (ASE) of parameters are obtained from the observed FIM to check the behavior of estimators obtained by the EM algorithm. The coverage probability (CP) based on 95% CI is the proportion of times the CIs contain the true parameter value.

For the true parameter values $\alpha_1 = 1.0$, $\alpha_2 = 1.0$, $\alpha_3 = 1.0$, $\beta = 1.5$, $\lambda = 1.0$ and different sample sizes $n = 25, 50, 100, 250, 500, 1000$ with initial guesses 0.5, 0.5, 0.5, 0.5, 0.5 of

$\alpha_1, \alpha_2, \alpha_3, \beta, \lambda$ respectively in each sample size, the simulation results obtained. The AE, MSE, CP based on 95% CI, ALCI, ASE and average number of iterations (AI) obtained from the EM algorithm based on 10^4 replications are given in Table 2. Under the same situation the simulation results for the parameter combination (0.6, 0.7, 0.8, 0.5, 0.5) are given in Table 3. Also, we have tried for other choice of initial guesses of parameters $\alpha_1, \alpha_2, \alpha_3, \beta, \lambda$, viz., (1.0, 1.0, 1.0, 1.0, 1.0) and (0.25, 0.25, 0.25, 0.25, 0.25) but have not reported the results here and in all cases, the AE and MSE converge to the same points under the same stopping criterion. From Tables 2 and 3, some of the points are quite clear from the simulation exercise. In each case, the average biases, MSEs and ALCIs decrease as the sample size increases, which indicates that the estimators obtained by the EM algorithm are consistent and efficient. In most of the cases the estimates are positively biased, mainly for the small sample sizes. If the sample size is not very small, the asymptotic normality results can be used for constructing the CIs. The bias and MSE of estimate of β are close to zero even if the sample size is small while the bias of estimate of λ is relatively large. Now from the above experimental results we have seen that the EM algorithm works quite effectively to find the MLEs of the parameters $\alpha_1, \alpha_2, \alpha_3, \beta$ and λ for the BMWD.

Also we have done an extensive simulation study to compare the performance of EM algorithm with the existing optim and nlm algorithms. For this purpose, we take 10^3 replications of different sample sizes, e.g. 25, 50, 100 from BMW(0.05, 0.04, 0.06, 1.5, 0.01). The AE of parameters and the corresponding MSE and the average log-likelihood (Av.LL) value with average time (in seconds) are given in Table 4. We see that the EM algorithm performs better than optim and nlm when the initial guess values are away from the true parameter values for both small and large samples. We have tried for another parameter combinations and initial guesses, but not reported these results here. The EM algorithm converges to the true parameter values regardless of how large or small initial guesses we taken. We have also seen the optim and nlm algorithms fail at some initial guesses that are away from true parameter values. For the large sample cases the optim and nlm become more sensitive with initial guesses. The nlm perform better than optim for both small and large samples. Sometimes the optim and nlm for some initial guesses give the negative value of determinant of the observed FIM.

Table 2: AE, MSE, CP, ALCI and AI for BMW(1.0, 1.0, 1.0, 1.5, 1.0)

$n \downarrow$		α_1	α_2	α_3	β	λ	AI
25	AE	1.2123	1.2178	1.2035	1.5031	1.2678	11.10
	MSE	1.0874	1.0740	1.0029	0.2063	1.1092	
	CP	0.84	0.84	0.84	0.96	0.97	
	ALCI	6.2200	6.2178	6.2918	2.3508	5.2656	
50	AE	1.1607	1.1619	1.1606	1.5085	1.1086	9.97
	MSE	0.5948	0.5940	0.5964	0.1191	0.5851	
	CP	0.88	0.88	0.88	0.96	0.97	
	ALCI	3.7349	3.7465	3.7875	1.5735	3.4618	
100	AE	1.1062	1.1053	1.1066	1.5084	1.0486	9.24
	MSE	0.3322	0.3345	0.3356	0.0683	0.3214	
	CP	0.90	0.90	0.90	0.97	0.97	
	ALCI	2.3689	2.3659	2.3952	1.0832	2.3641	
250	AE	1.0447	1.0453	1.0452	1.5038	1.0208	8.52
	MSE	0.1260	0.1269	0.1299	0.0288	0.1371	
	CP	0.93	0.93	0.93	0.95	0.95	
	ALCI	1.3700	1.3711	1.3844	0.6764	1.4674	
500	AE	1.0245	1.0239	1.0257	1.5026	1.0073	8.06
	MSE	0.0599	0.0603	0.0610	0.0148	0.0686	
	CP	0.94	0.94	0.94	0.95	0.95	
	ALCI	0.9434	0.9429	0.9536	0.4768	1.0325	
1000	AE	1.0116	1.0120	1.0122	1.5011	1.0046	7.68
	MSE	0.0294	0.0291	0.0298	0.0075	0.0351	
	CP	0.95	0.95	0.95	0.95	0.95	
	ALCI	0.6564	0.6566	0.6630	0.3367	0.7285	

6 Data Analysis

6.1 First data: UEFA Data

In this subsection, we apply the BMWD model to a real-life data set for illustrative purpose. We use the soccer data, where at least one goal is scored by the home team and at least one goal is scored directly from a kick goal (kick goal means a penalty kick, free kick, foul kick, or any other direct kick) by any team which was initially analyzed by Meintains [12] using the MOBE model. Many authors have re-analyzed the data by using different bivariate models, e.g., Kundu and Dey [7] to fit the MOBW model. We analyze the data set using the BMWD. We fit the model to the data by using the iterative procedure of the EM algorithm and compare these results with the results obtained by the direct optimization methods optim (Nelder-Mead

Table 3: AE, MSE, CP, ALCI and AI for BMW(0.6, 0.7, 0.8, 0.5, 0.5)

$n \downarrow$		α_1	α_2	α_3	β	λ	AI
25	AE	0.6212	0.7266	0.8184	0.5127	0.5929	10.17
	MSE	0.0824	0.1019	0.1112	0.0138	0.1026	
	CP	0.91	0.91	0.92	0.95	0.95	
	ALCI	1.0695	1.1919	1.2888	0.4528	1.1753	
50	AE	0.6099	0.7121	0.8064	0.5064	0.5436	9.54
	MSE	0.0356	0.0437	0.0519	0.0063	0.0407	
	CP	0.93	0.94	0.93	0.95	0.95	
	ALCI	0.7355	0.8156	0.8811	0.3131	0.7801	
100	AE	0.6032	0.7038	0.8036	0.5023	0.5226	9.13
	MSE	0.0171	0.0216	0.0247	0.0031	0.0193	
	CP	0.94	0.94	0.95	0.95	0.95	
	ALCI	0.5123	0.5673	0.6146	0.2186	0.5347	
250	AE	0.6020	0.7004	0.8021	0.5008	0.5086	8.89
	MSE	0.0069	0.0083	0.0097	0.0012	0.0073	
	CP	0.94	0.95	0.94	0.95	0.95	
	ALCI	0.3223	0.3559	0.3839	0.1374	0.3318	
500	AE	0.6009	0.7007	0.8000	0.5003	0.5043	8.90
	MSE	0.0033	0.0042	0.0048	0.0006	0.0035	
	CP	0.95	0.95	0.95	0.95	0.95	
	ALCI	0.2273	0.2512	0.2717	0.0969	0.2330	
1000	AE	0.6012	0.7009	0.8009	0.5004	0.5019	8.81
	MSE	0.0017	0.0021	0.0024	0.0003	0.0018	
	CP	0.95	0.95	0.95	0.95	0.95	
	ALCI	0.1607	0.1776	0.1922	0.0685	0.1643	

Method) and nlm available in R. Note that *optim()* and *nlm()* are popular functions available in R programming environment for optimization of a function. Also, we compare which model is an appropriate fit for the data set among the three choices, viz., MOBE, MOBW and BMWD. Here X_1 represents the time (in minutes) of first kick goal scored by any team and X_2 denotes the time (in minutes) at which the first goal of any type is scored by the home team. In this data set we have all three possibilities, e.g., $X_1 < X_2$, $X_1 > X_2$ or $X_1 = X_2 = X$(say). In the data set we have n = 37 observations, in which there are $n_1 = 6$ observations in the region $X_1 < X_2$, $n_2 = 17$ observations in the region $X_1 > X_2$ and $n_3 = 14$ observations in the region $X_1 = X_2$. We have $n_1 + n_2 + n_3 = 37$. The scatter plot of the soccer data is given in Figure 2. Several questions may arise from here, such as, what is the effect of kick goal as the first goal scored by any team? Do both the teams have the same number of first kick goals? Which

Table 4: The AE, MSE, Av.LL, average time in seconds for EM, optim and nlm

True Parameter: $\alpha_1 = 0.05$, $\alpha_2 = 0.04$, $\alpha_3 = 0.06$, $\beta = 1.5$, $\lambda = 0.01$								
$n \downarrow$		AE(α_1) (MSE)	AE(α_2) (MSE)	AE(α_3) (MSE)	AE(β) (MSE)	AE(λ) (MSE)	Av.LL	time(secs)
25	Ini.guess	0.9	0.9	0.9	0.9	0.9	–	–
	EM	0.0486	0.0391	0.0578	1.4915	0.0353	-104.1342	0.0143
		(0.0005)	(0.0004)	(0.0007)	(0.1549)	(0.0077)		
	nlm	0.0708	0.0576	0.0828	1.3798	0.0448	-106.1228	0.0217
		(0.0135)	(0.0124)	(0.0133)	(0.1739)	(0.0077)		
	optim	0.8331	0.7082	1.1033	0.1378	0.0723	-169.9854	0.0161
		(0.7433)	(0.5579)	(1.2729)	(1.8688)	(0.0053)		
50	Ini.guess	0.9	0.9	0.9	0.9	0.9	–	–
	EM	0.0500	0.0395	0.0598	1.4866	0.0242	-210.2612	0.0188
		(0.0003)	(0.0002)	(0.0004)	(0.0777)	(0.0033)		
	nlm	0.0839	0.0707	0.1006	1.3608	0.0331	-216.2919	0.0288
		(0.0235)	(0.0230)	(0.0230)	(0.1289)	(0.0036)		
	optim	0.8991	0.7408	1.1499	0.1393	0.0656	-346.8760	0.0199
		(0.8037)	(0.5581)	(1.3656)	(1.8584)	(0.0038)		
100	Ini.guess	0.6	0.6	0.6	1	1	–	–
	EM	0.0498	0.0398	0.0595	1.5014	0.0159	-423.2508	0.0297
		(0.0001)	(0.0001)	(0.0002)	(0.0421)	(0.0016)		
	nlm	0.0918	0.0790	0.0996	1.2815	0.0415	-437.6668	0.0394
		(0.0184)	(0.0174)	(0.0155)	(0.1934)	(0.0050)		
	optim	0.5643	0.4725	0.9246	0.2211	0.0731	-616.4281	0.0263
		(0.3036)	(0.2276)	(0.8627)	(1.6622)	(0.0062)		

model fit the data appropriately? We are going to address all such issues.

Before analyzing the data set using the BMWD, we first fit the MWDs 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. For this purpose we consider some model selection criteria, e.g., the WPP plot to check whether the MWD is an appropriate fit or not to X_1, X_2 and $\min(X_1, X_2)$ separately. The WPP plots, see Figure 3, of X_1, X_2 and $\min(X_1, X_2)$ for the considered data set are approximated by the convex curve for both marginals and for their minimum also. Thus from the WPP plots, we can fit MWDs to the marginals and minimum of the data set.

The MLEs of parameters of the MWD for each case obtained by using three dimensional standard numerical optimization techniques, e.g., optim and nlm. To start the numerical technique, we need some initial guesses of the unknown parameters. These initial guesses are obtained by the WPP plot, for the MWD that can be represented as the linear equation; $y = \log(\alpha) + \beta t_1 + \lambda t_2$,

Figure 2: Scatter plot of UEFA Data

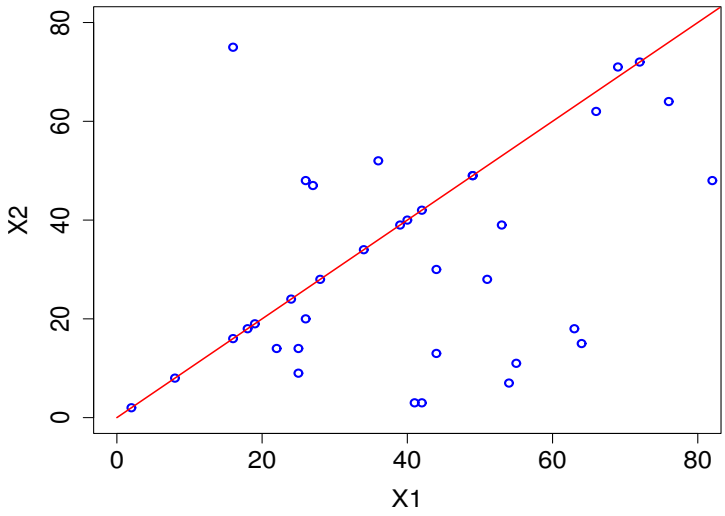
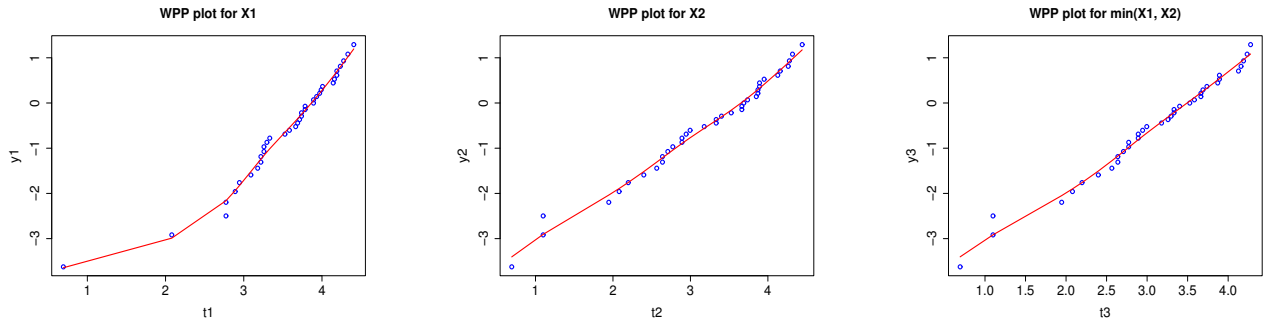


Figure 3: WPP Plots of marginals and min (X_1, X_2)



where $t_1 = t$, $t_2 = \exp(t)$ and $t = \log(x)$, $y_i = \log[-\log(1 - i/(n+1))]$. Note that t_1 and t_2 are not statistically independent. For the marginal $X_1 \sim MW(\alpha, \beta, \lambda)$, using the regression procedure, the estimated model is $y_1 = -4.5235 + 0.7700t_1 + 0.0308\exp(t_1)$, where $t_1 = \log(x_1)$ with coefficient of determination, $R^2 = 0.976$ and hence the initial guess values of (α, β, λ) are (0.0109, 0.7700, 0.0308). Similarly for the marginal X_2 and $Z = \min(X_1, X_2)$ cases, the estimated respective models are $y_2 = -4.0959 + 1.0488t_2 + 0.0072\exp(t_2)$, where $t_2 = \log(x_2)$ with $R^2 = 0.987$ and $y_3 = -4.2442 + 1.1397t_3 + 0.0069\exp(t_3)$, where $t_3 = \log(z)$ with $R^2 = 0.987$. Hence the corresponding initial guess values of (α, β, λ) for X_2 and $\min(X_1, X_2)$ are (0.016, 1.0488, 0.0072) and (0.0143, 1.1397, 0.0069) respectively. Using these initial guess values, we fit the MWD to X_1, X_2 and $\min(X_1, X_2)$ separately and for each case the MLEs and corresponding

Figure 4: KS distance plots of marginals and minimum

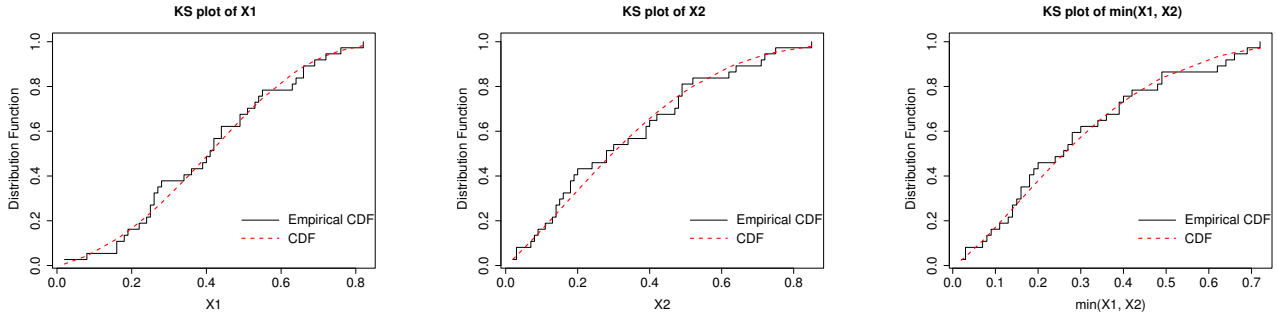
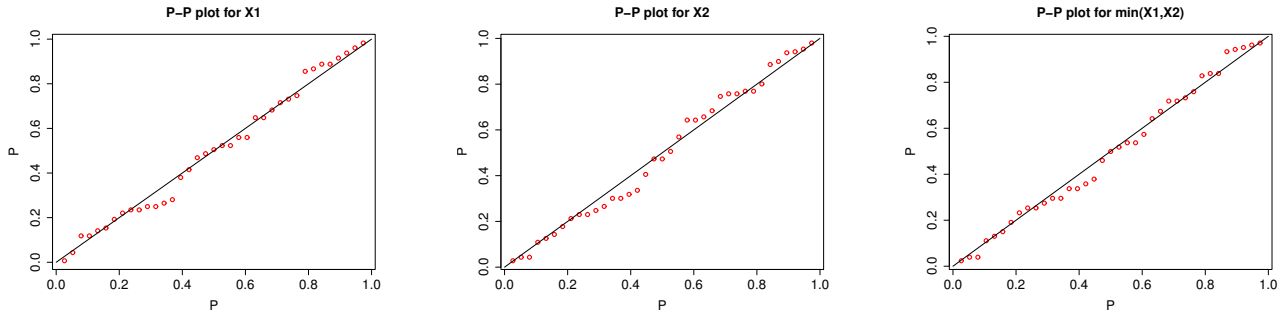


Figure 5: P-P plots of marginals and minimum



standard error (SE) and log-likelihood (LL) values with KS distance and associated p-values are given in Table 5.

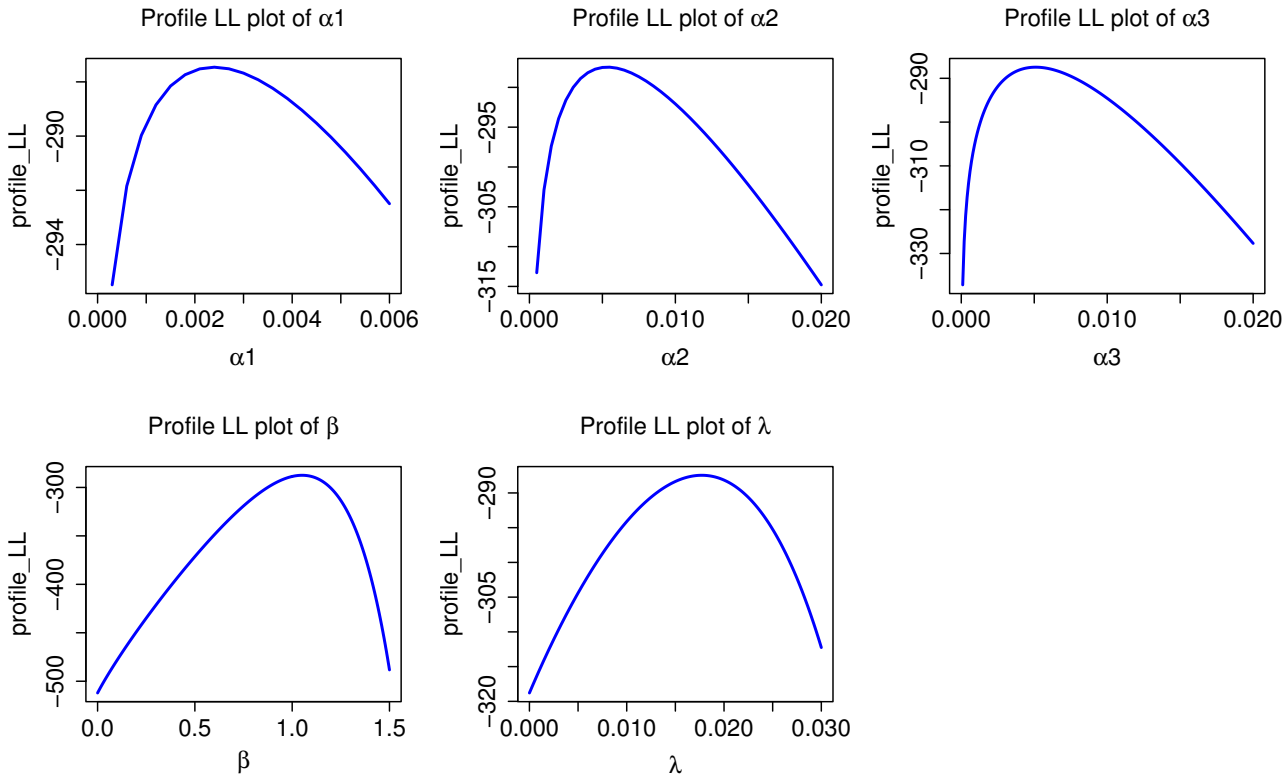
Table 5: MLEs, SE, LL, KS distance with p-value of MW for X_1, X_2 and $\min(X_1, X_2)$

Model	MLEs	(Std Error)	LL	KS	p-value
X_1	$\alpha = 0.002739$	0.000681	-161.9119	0.0979	0.8698
	$\beta = 1.256448$	0.170828			
	$\lambda = 0.021497$	0.009539			
X_2	$\alpha = 0.013188$	0.010595	-162.6188	0.0965	0.8811
	$\beta = 1.074898$	0.297391			
	$\lambda = 0.010765$	0.009107			
$\min(X_1, X_2)$	$\alpha = 0.010086$	0.007634	-157.8155	0.0805	0.9702
	$\beta = 1.232323$	0.296809			
	$\lambda = 0.008166$	0.010246			

Based on the p-values, KS distance plots, see Figure 4, and the percentile-percentile (P-P) plots, see Figure 5, the MWD cannot be rejected to fit the marginals and minimum of the data set also.

Now we will fit the BMWD for the data. The initial guess value for β is the mean of (1.2564, 1.0749, 1.2323), which is 1.1879 and for λ, it is the mean of (0.0215, 0.0108, 0.0082), which is 0.0135. To find the initial guess values for the parameters $(\alpha_1, \alpha_2, \alpha_3)$ we again fit the $MW(\alpha_i + \alpha_3, 1.1879, 0.0135)$ to X_i, $i = 1, 2$ and the $MW(\alpha_1 + \alpha_2 + \alpha_3, 1.1879, 0.0135)$ to $\min(X_1, X_2)$ separately. The respective MLEs are obtained as $\alpha_1+\alpha_3 = 0.0057$, $\alpha_2+\alpha_3 = 0.0073$ and $\alpha_1 + \alpha_2 + \alpha_3 = 0.0092$. By solving these three linear equations the initial guess values of (α_1, α_2, α_3), are obtained as (0.0019, 0.0035, 0.0038). Using these initial guess values, viz., (0.0019, 0.0035, 0.0038, 1.1879, 0.0135) 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 log-likelihood function at the k^{th} iteration. In each iteration we need to estimate β and λ by using the 2-D optimization technique, eg., optim and nlm. The iteration stops after 6 cycles in 2.23 seconds and we get the MLEs of $\alpha_1, \alpha_2, \alpha_3, \beta$ and λ and the corresponding SE, 95% and 99% CIs which are given in Table 6. The corresponding LL and pseudo log-likelihood (pLL) values are LL = -287.4584 and pLL = -302.2672, respectively. For some parameter estimates, we get lower limit of asymptotic CI as negative. In such cases, we take the lower limit as 0 since the parameters take only positive real values. We have also tried for some other initial guesses, e.g., (5, 5, 5, 5, 5, 5), (1.0, 1.0, 1.0, 1.0, 1.0), (0.01, 0.01, 0.01, 0.01, 0.01) and (0.5, 0.5, 0.5, 0.5, 0.5) for $(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$. The iteration stops after k= 10th. The EM algorithm converges to the

Figure 6: Profile log likelihood plots for UEFA Data



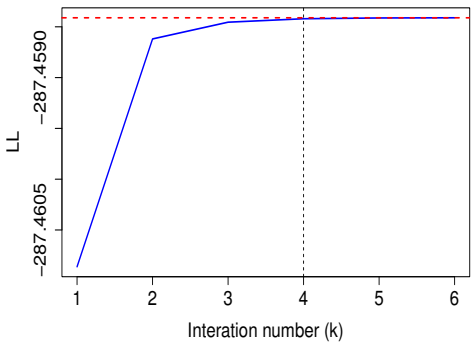
same point when we use the the same stopping criterion. The profile log-likelihood function of $\alpha_1, \alpha_2, \alpha_3, \beta$ and λ of the BMWD for soccer data are plotted, see Figures 6 separately. These plots of the profile log-likelihood functions show that there is unique MLEs.

Table 6: MLEs, SE, 95%, 99% CIs for the BMWD

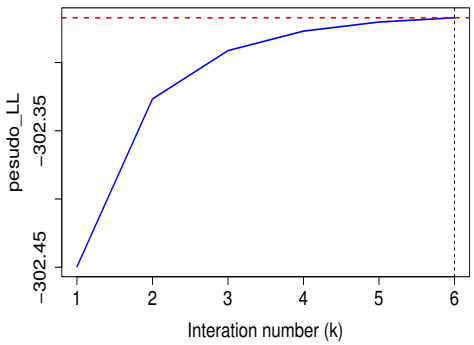
Par	MLE	SE	95% CI	99% CI
α_1	0.002393	0.002422	(0.000000 , 0.007141)	(0.000000 , 0.008633)
α_2	0.005344	0.005123	(0.000000 , 0.015385)	(0.000000 , 0.018540)
α_3	0.005097	0.004929	(0.000000 , 0.014757)	(0.000000 , 0.017793)
β	1.051223	0.334752	(0.395121 , 1.707325)	(0.188958 , 1.913487)
λ	0.017715	0.008760	(0.000547 , 0.034885)	(0.000000 , 0.040279)

It is interesting to see from Figure 7a that at each iteration the log-likelihood function is gradually increasing. From Figure 7a it is also clear that the log-likelihood value almost stabilizes after 4-th iteration. The pseudo log-likelihood function, see Figure 7b, increases slower than the

Figure 7: Log-likelihood and profile log-likelihood plots for the soccer data



(a) Log-likelihood plot



(b) Profile log-likelihood plot

log-likelihood function. Now the natural question arises whether the BMWD fits the data well or not. For this purpose we use the bivariate GoF. The KS distance is obtained using the real data set. The corresponding critical values and p-values for different model are obtained by parametric bootstrap method. In bootstrap we generate 10^5 samples each of size $n = 37$ from $BMW(\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}, \hat{\beta}, \hat{\lambda})$, where $(\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}, \hat{\beta}, \hat{\lambda})$ are MLEs based on real data set. In each sample we obtained the KS test statistic D_n and hence we get the sampling distribution of D_n, see [20]. By using this sampling distribution, the critical value is obtained by taking the q^{th} quantile. The critical and p-value are given in Table 7. Based on the p-value the BMWD cannot be rejected for the data.

Table 7: GoF test to UEFA Data

Model	KS	p-value	Critical
MOBE	0.2660	0.0006	0.1755
MOBW	0.0726	0.7943	0.3590
BMW	0.0742	0.5720	0.1338

Table 8: LL, AIC, BIC, CAIC AND HQIC values for different models

Model	LL	AIC	BIC	CAIC	HQIC
MOBE	-299.0672	604.1344	608.9672	606.5508	605.8382
MOBW	-289.4283	586.8566	593.3003	590.0784	589.1283
BMW	-287.4584	584.9168	592.9714	588.9441	587.7564

For the comparison of different models, as to which one given the best fit among the three, we

consider different model selection and relative GoF measure criteria, e.g., the Akaike information criterion (AIC), $AIC = -2LL + 2p$, where p is the number of parameters to be estimated, Bayesian information criteria $BIC = -2LL + p\ln(n)$, consistent AIC, $CAIC = w * AIC + (1 - w) * BIC$, where w is the weight and $0 < w < 1$ with default value is w = 0.5 and Hannan–Quinn information criterion $(HQIC) = -2LL + 2p\ln(\ln(n))$. The model with lower AIC/BIC/CAIC/HOIC is better than other models fitted for the same data set. From Table 8 we observed that the BMWD is better fits the data than MOBE and MOBW models. Now we develop some following tests;

(1) We test whether the marginal distributions (times at which the first goal scored by the home team and time at which the fist kick goal scored by any team) are same or not, i.e., $H_0 : \alpha_1 = \alpha_2 \; versus \; H_a : \alpha_1 \neq \alpha_2$. Then $\chi^2_{cal} = 3.8542$ with p-value = 0.0496 and we reject the null hypothesis at 5% level of significance. Thus the marginals do not have the same distribution.

(2) We test whether the addition of parameter λ is significant for the data or not, i.e., $H_0 : \lambda = 0 \; (MOBW) \; versus \; H_a : \lambda > 0 \; (BMWD)$. Here $\chi^2_{cal} = 3.9398$ with p-value = 0.0236 and we reject the null hypothesis at 5% level of significance. Thus λ is statistically significant and hence the BMWD is appropriate fit the soccer data.

(3) We test the hypothesis $H_0 : \lambda = 0, \beta = 1 \; (MOBE) \; versus \; H_a : \lambda > 0, \beta \neq 1 \; (BMW)$. Then $\chi^2_{cal} = 23.2176$ with p-value = 5.2662e-06 and we reject the null hypothesis at 5% level of significance. Hence the BMWD is appropriate fit the data.

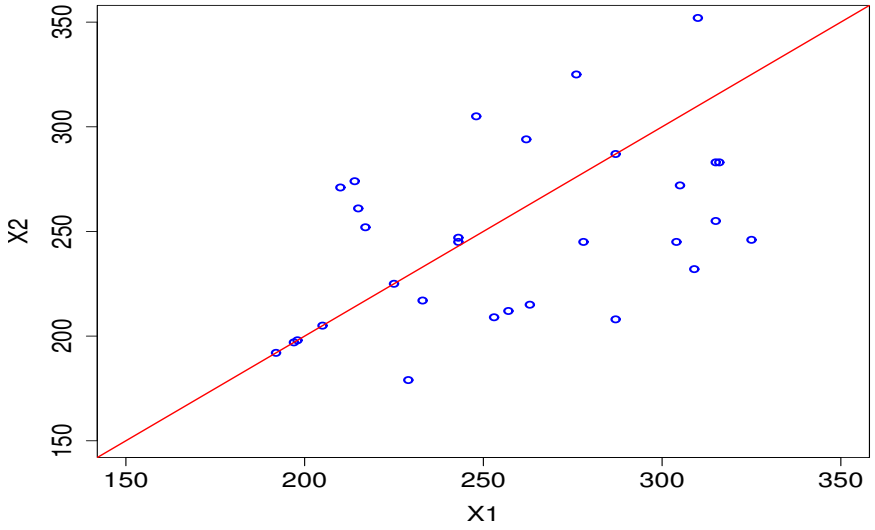
(4) We test whether bivariate Rayleigh model is appropriate fit to the data or not, i.e., $H_0 : \lambda = 0, \beta = 2 \; (BRD) \; versus \; H_a : \lambda > 0, \beta \neq 2 \; (BMW)$. Then $\chi^2_{cal} = 6.4807$ with p-value = 0.0025 and we reject H_0 at 5% level of significance. Hence the BMWD is appropriate fit the data.

(5) We test the hypothesis $H_0 : \beta = 0 \; versus \; H_a : \beta > 0 \; (BMW)$. The $\chi^2_{cal} = 22.2669$ with p-value = 1.1863e-06 and we reject H_0 at 5% level of significance. Hence the BMWD is appropriate fit.

6.2 Second data: Cholesterol Level Data

This data set contains cholesterol levels at 5 and 25 weeks after treatment of 30 patients. This data set was used by Shoaee [18] and is represented in Table 9. Let the cholesterol level at

Figure 8: Scatter plot of Cholesterol level Data



5^{th} week be denoted by Y_1 and at 25^{th} week by Y_2. Before analyzing the data, we apply the transformation $X_1 = (Y_1 - 178.5)/100$ and $X_2 = (Y_2 - 178.5)/100$ to all data points. The scatter plot of the data is given in Figure 8.

Table 9: Cholesterol levels at 5 and 25 weeks after treatment in 30 patients.

S.N.	5-th	25-th	S.N.	5-th	25-th	S.N.	5-th	25-th
1	325	246	2	278	245	3	257	212
4	192	192	5	276	325	6	262	294
7	309	232	8	287	287	9	304	245
10	215	261	11	217	252	12	248	305
13	225	225	14	287	208	15	233	217
16	198	198	17	229	179	18	310	352
19	214	274	20	253	209	21	316	283
22	243	245	23	305	272	24	197	197
25	243	247	26	315	283	27	205	205
28	315	255	29	263	215	30	210	271

We analyze the data set using the BMWD to study the effect of treatment in cholesterol levels of patients. We fit the model to the data by using the iterative procedure of the EM algorithm and compare these results with those results obtained by direct optimization methods optim (Nelder-Mead Method) and nlm available in R. In this data set we have all three possibilities,

e.g., $X_1 < X_2$ with $n_1 = 10$ patients, $X_1 > X_2$ with $n_2 = 14$ patients and $X_1 = X_2 = X$ (say) with $n_3 = 6$ patients.

Before analyzing the data set using the BMWD, we first fit the MWD 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. For this purpose we consider some model selection criteria, e.g., the total time test (TTT) plot and the Weibull probability paper (WPP) plot to check whether the MWD is an appropriate fit or not. From scaled TTT plots, see Figure 9, and WWP plots, see Figure 10, of X_1, X_2 and $\min(X_1, X_2)$ it is quite clear that the MWD model is appropriate fit to both the variables X_1 and X_2, and their $\min\{X_1, X_2\}$.

Figure 9: TTT Plots of marginals and min (X_1, X_2)

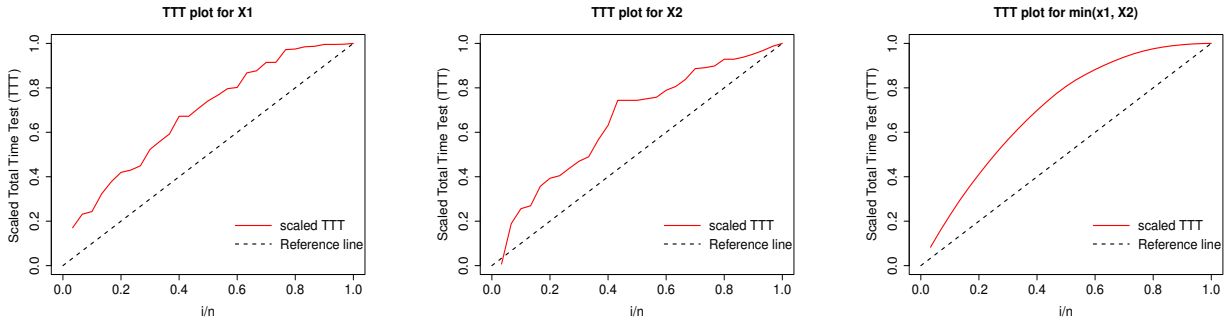
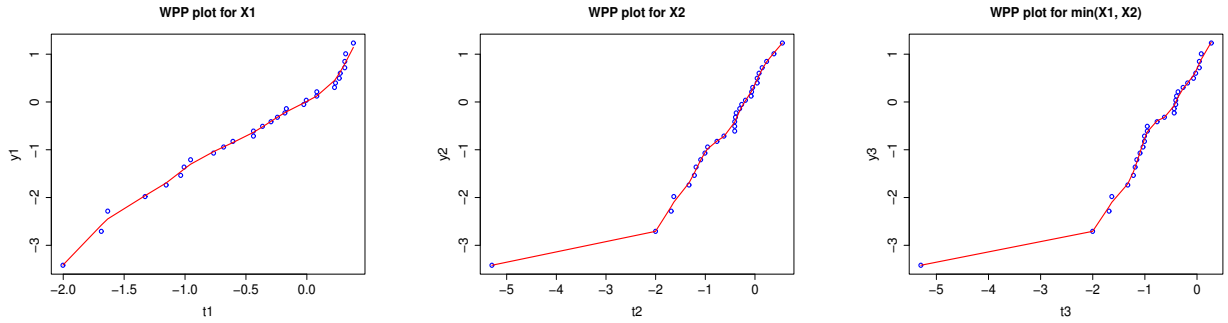


Figure 10: WPP Plots of marginals and min (X_1, X_2)



The MLEs of parameters of the MWD for each case are obtained by using 3-D standard numerical optimization techniques, e.g., optim and nlm (packages available in R). To start the numerical technique, we need some initial guesses of the unknown parameters. The initial guesses are obtained from the WPP plot, for the MWD that can be represented as a linear equation, like, $y = \log(\alpha) + \beta t_1 + \lambda t_2$, where $t_1 = t$, $t_2 = \exp(t)$ and $t = \log(x)$, $y_i = \log[-\log(1 - i/(n+1))]$.

Note that t_1 and t_2 are not statistically independent. For the marginal $X_1 \sim MW(\alpha, \beta, \lambda)$, using the regression procedure, the estimated model is $y_1 = -0.1001 + 1.5097t_1 + 0.2455\exp(t_1)$, where $t_1 = \log(x_1)$ with coefficient of determination, $R^2 = 0.9802$ and hence the initial guess values of (α, β, λ) are (0.9048, 0.1.5097, 0.2455). Similarly for the marginal X_2 and $Z = \min(X_1, X_2)$ cases, the estimated models are $y_2 = -1.3818 + 0.4442t_2 - 1.6559\exp(t_2)$, where $t_2 = \log(x_2)$ with $R^2 = 0.9609$ and $y_3 = -1.3702 + 0.4455t_3 + 2.1554\exp(t_3)$, where $t_3 = \log(z)$ with $R^2 = 0.9405$, respectively. Hence the corresponding initial guess values of (α, β, λ) for X_2 and $\min(X_1, X_2)$ are (0.2511, 0.4442, 1.6559) and (0.2540, 0.4455, 2.1554) respectively. Using these initial guess values, we fit the MWD to X_1, X_2 and $\min(X_1, X_2)$ separately and for each case the MLEs and corresponding standard error (SE) and log-likelihood (LL) values with KS distance and associated p-values are given in Table 10. Based on the p-values, KS distance plots, see Figure

Table 10: MLEs, SE, LL, KS distance with p-value of MW for X_1, X_2 and $\min(X_1, X_2)$

Model	MLEs	(Std Error)	LL	KS	p-value
X_1	$\alpha = 0.4209$	0.3806	-14.2876	0.1285	0.7052
	$\beta = 1.2450$	0.6265			
	$\lambda = 1.0081$	0.8131			
X_2	$\alpha = 0.5115$	0.3411	-13.6642	0.1026	0.9105
	$\beta = 0.9019$	0.4167			
	$\lambda = 1.0216$	0.5882			
$\min(X_1, X_2)$	$\alpha = 0.6048$	0.5067	-6.7354	0.1350	0.6450
	$\beta = 0.9576$	0.4739			
	$\lambda = 1.3194$	0.8346			

11, and the percentile-percentile (P-P) plots, see Figure 12, the MWD cannot be rejected to fit the marginals and minimum of the data set also.

Figure 11: KS distance plots of marginals and minimum

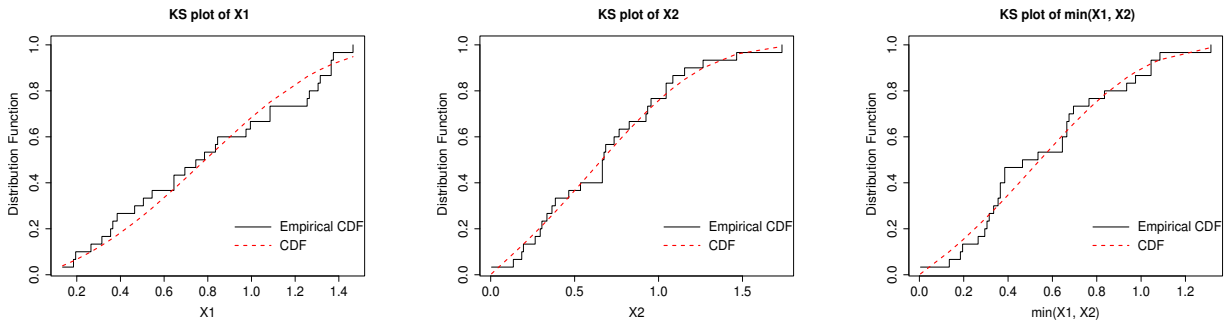
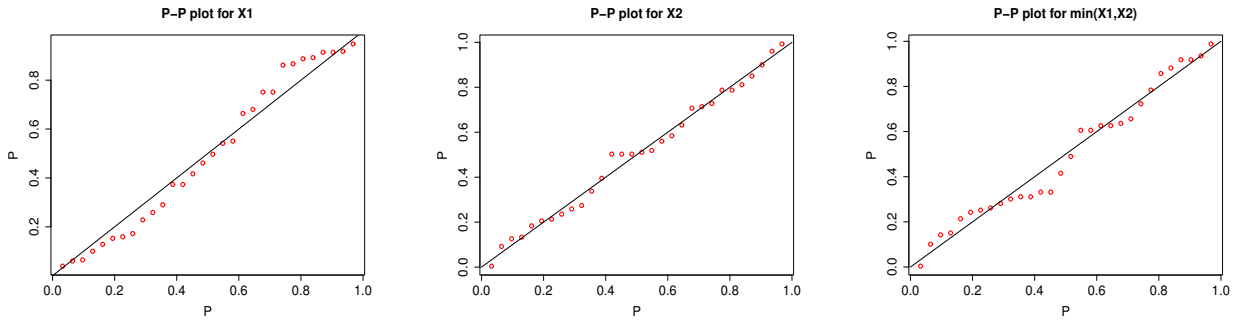


Figure 12: P-P plots of marginals and minimum



Now we will fit the BMWD for the data. The initial guess value for β is the mean of (0.1.2450, 0.9019, 0.9576), which is 1.0348 and for λ, it is the mean of (1.0081, 1.0216, 1.3194), which is 1.1164. The respective MLEs are obtained as $\alpha_1 + \alpha_3 = 0.4209$, $\alpha_2 + \alpha_3 = 0.5115$ and $\alpha_1 + \alpha_2 + \alpha_3 = 0.6048$. By solving these three linear equations the initial guess values of (α_1, α_2, α_3), are obtained as (0.0934, 0.1839, 0.3275). Using all five initial guess values, viz., (0.0934, 0.1839, 0.3275, 1.0348, 1.1164) 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 log-likelihood function at the k^{th} iteration. In each iteration we need to estimate β and λ by using the 2-D optimization technique, eg., optim and nlm. The iteration stops after 14 cycles in 0.28 seconds and we get the MLEs of $\alpha_1, \alpha_2, \alpha_3, \beta$ and λ and the corresponding SE, 95% and 99% CIs which are given in Table 11. The corresponding LL and pseudo log-likelihood (pLL) values are LL = -38.2120 and pLL = -53.4373, respectively. For some parameter estimates, we get lower limit of asymptotic CIs as negative. In such cases, we take the lower limit as 0. We have also tried for some other initial guesses, e.g., (5, 5, 5, 5, 5, 5), (1.0, 1.0, 1.0, 1.0, 1.0), (0.01, 0.01, 0.01, 0.01, 0.01) and (0.5, 0.5, 0.5, 0.5, 0.5) for $(\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$. The iteration stops after k = 13th. The EM algorithm converges to the same point when we use the the same stopping criterion irrespective of the initial guess values. The profile log-likelihood function of $\alpha_1, \alpha_2, \alpha_3, \beta$ and λ of the BMWD for cholesterol data are plotted, see Figure 13 separately. These plots indicate the MLEs are unique.

It is interesting to see from Figure 14a that at each iteration the log-likelihood function is gradually increasing. From Figure 14a it is also clear that the log-likelihood value almost stabilizes after 5-th iteration. The pseudo log-likelihood function, see Figure 14b, increases slower than the log-likelihood function.

Now the natural question arises whether the BMWD fits the data well or not. For this pur-

Figure 13: Profile log likelihood plots for Cholesterol level Data

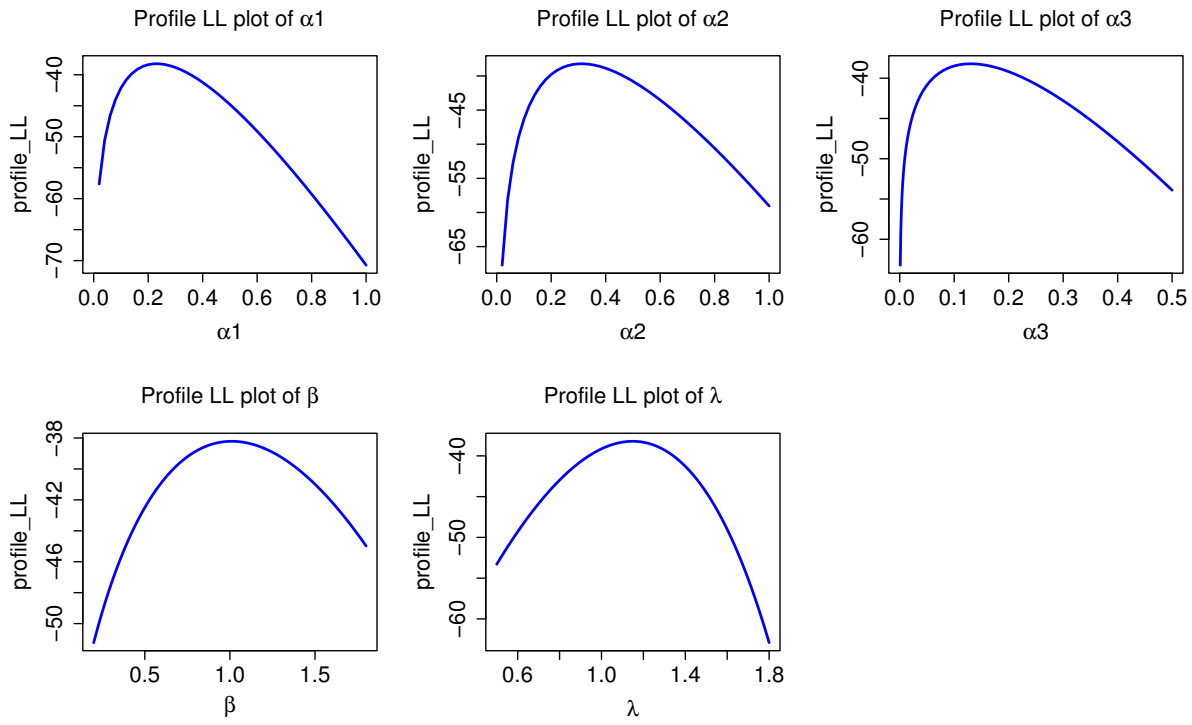
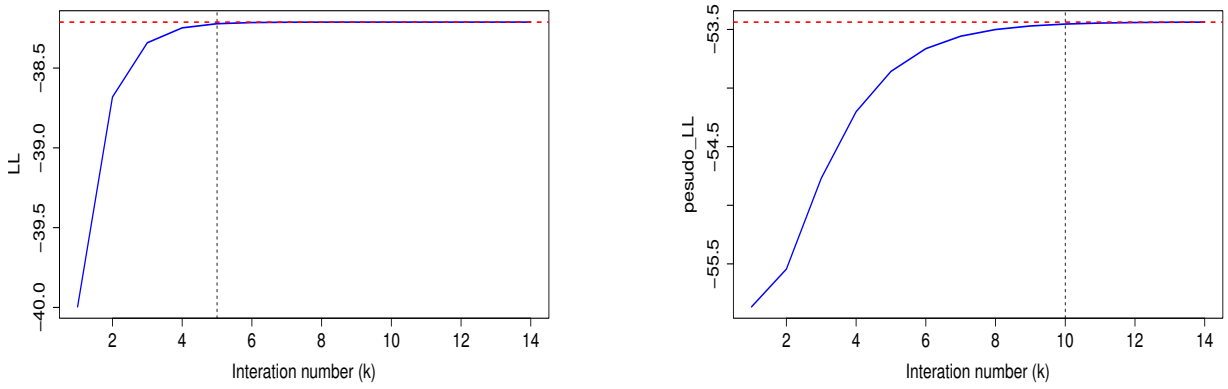


Figure 14: Log-likelihood and profile log-likelihood plots for Cholesterol level Data



(a) Log-likelihood plot

(b) Profile log-likelihood plot

Table 11: MLEs, SE, 95%, 99% CIs for the BMWD

Par	MLE	SE	95% CI	99% CI
α_1	0.2307	0.1452	(0.0000, 0.5152)	(0.0000, 0.6046)
α_2	0.3113	0.1953	(0.0000, 0.6940)	(0.0000, 0.8143)
α_3	0.1304	0.0927	(0.0000, 0.3120)	(0.000, 0.3691)
β	1.0110	0.4063	(0.2146, 1.8074)	(0.0000, 2.0577)
λ	1.1490	0.5314	(0.1076, 2.1905)	(0.0000, 2.5177)

pose we use the univariate GoF to marginals and the bivariate GoF to the BMW model. For the marginals the KS distances and the corresponding p-values (reported in brackets) between X_1, X_2 and $\min\{X_1, X_2\}$ with MW(0.3611, 1.0110, 1.1490), MW(0.4417, 1.0110, 1.1490) and MW(0.6724, 1.0110, 1.1490) are 0.1203 (0.7784), 0.1029 (0.9084) and 0.1379 (0.6186) respectively. It indicates that the MWD can used to analyzed X_1, X_2 and $\min\{X_1, X_2\}$. It does not guarantee that (X_1, X_2) will have BMWD, but gives an indication that BMWD may be used to analyze this biavariate data set. We consider a bivariate GoF for this purpose.

The KS distance is obtained using the real data set. The corresponding critical values and p-values for different model are obtained by parametric bootstrap method. In bootstrap we generate 10^5 samples each of size n = 32 from $BMW(\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}, \hat{\beta}, \hat{\lambda})$, where $(\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}, \hat{\beta}, \hat{\lambda})$ are MLEs based on real data set. In each sample we obtained the KS test statistic D_n and hence we get the sampling distribution of D_n, see [20]. By using this sampling distribution, the critical value is obtained by taking the q^{th} quantile. The critical and p-values are given in Table 12. Based on the p-value the BMWD cannot be rejected for the data.

Table 12: Bivariate GoF test to Cholesterol Level Data

Model	KS	p-value	Critical
MOBE	0.3149	0.0002	0.1968
MOBW	0.0926	0.5493	0.1721
BMW	0.0639	0.7748	0.1437

Table 13: LL, AIC, BIC, CAIC AND HQIC values for different models

Model	LL	AIC	BIC	CAIC	HQIC
MOBE	-52.1817	110.3634	114.5670	112.4652	111.7082
MOBW	-40.3992	88.7984	94.4032	91.6008	90.5914
BMW	-38.2120	86.4240	93.4300	89.9270	88.6653

For the comparison of different models, as to which one gives the best fit among the three, we

consider different model selection and relative GoF measure criteria, like, AIC, BIC, CAIC and HQIC. From Table 13 we see that the BMWD better fits the data than MOBE and MOBW models. Now we develop the following tests;

(1) We test whether the marginal distributions are same or not, i.e., $H_0 : \alpha_1 = \alpha_2 \; versus \; H_a : \alpha_1 \neq \alpha_2$. Then $\chi^2_{cal} = 0.7729$ with p-value $= 0.3793$ and we my accept the null hypothesis at 5% level of significance. Thus the marginals have the same distribution.

(2) We test whether the addition of parameter λ is significant for the data or not, i.e., $H_0 : \lambda = 0 \; (MOBW) \; versus \; H_a : \lambda > 0 \; (BMWD)$. Here $\chi^2_{cal} = 4.3744$ with p-value $= 0.0182$ and we reject the null hypothesis at 5% level of significance. Thus λ is statistically significant and hence the BMWD is appropriate fit the motorcycle accident insurance data.

(3) We test the hypothesis, $H_0 : \lambda = 0, \beta = 1 \; (MOBE) \; versus \; H_a : \lambda > 0, \beta \neq 1 \; (BMW)$. Then $\chi^2_{cal} = 27.9394$ with p-value $= 0.4.9114\text{e-}07$ and we reject the null hypothesis at 5% level of significance. Hence the BMWD is appropriate fit the data.

(4) We test the hypothesis $H_0 : \beta = 0 \; versus \; H_a : \beta > 0 \; (BMW)$. The $\chi^2_{cal} = 14.8859$ with p-value $= 5.7106\text{e-}05$ and we reject H_0 at 5% level of significance. Hence the BMWD is appropriate fit.

7 Conclusion

In this paper, we study the BMWD and discuss the applicability and efficiency of the EM algorithm for estimating the maximum likelihood estimators of the five unknown parameters of the BMWD. It has been seen that using the EM algorithm in practice is not very difficult and is actually only a 2-D optimization problem. In essence, the EM algorithm turns a 5-D optimization problem to a 2-D problem. The results of the simulations show that the EM algorithm works quite well. We have also obtained the asymptotic CIs of the estimated parameters using the idea of Louis [9] and the asymptotic normality property of the MLEs. We have seen that the asymptotic results can be used effectively to obtain CIs even when sample size is small. The LR tests are constructed at the boundary point of the parameter space, whenever required. We have analyzed one real data set, which was initially studied by Meintains [12] using MOBE distribution, and Kundu and Dey [7] using MOBW distribution. We have constructed bivariate GoF test using SEB [1] and observed that the BMWD fits appropriately to this data set. Using specific criteria for choosing a model, such as AIC, BIC, CAIC etc., we see that the BMWD fits

the data better than the MOBW and MOBE models. irrespective of the initial guesses of the parameters α_1, α_2 and α_3, the proposed EM algorithm converges to the true parameter unlike optim and nlm algorithms. The usual standard numerical optimization methods, e.g., optim and nlm cannot be used when the initial guesses are far from the true parameter values.

References

[1] Alam, K., Abernathy, R., and Williams, C. L. (1993). Multivariate goodness-of-fit tests based on statistically equivalent blocks. *Communications in Statistics-Theory and Methods*, 22(6):1515–1533.

[2] Alam, K. and Williams, C. L. (1995). A multivariate goodness-of-fit test for stochastically ordered distributions. *Biometrical journal*, 37(8):945–956.

[3] Anderson, T. W. (1966). Some nonparametric multivariate procedures based on statistically equivalent blocks. *Proccedings of the International Symposium on Multivariate Analysis, Academic Press, New York*, pages 5–27.

[4] Bemis, B. M., 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(340):927–929.

[5] El-Bassiouny, A., Shahen, H., and Abouhawwash, M. (2018). A new bivariate modified Weibull distribution and its extended distribution. *Journal of Statistics Applications & Probability*, 7(2):217–231.

[6] Hallin, M., Mordant, G., and Segers, J. (2021). Multivariate goodness-of-fit tests based on wasserstein distance.

[7] Kundu, D. and Dey, A. K. (2009). Estimating the parameters of the Marshall–Olkin bivariate Weibull distribution by EM algorithm. *Computational Statistics & Data Analysis*, 53(4):956–965.

[8] Lai, C., Xie, M., and Murthy, D. (2003). A modified Weibull distribution. *IEEE Transactions on Reliability*, 52(1):33–37.

[9] Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):226–233.

[10] Lu, J.-C. (1989). Weibull extensions of the Freund and Marshall-Olkin bivariate exponential models. *IEEE Transactions on Reliability*, 38(5):615–619.

[11] Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44.

[12] Meintanis, S. G. (2007). Test of fit for Marshall–Olkin distributions with applications. *Journal of Statistical Planning and Inference*, 137(12):3954–3963.

[13] Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.

[14] Pyke, R. (1965). Spacings. *Journal of the Royal Statistical Society: Series B (Methodological)*, 27(3):395–436.

[15] Sarhan, A. M., Apaloo, J., and Kundu, D. (2022). A new bivariate lifetime distribution: properties, estimations and its extension. *Communications in Statistics-Simulation and Computation*, pages 1–18.

[16] Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610.

[17] Shahen, H., El-Bassiouny, A., and Abouhawwash, M. (2019). Bivariate exponentiated modified Weibull distribution. *Journal of Statistics Applications and Probability*, 8(1):27–39.

[18] Shoaee, S. (2020). On a new class of bivariate survival distributions based on the model of dependent lives and its generalization. *Applications and Applied Mathematics: An International Journal (AAM)*, 15(2):6.

[19] Singh, R., Dutta, S., and Misra, N. (2022). Some multivariate goodness of fit tests based on data depth. *Journal of Nonparametric Statistics*, 34(2):428–447.

[20] Stute, W., Manteiga, W. G., and Quindimil, M. P. (1993). Bootstrap based goodness-of-fit-tests. *Metrika*, 40:243–256.

[21] Zhang, C., Xiang, Y., and Shen, X. (2012). Some multivariate goodness-of-fit tests based on data depth. *Journal of Applied Statistics*, 39(2):385–397.

Appendix I: Observed Fisher Information matrix

To compute the FIM, we use the same notation as of Louis [9], where the matrix $H = ((H_{ij}))$ denotes the Hessian matrix and the vector $U = (U_i)$ denotes the gradient (or score) vector of the pseudo log-likelihood function. Let $\hat{\theta} = (\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}, \hat{\beta}, \hat{\lambda})$ be MLEs, then the observed FIM is $\mathcal{F}_{\hat{\theta}} = I_p - UU^T$, where $I_p = -H$. The elements of the gradient vector $U = (U_1, U_2, U_3, U_4, U_5)^T$ are

$$
\begin{aligned}
U_1 =& \frac{\partial l_p}{\partial \alpha_1} = \frac{n_1}{\hat{\alpha}_1} + \frac{n_2}{\hat{\alpha}_1 + \hat{\alpha}_3} - \Bigg[\sum_{i \epsilon I_1 \cup I_2} x_{1i}^{\hat{\beta}} e^{\hat{\lambda} x_{1i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \Bigg], \\
U_2 =& \frac{\partial l_p}{\partial \alpha_2} = \frac{n_1}{\hat{\alpha}_2 + \hat{\alpha}_3} + \frac{n_2}{\hat{\alpha}_2} - \Bigg[\sum_{i \epsilon I_1 \cup I_2} x_{2i}^{\hat{\beta}} e^{\hat{\lambda} x_{2i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \Bigg], \\
U_3 =& \frac{\partial l_p}{\partial \alpha_3} = \frac{n_1}{\hat{\alpha}_2 + \hat{\alpha}_3} + \frac{n_2}{\hat{\alpha}_1 + \hat{\alpha}_3} + \frac{n_3}{\hat{\alpha}_3} - \Bigg[\sum_{i \epsilon I_1} x_{2i}^{\hat{\beta}} e^{\hat{\lambda} x_{2i}} + \sum_{i \epsilon I_2} x_{1i}^{\hat{\beta}} e^{\hat{\lambda} x_{1i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \Bigg],
\end{aligned}
$$

$$
\begin{aligned}
U_4 =& \frac{\partial l_p}{\partial \beta} = \Bigg[\sum_{i \epsilon I_1 \cup I_2} \ln x_{1i} + \sum_{i \epsilon I_1 \cup I_2} \ln x_{2i} + \sum_{i \epsilon I_3} \ln x_i \Bigg] \\
&+ \Bigg[\sum_{i \epsilon I_1 \cup I_2} (\hat{\beta} + \hat{\lambda} x_{1i})^{-1} + \sum_{i \epsilon I_1 \cup I_2} (\hat{\beta} + \hat{\lambda} x_{2i})^{-1} + \sum_{i \epsilon I_3} (\hat{\beta} + \hat{\lambda} x_i)^{-1} \Bigg] \\
&- \hat{\alpha}_1 \Bigg[\sum_{i \epsilon I_1 \cup I_2} x_{1i}^{\hat{\beta}} e^{\hat{\lambda} x_{1i}} \ln x_{1i} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \ln x_i \Bigg] - \hat{\alpha}_2 \Bigg[\sum_{i \epsilon I_1 \cup I_2} x_{2i}^{\hat{\beta}} e^{\hat{\lambda} x_{2i}} \ln x_{2i} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \ln x_i \Bigg] \\
&- \hat{\alpha_3} \Bigg[\sum_{i \epsilon I_1} x_{2i}^{\hat{\beta}} e^{\hat{\lambda} x_{2i}} \ln x_{2i} + \sum_{i \epsilon I_2} x_{1i}^{\hat{\beta}} e^{\hat{\lambda} x_{1i}} \ln x_{1i} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \ln x_i \Bigg], \\
U_5 =& \frac{\partial l_p}{\partial \lambda} = \Bigg[\sum_{i \epsilon I_1 \cup I_2} x_{1i} (\hat{\beta} + \hat{\lambda} x_{1i})^{-1} + \sum_{i \epsilon I_1 \cup I_2} x_{2i} (\hat{\beta} + \hat{\lambda} x_{2i})^{-1} + \sum_{i \epsilon I_3} x_i (\hat{\beta} + \hat{\lambda} x_i)^{-1} \Bigg] \\
&+ \Bigg[\sum_{i \epsilon I_1 \cup I_2} x_{1i} + \sum_{i \epsilon I_1 \cup I_2} x_{2i} + \sum_{i \epsilon I_3} x_i \Bigg] - \hat{\alpha}_1 \Bigg[\sum_{i \epsilon I_1 \cup I_2} x_{1i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{1i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i} \Bigg] \\
&- \hat{\alpha}_2 \Bigg[\sum_{i \epsilon I_1 \cup I_2} x_{2i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{2i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i} \Bigg] - \hat{\alpha_3} \Bigg[\sum_{i \epsilon I_1} x_{2i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{2i}} + \sum_{i \epsilon I_2} x_{1i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{1i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i} \Bigg].
\end{aligned}
$$

The elements of the Hessian matrix H, which is symmetric matrix, are given below by

$$H_{11} = \frac{\partial^2 l_p}{\partial \alpha_1^2} = -\left[\frac{n_1}{\hat{\alpha}_1^2} + \frac{n_2}{(\hat{\alpha}_1 + \hat{\alpha}_3)^2}\right], \quad H_{12} = H_{21} = \frac{\partial^2 l_p}{\partial \alpha_1 \partial \alpha_2} = 0, \quad H_{13} = H_{31} = \frac{\partial^2 l_p}{\partial \alpha_1 \partial \alpha_3} = -\left[\frac{n_2}{(\hat{\alpha}_1 + \hat{\alpha}_3)^2}\right],$$

$$H_{14} = H_{41} = \frac{\partial^2 l_p}{\partial \alpha_1 \partial \beta} = -\left[\sum_{i \epsilon I_1 \cup I_2} x_{1i}^{\hat{\beta}} e^{\hat{\lambda} x_{1i}} \ln x_{1i} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \ln x_i\right],$$

$$H_{15} = H_{51} = \frac{\partial^2 l_p}{\partial \alpha_1 \partial \lambda} = -\left[\sum_{i \epsilon I_1 \cup I_2} x_{1i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{1i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i}\right],$$

$$H_{22} = \frac{\partial^2 l_p}{\partial \alpha_2^2} = -\left[\frac{n_1}{(\hat{\alpha}_2 + \hat{\alpha}_3)^2} + \frac{n_2}{\hat{\alpha}_2^2}\right], \quad H_{23} = H_{32} = \frac{\partial^2 l_p}{\partial \alpha_2 \partial \alpha_3} = -\left[\frac{n_1}{(\hat{\alpha}_2 + \hat{\alpha}_3)^2}\right],$$

$$H_{24} = H_{42} = \frac{\partial^2 l_p}{\partial \alpha_2 \partial \beta} = -\left[\sum_{i \epsilon I_1 \cup I_2} x_{2i}^{\hat{\beta}} e^{\hat{\lambda} x_{2i}} \ln x_{2i} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \ln x_i\right],$$

$$H_{25} = H_{52} = \frac{\partial^2 l_p}{\partial \alpha_2 \partial \lambda} = -\left[\sum_{i \epsilon I_1 \cup I_2} x_{2i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{2i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i}\right], H_{33} = \frac{\partial^2 l_p}{\partial \alpha_3^2} = -\left[\frac{n_1}{(\hat{\alpha}_2 + \hat{\alpha}_3)^2} + \frac{n_2}{(\hat{\alpha}_1 + \hat{\alpha}_3)^2} + \frac{n_3}{\hat{\alpha}_3^2}\right],$$

$$H_{34} = H_{43} = \frac{\partial^2 l_p}{\partial \alpha_4 \partial \beta} = -\left[\sum_{i \epsilon I_1} x_{2i}^{\hat{\beta}} e^{\hat{\lambda} x_{2i}} \ln x_{2i} + \sum_{i \epsilon I_2} x_{1i}^{\hat{\beta}} e^{\hat{\lambda} x_{1i}} \ln x_{1i} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} \ln x_i\right],$$

$$H_{35} = H_{53} = \frac{\partial^2 l_p}{\partial \alpha_3 \partial \lambda} = -\left[\sum_{i \epsilon I_1} x_{2i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{2i}} + \sum_{i \epsilon I_2} x_{1i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{1i}} + \sum_{i \epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i}\right],$$

$$H_{44} = \frac{\partial^2 l_p}{\partial \beta^2} = -\Big[\sum_{i\epsilon I_1 \cup I_2} (\hat{\beta} + \hat{\lambda} x_{1i})^{-2} + \sum_{i\epsilon I_1 \cup I_2} (\hat{\beta} + \hat{\lambda} x_{2i})^{-2} + \sum_{i\epsilon I_3} (\hat{\beta} + \hat{\lambda} x_i)^{-2}\Big]$$
$$- \hat{\alpha}_1\Big[\sum_{i\epsilon I_1 \cup I_2} x_{1i}^{\hat{\beta}} e^{\hat{\lambda} x_{1i}} (\ln x_{1i})^2 + \sum_{i\epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} (\ln x_i)^2\Big] - \hat{\alpha}_2\Big[\sum_{i\epsilon I_1 \cup I_2} x_{2i}^{\hat{\beta}} e^{\hat{\lambda} x_{2i}} (\ln x_{2i})^2 + \sum_{i\epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} (\ln x_i)^2\Big]$$
$$- \hat{\alpha_3}\Big[\sum_{i\epsilon I_1} x_{2i}^{\hat{\beta}} e^{\hat{\lambda} x_{2i}} (\ln x_{2i})^2 + \sum_{i\epsilon I_2} x_{1i}^{\hat{\beta}} e^{\hat{\lambda} x_{1i}} (\ln x_{1i})^2 + \sum_{i\epsilon I_3} x_i^{\hat{\beta}} e^{\hat{\lambda} x_i} (\ln x_i)^2\Big],$$
$$H_{45} = H_{54} = \frac{\partial^2 l_p}{\partial \beta \partial \lambda} = -\Big[\sum_{i\epsilon I_1 \cup I_2} x_{1i} (\hat{\beta} + \hat{\lambda} x_{1i})^{-2} + \sum_{i\epsilon I_1 \cup I_2} x_{2i} (\hat{\beta} + \hat{\lambda} x_{2i})^{-2} + \sum_{i\epsilon I_3} x_i (\hat{\beta} + \hat{\lambda} x_i)^{-2}\Big]$$
$$- \hat{\alpha}_1\Big[\sum_{i\epsilon I_1 \cup I_2} x_{1i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{1i}} \ln x_{1i} + \sum_{i\epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i} \ln x_i\Big] - \hat{\alpha}_2\Big[\sum_{i\epsilon I_1 \cup I_2} x_{2i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{2i}} \ln x_{2i} + \sum_{i\epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i} \ln x_i\Big]$$
$$- \hat{\alpha_3}\Big[\sum_{i\epsilon I_1} x_{2i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{2i}} \ln x_{2i} + \sum_{i\epsilon I_2} x_{1i}^{\hat{\beta}+1} e^{\hat{\lambda} x_{1i}} \ln x_{1i} + \sum_{i\epsilon I_3} x_i^{\hat{\beta}+1} e^{\hat{\lambda} x_i} \ln x_i\Big],$$
$$H_{55} = \frac{\partial^2 l_p}{\partial \lambda^2} = -\Big[\sum_{i\epsilon I_1 \cup I_2} x_{1i}^2 (\hat{\beta} + \hat{\lambda} x_{1i})^{-2} + \sum_{i\epsilon I_1 \cup I_2} x_{2i}^2 (\hat{\beta} + \hat{\lambda} x_{2i})^{-2} + \sum_{i\epsilon I_3} x_i^2 (\hat{\beta} + \hat{\lambda} x_i)^{-2}\Big]$$
$$- \hat{\alpha}_1\Big[\sum_{i\epsilon I_1 \cup I_2} x_{1i}^{\hat{\beta}+2} e^{\hat{\lambda} x_{1i}} + \sum_{i\epsilon I_3} x_i^{\hat{\beta}+2} e^{\hat{\lambda} x_i}\Big] - \hat{\alpha}_2\Big[\sum_{i\epsilon I_1 \cup I_2} x_{2i}^{\hat{\beta}+2} e^{\hat{\lambda} x_{2i}} + \sum_{i\epsilon I_3} x_i^{\hat{\beta}+2} e^{\hat{\lambda} x_i}\Big]$$
$$- \hat{\alpha_3}\Big[\sum_{i\epsilon I_1} x_{2i}^{\hat{\beta}+2} e^{\hat{\lambda} x_{2i}} + \sum_{i\epsilon I_2} x_{1i}^{\hat{\beta}+2} e^{\hat{\lambda} x_{1i}} + \sum_{i\epsilon I_3} x_i^{\hat{\beta}+2} e^{\hat{\lambda} x_i}\Big].$$

Appendix II : Pseudo code for EM algorithm

The pseudo-code of the EM algorithm

- For given n observations, $\mathcal{D} = \{(x_{11}, x_{21}), (x_{12}, x_{22}), ..., (x_{1n}, x_{2n})\}$.
- Take $\mathbf{x}_1 = (x_{11}, x_{12}, ..., x_{1n})$ and $\mathbf{x}_2 = (x_{21}, x_{22}, ..., x_{2n})$.
- For $I_1 = \{(x_{1i}, x_{2i}) : x_{1i} < x_{2i}\}$ with $n_1 = |I_1|$,
  take $\mathbf{x}_{11} = (x_{1l_1}, x_{1l_2}, ..., x_{1l_{n_1}})$ and $\mathbf{x}_{12} = (x_{2l_1}, x_{2l_2}, ..., x_{2l_{n_1}})$, where $\{l_1, l_2, ..., l_{n_1}\} \subset \{1, 2, 3..., n\}$.
- For $I_2 = \{(x_{1i}, x_{2i}) : x_{1i} > x_{2i}\}$ with $n_2 = |I_2|$,
  take $\mathbf{x}_{21} = (x_{1m_1}, x_{1m_2}, ..., x_{1m_{n_2}})$ and $\mathbf{x}_{22} = (x_{2m_1}, x_{2m_2}, ..., x_{2m_{n_2}})$, where $\{m_1, m_2, ..., m_{n_2}\} \subset \{1, 2, 3..., n\}$.
- For $I_3 = \{(x_{1i}, x_{2i}) : x_{1i} = x_{2i} = x_i\}$ with $n_3 = |I_3|$, take $\mathbf{x}_3 = (x_{j_1}, x_{j_2}, ..., x_{j_{n_3}})$, where $\{j_1, j_2, ..., j_{n_3}\} \subset \{1, 2, 3..., n\}$.

- Take initial guesses for five unknown parameters $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta, \lambda)$ as $\theta^{(0)} = (\alpha_1^{(0)}, \alpha_2^{(0)}, \alpha_3^{(0)}, \beta^{(0)}, \lambda^{(0)})$
- **The k^{th}, $k = 1, 2, 3, ...$, iteration of the EM Algorithm**

Step1. Calculate

$$r_1^{(k)} = \frac{\alpha_3^{(k-1)}}{\alpha_2^{(k-1)} + \alpha_3^{(k-1)}}, \qquad r_2^{(k)} = \frac{\alpha_2^{(k-1)}}{\alpha_2^{(k-1)} + \alpha_3^{(k-1)}}$$

$$s_1^{(k)} = \frac{\alpha_3^{(k-1)}}{\alpha_1^{(k-1)} + \alpha_3^{(k-1)}} \qquad s_2^{(k)} = \frac{\alpha_1^{(k-1)}}{\alpha_2^{(k-1)} + \alpha_3^{(k-1)}}$$

using equations (6) & (7).

Step2. Calculate the profile log-likelihood (pll) function of $\eta = (\beta, \lambda)$

$$pll(\eta) = l_p\big(\alpha_1^{(k-1)}(\eta), \alpha_2^{(k-1)}(\eta), \alpha_3^{(k-1)}(\eta), \beta^{(k-1)}, \lambda^{(k-1)}\big)$$

using equation (8). This is two-dimensional optimization problem. The MLE of η obtained by some standard optimization techniques, e.g., Nelder-Mead method, Newton-Raphson method etc. Here we use the most famous Nelder-Mead method to solve two-dimensional problem available in R under the function "optim()". Also the method is available in another programming environments, say, in Mathemica under the function "NMinimum()", in SAS under the function "nlpms()", in Matlab under the function "fminsearch()", in Python under the function "minimize()", etc. The MLE is obtained as

$$\eta^{(k)} = \arg\max_{\eta} pll(\eta).$$

Step3. Calculate the MLEs of α_1, α_2 and α_3 as

$$\alpha_1^{(k)} = \frac{n_1 + s_2^{(k)} n_2}{\sum_{i=1}^{n_1} x_{1l_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{1l_i}} + \sum_{i=1}^{n_2} x_{1m_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{1m_i}} + \sum_{i=1}^{n_3} x_{j_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{j_i}}}$$

$$\alpha_2^{(k)} = \frac{r_2^{(k)} n_1 + n_2}{\sum_{i=1}^{n_1} x_{2l_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{2l_i}} + \sum_{i=1}^{n_2} x_{2m_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{2m_i}} + \sum_{i=1}^{n_3} x_{j_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{j_i}}}$$

and

$$\alpha_1^{(k)} = \frac{r_1^{(k)} n_1 + s_1^{(k)} n_2}{\sum_{i=1}^{n_1} x_{2l_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{2l_i}} + \sum_{i=1}^{n_2} x_{1m_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{1m_i}} + \sum_{i=1}^{n_3} x_{j_i}^{\beta^{(k)}} e^{\lambda^{(k)} x_{j_i}}}$$

using equation (9).

Step4. Calculate the log-likelihood value at k^{th} iteration

$$l(k) = l\big(\alpha_1^{(k)}, \alpha_2^{(k)}, \alpha_3^{(k)}, \beta^{(k)}, \lambda^{(k)}\big)$$

using equation (4). And the pseudo log-likelihood value is

$$l_p(k) = l_p\big(\alpha_1^{(k)}, \alpha_2^{(k)}, \alpha_3^{(k)}, \beta^{(k)}, \lambda^{(k)}\big)$$

using equation (8).

Step5. Calculate the ratio

$$r(k) = \left| \frac{l_p(k) - l_p(k-1)}{l_p(k-1)} \right|.$$

Step6. Check if $r(k) < 10^{-8}$, then stop the algorithm and the MLE is obtained as $\hat{\theta} = \theta^{(k)}$.

Step7. Otherwise repeat all the steps from 1 to 6 until the convergence criteria met.