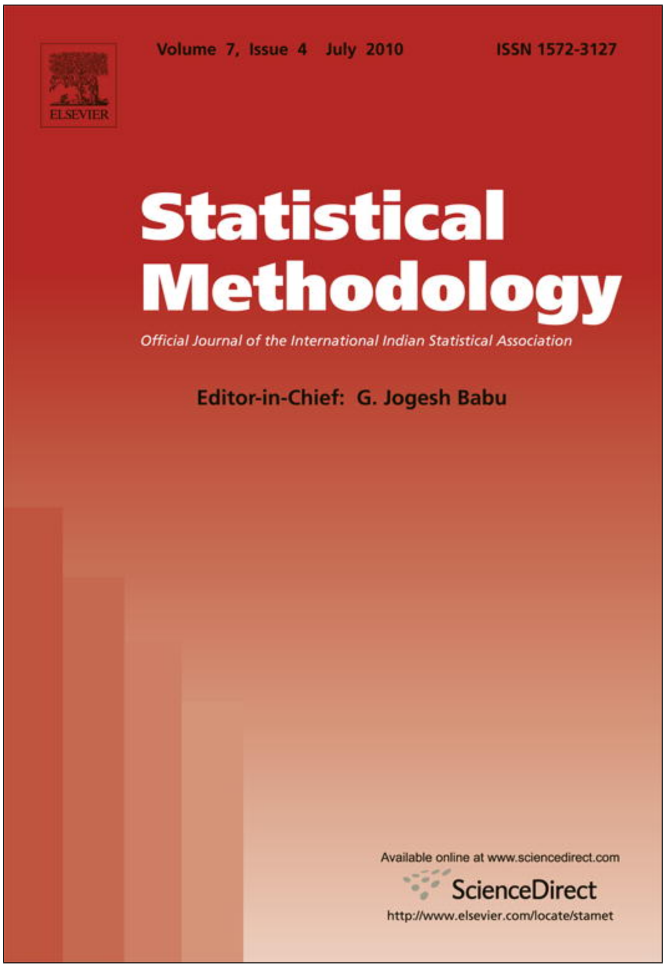
Volume 7, Issue 4 July 2010
ISSN 1572-3127
ELSEVIER
Statistical Methodology
Official Journal of the International Indian Statistical Association
Editor-in-Chief: G. Jogesh Babu
Available online at www.sciencedirect.com
ScienceDirect
http://www.elsevier.com/locate/stamet

Statistical Methodology 7 (2010) 464–477

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

A class of absolutely continuous bivariate distributions

Debasis Kundu [a,*], Rameshwar D. Gupta [b]

[a] Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, Pin 208016, India

[b] Department of Computer Science and Statistics, The University of New Brunswick at Saint John, New Brunswick, Canada E2L 4L5

ARTICLE INFO

Article history:
Received 27 September 2009
Received in revised form
22 December 2009
Accepted 13 January 2010

Keywords:
Block and Basu bivariate exponential model
Joint probability density function
Conditional probability density function
Maximum likelihood estimators
Asymptotic distribution
Fisher information matrix
EM algorithm

ABSTRACT

Block and Basu bivariate exponential distribution is one of the most popular absolutely continuous bivariate distributions. Extensive work has been done on the Block and Basu bivariate exponential model over the past several decades. Interestingly it is observed that the Block and Basu bivariate exponential model can be extended to the Weibull model also. We call this new model as the Block and Basu bivariate Weibull model. We consider different properties of the Block and Basu bivariate Weibull model. The Block and Basu bivariate Weibull model has four unknown parameters and the maximum likelihood estimators cannot be obtained in closed form. To compute the maximum likelihood estimators directly, one needs to solve a four dimensional optimization problem. We propose to use the EM algorithm for computing the maximum likelihood estimators of the unknown parameters. The proposed EM algorithm can be carried out by solving one non-linear equation at each EM step. Our method can be also used to compute the maximum likelihood estimators for the Block and Basu bivariate exponential model. One data analysis has been preformed for illustrative purpose.

1. Introduction

Block and Basu [3] obtained the bivariate exponential distribution (BBBE) from the Marshall–Olkin bivariate exponential (MOBE) distribution by removing the singular part and retaining only the absolutely continuous part. Although MOBE is a singular bivariate exponential distribution, BBBE distribution enjoys all the properties of an absolutely continuous distribution. Because of this reason,

* Corresponding author. Tel.: +91 512 2597141; fax: +91 512 2597500.
E-mail address: kundu@iitk.ac.in (D. Kundu).

1572-3127/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2010.01.004

BBBE is a very popular bivariate distribution. It has been used extensively for data analysis purposes, even though it is known that the marginals of BBBE are not exponential unlike MOBE model.

Along the same line as MOBE distribution, see for example Marshall and Olkin [11], Block and Basu bivariate Weibull (BBBW) distribution has been defined. BBBW model has been obtained from the Marshall–Olkin bivariate Weibull (MOBW) model by removing the singular part and that makes BBBW distribution as an absolutely continuous bivariate distribution. Clearly, it is a more flexible model than BBBE model because of the presence of the shape parameter. Although extensive work has been done on BBBE model, not that much of attention has been paid on BBBW model. The reason might be, in spite of the fact BBBW is more flexible than BBBE, computationally it may not be very tractable. In fact computing the maximum likelihood estimators (MLEs) of the unknown parameters of BBBW model is not a trivial issue.

In this paper we provide different properties of BBBW model. We discuss different computational issues associated in computing the parameters of BBBW model. First we consider the computation of MLEs of the four unknown parameters of BBBW model. It is observed that MLEs cannot be obtained in explicit form as expected, and they can be obtained by solving a multidimensional optimization problem. It is observed that the EM algorithm can be used quite effectively to compute MLEs of BBBW parameters. At each step (iteration), one needs to solve only a one dimensional optimization problem, and we have proposed a simple procedure to solve this problem. We have also provided the observed and expected Fisher information matrix. The expected Fisher information matrix provides the dispersion matrix of the asymptotic distribution of MLEs and the observed Fisher information matrix is needed to compute the approximate confidence intervals of the unknown parameters.

Although MOBE or BBBE model has been proposed quite some times back, only recently Karlis [7] proposed an efficient estimation technique to compute the unknown parameters of MOBE model. Moreover, in spite of the fact BBBE model has been derived from the Marshall–Olkin bivariate exponential model, it is not clear how the method of Karlis [7] can be used to compute MLEs of BBBE model. Note that BBBE model can be obtained as a special case of BBBW model and the proposed EM algorithm can be used very easily for BBBE model also. In fact in case of BBBE model, at each EM step no non-linear equation needs to be solved.

The remaining part of the paper is organized as follows. In Section 2 we introduce the model and provide different properties of the model in Section 3. The maximum likelihood estimators are discussed in Section 4. Analysis of one data set has been presented in Section 5 and finally we conclude the paper in Section 6.

2. BBBW: Model description

We use the following notation for the rest of the paper. If X has a univariate Weibull distribution with the shape and scale parameters as $\alpha > 0$ and $\lambda > 0$ respectively, then for $x > 0$, the probability density function (PDF) is defined as follows;

$$f_{\mathrm{WE}}(x; \alpha, \lambda) = \alpha\lambda x^{\alpha-1} \mathrm{e}^{-\lambda x^{\alpha}}. \tag{1}$$

The corresponding survival function (SF) and the hazard function (HF) will be denoted by $S_{\mathrm{WE}}(x; \alpha, \lambda)$ and $h_{\mathrm{WE}}(x; \alpha, \lambda)$ respectively. A Weibull distribution with the shape parameter α and the scale parameter λ will be denoted by $\mathrm{WE}(\alpha, \lambda)$.

Suppose U_0 follows ($\sim$) $\mathrm{WE}(\alpha, \lambda_0)$, $U_1 \sim \mathrm{WE}(\alpha, \lambda_1)$ and $U_2 \sim \mathrm{WE}(\alpha, \lambda_2)$ and they are mutually independent. If $X_1 = \min\{U_0, U_1\}$ and $X_2 = \min\{U_0, U_2\}$, then the joint distribution function of (X_1, X_2) is called the Marshall–Olkin bivariate Weibull distribution. The joint survival function of (X_1, X_2) can be written for $z = \max\{x_1, x_2\}$ as;

$$\begin{aligned} S_{MO}(x_1, x_2) &= P(X_1 > x_1, X_2 > x_2) = P(U_0 > z, U_1 > x_1, U_2 > x_2) \\ &= S_{\mathrm{WE}}(x_1; \alpha, \lambda_1) S_{\mathrm{WE}}(x_2; \alpha, \lambda_2) S_{\mathrm{WE}}(z; \alpha, \lambda_0) \\ &= \begin{cases} S_{\mathrm{WE}}(x_1; \alpha, \lambda_1) S_{\mathrm{WE}}(x_2; \alpha, \lambda_0 + \lambda_2) & \text{if } 0 < x_1 < x_2 < \infty \\ S_{\mathrm{WE}}(x_1; \alpha, \lambda_0 + \lambda_1) S_{\mathrm{WE}}(x_2; \alpha, \lambda_2) & \text{if } 0 < x_2 < x_1 < \infty \\ S_{\mathrm{WE}}(x; \alpha, \lambda_0 + \lambda_1 + \lambda_2) & \text{if } 0 < x_1 = x_2 = x < \infty. \end{cases} \end{aligned} \tag{2}$$

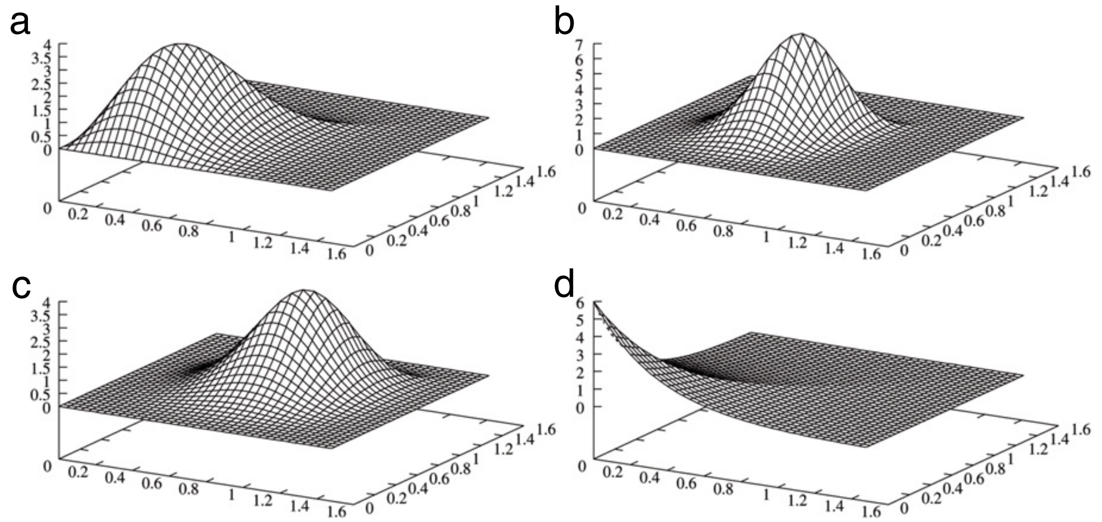

Fig. 1. The joint PDFs of (Y_1, Y_2). (a) $\alpha = 3$, $\lambda_0 = 4.0$, (b) $\alpha = 4$, $\lambda_0 = 4.0$, (c) $\alpha = 4$, $\lambda_0 = 1.0$, (d) $\alpha = 1$, $\lambda_0 = 2.0$.

It may be observed that the joint survival function of (X_1, X_2) can be written as a mixture of an absolutely continuous part and a singular part as follows;

$$S_{MO}(x_1, x_2) = \frac{\lambda_1 + \lambda_2}{\lambda_0 + \lambda_1 + \lambda_2} S_a(x_1, x_2) + \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} S_s(x_1, x_2), \tag{3}$$

where $S_a(\cdot, \cdot)$ is the absolutely continuous part and $S_s(\cdot, \cdot)$ is the singular part. Note that for $z = \max\{x_1, x_2\}$,

$$S_s(x_1, x_2) = S_{\text{WE}}(z; \alpha, \lambda_0 + \lambda_1 + \lambda_1)$$

and $S_a(x_1, x_2)$ can be obtained by subtraction as

$$S_a(x_1, x_2) = \frac{\lambda_0 + \lambda_1 + \lambda_2}{\lambda_1 + \lambda_2} e^{-\lambda_1 x_1^{\alpha}} e^{-\lambda_2 x_2^{\alpha}} e^{-\lambda_0 z^{\alpha}} - \frac{\lambda_0}{\lambda_1 + \lambda_2} e^{-(\lambda_0 + \lambda_1 + \lambda_2) z^{\alpha}}. \tag{4}$$

Note that BBBW distribution can be obtained from MOBW distribution by removing the singular part and keeping only the continuous part. The joint PDF of BBBW can be written as

$$f_{\text{BB}}(y_1, y_2) = \begin{cases} c f_1(y_1, y_2) = c f_{\text{WE}}(y_1; \alpha, \lambda_1) f_{\text{WE}}(y_2; \alpha, \lambda_0 + \lambda_2) & \text{if } 0 < y_1 < y_2 \\ c f_2(y_1, y_2) = c f_{\text{WE}}(y_1; \alpha, \lambda_0 + \lambda_1) f_{\text{WE}}(y_2; \alpha, \lambda_2) & \text{if } 0 < y_2 < y_1, \end{cases} \tag{5}$$

here c is the normalizing constant and $c = \frac{\lambda_0 + \lambda_1 + \lambda_2}{\lambda_1 + \lambda_2}$. Therefore, the joint PDF of (Y_1, Y_2) can be written as (5) and it will be denoted by BBBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$. The joint survival function of Y_1 and Y_2 is $S_a(\cdot, \cdot)$. The joint PDF of (Y_1, Y_2) is unimodal and the surface plot of $f_{\text{BB}}(y_1, y_2)$ for different values of α and λ_0, keeping $\lambda_1 = \lambda_2 = 1$, are provided in Fig. 1. From Fig. 1, it is clear that it can take different shapes, which can be very useful for data analysis purposes.

From the joint PDF $f_{\text{BB}}(y_1, y_2)$, it is immediate that when $\lambda_0 = 0$, then Y_1 and Y_2 become independent and both of them have Weibull distributions. For $\lambda_0 > 0$, Y_1 and Y_2 are dependent. It will be proved later that for $\lambda_0 > 0$, Y_1 and Y_2 are positively correlated.

From the construction of the Block and Basu bivariate Weibull distribution, it is immediate that if $(X_1, X_2) \sim$ MOBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$, and $(Y_1, Y_2) = (X_1, X_2)$, given that $X_1 \neq X_2$, then $(Y_1, Y_2) \sim$ BBBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$. Therefore, the generation from the BBBW is very simple. We can adopt the following simple generation technique to generate random samples from BBBW.

Algorithm to generate from BBBW.

- Step 1: Generate $U_0 \sim \text{WE}(\alpha, \lambda_0)$, $U_1 \sim \text{WE}(\alpha, \lambda_1)$ and $U_2 \sim \text{WE}(\alpha, \lambda_2)$ independently.
- Step 2: If $U_0 < U_1$ and $U_0 < U_2$ go back to Step 1 otherwise set $Y_1 = \min\{U_0, U_1\}$ and $Y_2 = \min\{U_0, U_2\}$.

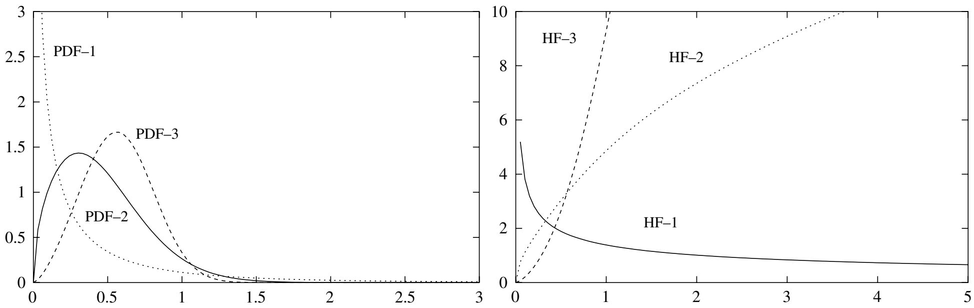

Fig. 2. The PDFs and HFs of Y_1 for, PDF-1/HF-1: $\alpha = 0.5$, $\lambda_0 = 2.0$, PDF-2/HF-2: $\alpha = 1.5$, $\lambda_0 = 2.5$, PDF-3/HF-3: $\alpha = 2.5$, $\lambda_0 = 3.0$.

3. Different properties

In this section we provide different basic properties of the BBBW model. First we provide the marginal and conditional distributions of BBBW model.

Theorem 3.1. *If* $(Y_1, Y_2) \sim \text{BBBW}(\alpha, \lambda_0, \lambda_1, \lambda_2)$, *then the marginal PDFs of* Y_1 *and* Y_2 *are*

$$f_{Y_1}(y_1) = cf_{\text{WE}}(y_1; \alpha, \lambda_0 + \lambda_1) - c\frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} f_{\text{WE}}(y_1; \alpha, \lambda_0 + \lambda_1 + \lambda_2) \quad (6)$$

and

$$f_{Y_2}(y_2) = cf_{\text{WE}}(y_2; \alpha, \lambda_0 + \lambda_2) - c\frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} f_{\text{WE}}(y_2; \alpha, \lambda_0 + \lambda_1 + \lambda_2) \quad (7)$$

respectively, where c is the same as before.

Proof. They can be obtained by routine calculations. ■

Now we will discuss some of the basic properties of the marginals of BBBW. We mainly mention the properties of Y_1. The properties of Y_2 are exactly the same. From Theorem 3.1, it is clear as expected that the distribution of Y_1 is not Weibull distribution in general. For $\lambda_0 = 0$ or $\lambda_2 = 0$, the distribution of Y_1 becomes Weibull. Although the PDF of Y_1 is not Weibull in general, but the shape of the PDF of Y_1 is very similar to the PDF of a Weibull distribution. The hazard function of Y_1 is

$$h_{Y_1}(y) = (\lambda_0 + \lambda_1 + \lambda_2)\alpha y^{\alpha-1}\left[1 - \frac{\lambda_2}{\lambda_0 + \lambda_1 + \lambda_2 - \lambda_0 e^{-\lambda_2 y^{\alpha}}}\right]. \quad (8)$$

The hazard function (HF) of Y_1 can be increasing, decreasing or constant depending on the values of α and λ's. The PDFs and HFs of Y_1 for different values of α and λ_0 for $\lambda_1 = \lambda_2 = 1$ are provided in Fig. 2. The PDF of Y_1 can be written as a weighted version of $\text{WE}(\alpha, \lambda_0 + \lambda_1)$ distribution with the weight function proportional to

$$w_1(y) = \lambda_0 + \lambda_1 - \lambda_0 e^{-\lambda_2 y^{\alpha}}.$$

Therefore many properties which are available for the weighted distributions can be applied here also. Moreover, all the moments of Y_1 exist and they can be written in terms of gamma functions. Now we will discuss the conditional PDFs.

Theorem 3.2. *If* $(Y_1, Y_2) \sim \text{BBBW}(\alpha, \lambda_0, \lambda_1, \lambda_2)$, *then the conditional PDFs of* $Y_1|Y_2 = y_2$ *and* $Y_2|Y_1 = y_1$ *are*

$$f_{Y_1|Y_2=y_2}(y_1) = \begin{cases} \dfrac{\lambda_0 + \lambda_2}{\lambda_2 + \lambda_0(1 - e^{-\lambda_1 y_2^{\alpha}})} f_{\text{WE}}(y_1; \alpha, \lambda_1) & \text{if } y_1 < y_2 \\ \dfrac{\lambda_2}{e^{-\lambda_0 y_2^{\alpha}}\left(\lambda_2 + \lambda_0(1 - e^{-\lambda_1 y_2^{\alpha}})\right)} f_{\text{WE}}(y_1; \alpha, \lambda_0 + \lambda_1) & \text{if } y_2 < y_1 \end{cases} \quad (9)$$

and

$$f_{Y_2|Y_1=y_1}(y_2) = \begin{cases} \dfrac{\lambda_1}{e^{-\lambda_0 y_1^\alpha}\left(\lambda_1 + \lambda_0(1 - e^{-\lambda_2 y_1^\alpha})\right)} f_{WE}(y_2; \alpha, \lambda_0 + \lambda_2) & \text{if } y_1 < y_2 \\ \dfrac{\lambda_0 + \lambda_1}{\lambda_1 + \lambda_0(1 - e^{-\lambda_2 y_1^\alpha})} f_{WE}(y_2; \alpha, \lambda_2) & \text{if } y_2 < y_1 \end{cases} \quad (10)$$

respectively.

Proof. They can be obtained by routine calculation. ■

Theorems 3.1 and 3.2 also can be used to generate samples from BBBW. The bivariate survival function of Y_1 and Y_2 is

$$S_{BB}(y_1, y_2) = \begin{cases} c e^{-\lambda_1 y_1^\alpha} e^{-(\lambda_0+\lambda_2) y_2^\alpha} - \dfrac{\lambda_0}{\lambda_1 + \lambda_2} e^{-(\lambda_0+\lambda_1+\lambda_2) y_2^\alpha} & \text{if } y_1 < y_2 \\ c e^{-(\lambda_0+\lambda_1) y_1^\alpha} e^{-\lambda_2 y_2^\alpha} - \dfrac{\lambda_0}{\lambda_1 + \lambda_2} e^{-(\lambda_0+\lambda_1+\lambda_2) y_1^\alpha} & \text{if } y_2 < y_1 \end{cases} \quad (11)$$

and the marginal survival functions of Y_1 and Y_2 are

$$S_{Y_1}(y_1) = c e^{-(\lambda_0+\lambda_1) y_1^\alpha} - \frac{\lambda_0}{\lambda_1 + \lambda_2} e^{-(\lambda_0+\lambda_1+\lambda_2) y_1^\alpha} \quad \text{and} \quad (12)$$

$$S_{Y_2}(y_2) = c e^{-(\lambda_0+\lambda_2) y_2^\alpha} - \frac{\lambda_0}{\lambda_1 + \lambda_2} e^{-(\lambda_0+\lambda_1+\lambda_2) y_2^\alpha}, \quad (13)$$

respectively.

The bivariate failure rate as defined by Basu [2] can be written as

$$r(y_1, y_2) = \frac{f_{BB}(y_1, y_2)}{S_{BB}(y_1, y_2)} = \begin{cases} \dfrac{h_{WE}(y_1; \alpha, \lambda_1) h_{WE}(y_2; \alpha, \lambda_0 + \lambda_2)}{1 - \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} e^{-\lambda_1 (y_2^\alpha - y_1^\alpha)}} & \text{if } y_1 < y_2 \\ \dfrac{h_{WE}(y_1; \alpha, \lambda_0 + \lambda_1) h_{WE}(y_2; \alpha, \lambda_2)}{1 - \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} e^{-\lambda_1 (y_1^\alpha - y_2^\alpha)}} & \text{if } y_1 > y_2. \end{cases} \quad (14)$$

It is clear that when $\alpha = 1$, it reduces to Block and Basu bivariate failure rate. When they are independent (14) reduces to product of two Weibull hazards as expected. Moreover, interestingly, along the curve $y_1^\alpha = y_2^\alpha + b$, for some constant b, $r(y_1, y_2)$ is an increasing, decreasing or constant according as $\alpha > 1$, $\alpha < 1$ or $\alpha = 1$ respectively.

The hazard gradients, see [5], of BBBW are

$$\begin{aligned} h_1(y_1, y_2) &= -\frac{\partial}{\partial y_1} \ln S(y_1, y_2) \\ &= \begin{cases} \dfrac{h_{WE}(y_1; \alpha, \lambda_1)}{1 - \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} e^{-\lambda_1 (y_2^\alpha - y_1^\alpha)}} & \text{if } y_1 < y_2 \\ \dfrac{h_{WE}(y_1; \alpha, \lambda_0 + \lambda_1)}{1 - \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} e^{-\lambda_2 (y_1^\alpha - y_2^\alpha)}} + \dfrac{h_{WE}(y_1; \alpha, \lambda_0 + \lambda_1 + \lambda_2)}{1 - \frac{\lambda_0+\lambda_1+\lambda_2}{\lambda_0} e^{-\lambda_2 (y_2^\alpha - y_1^\alpha)}} & \text{if } y_2 < y_1 \end{cases} \end{aligned}$$

and

$$\begin{aligned} h_2(y_1, y_2) &= -\frac{\partial}{\partial y_2} \ln S(y_1, y_2) \\ &= \begin{cases} \dfrac{h_{WE}(y_2; \alpha, \lambda_0 + \lambda_2)}{1 - \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} e^{-\lambda_1 (y_2^\alpha - y_1^\alpha)}} + \dfrac{h_{WE}(y_1; \alpha, \lambda_0 + \lambda_1 + \lambda_2)}{1 - \frac{\lambda_0+\lambda_1+\lambda_2}{\lambda_0} e^{-\lambda_1 (y_1^\alpha - y_2^\alpha)}} & \text{if } y_1 < y_2 \\ \dfrac{h_{WE}(y_2; \alpha, \lambda_2)}{1 - \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2} e^{-\lambda_2 (y_1^\alpha - y_2^\alpha)}} & \text{if } y_1 > y_2. \end{cases} \end{aligned}$$

Now we provide the total positivity result of Y_1 and Y_2, for identical marginals.

Theorem 3.3. *If* $(Y_1, Y_2) \sim \text{BBW}(\alpha, \lambda_0, \lambda, \lambda)$, *then* (Y_1, Y_2) *has the total positivity of order two* (TP_2) *property.*

Proof. Note that (Y_1, Y_2) has the TP_2 property if and only if for any $y_{11}, y_{12}, y_{21}, y_{22}$, whenever, $0 < y_{11} < y_{12}$ and $0 < y_{21} < y_{22}$, then

$$f_{Y_1,Y_2}(y_{11}, y_{21}) f_{Y_1,Y_2}(y_{12}, y_{22}) \geq f_{Y_1,Y_2}(y_{12}, y_{21}) f_{Y_1,Y_2}(y_{11}, y_{22}). \tag{15}$$

Now we consider different cases as follows;

Case 1: $y_{21} < y_{22} < y_{11} < y_{12}$.
In this case

$$f_{Y_1,Y_2}(y_{11}, y_{21}) f_{Y_1,Y_2}(y_{12}, y_{22}) - f_{Y_1,Y_2}(y_{12}, y_{21}) f_{Y_1,Y_2}(y_{11}, y_{22}) = 0.$$

Case 2: $y_{21} < y_{11} < y_{22} < y_{12}$.

In this case it can be easily observed by simple calculation that, to prove (15) is equivalent to prove

$$e^{-\lambda_0 y_{11}^{\alpha}} \geq e^{-\lambda_0 y_{22}^{\alpha}}. \tag{16}$$

Clearly (16) is true as $y_{11} < y_{22}$. Similarly, for other cases also it can be proved along the similar line. ■

Since (Y_1, Y_2) have TP_2 property, therefore, for any non-decreasing function $U(\cdot, \cdot)$ and $V(\cdot, \cdot)$, $\text{Cov}(U(Y_1, Y_2), V(Y_1, Y_2)) > 0$. It implies that Y_1 and Y_2 are also positively correlated. (Y_1, Y_2) will also be positively likelihood ratio dependent, *i.e.* it satisfies (15), see [9]. Moreover, since the joint density function is TP_2, the joint survival function is also TP_2, which is equivalent to right corner set increasing, *i.e.* $P(Y_1 > y_1, Y_2 > y_2 | Y_1 > y_1', Y_2 > y_2')$ is non-decreasing in y_1', y_2', for every choice of y_1, y_2, see [4].

The following results will be useful for data analysis purposes. The proofs are quite trivial and therefore omitted.

Theorem 3.4. *Suppose* $(Y_1, Y_2) \sim \text{BBBW}(\alpha, \lambda_0, \lambda_1, \lambda_2)$.

(a) *The stress–strength parameter* $R = P(Y_1 < Y_2)$, *has the following form;*

$$R = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

(b) $\min\{Y_1, Y_2\} \sim \text{WE}(\alpha, \lambda_0 + \lambda_1 + \lambda_2)$.
(c) $Y_1 | \{Y_1 < Y_2\} \sim \text{WE}(\alpha, \lambda_0 + \lambda_1 + \lambda_2)$.
(d) $Y_2 | \{Y_2 < Y_1\} \sim \text{WE}(\alpha, \lambda_0 + \lambda_1 + \lambda_2)$.

The BBBW model satisfies all the regularity conditions for the MLEs to be consistent and asymptotically normal and we can state the following result:

Theorem 3.5. *If* $\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1$ *and* $\widehat{\lambda}_2$ *are the MLEs of* $\alpha, \lambda_0, \lambda_1$ *and* λ_2 *respectively, then*

$$\sqrt{n}\left\{\widehat{\alpha} - \alpha, \widehat{\lambda}_0 - \lambda_0, \widehat{\lambda}_1 - \lambda_1, \widehat{\lambda}_2 - \lambda_2\right\} \to N_4(0, I^{-1}), \tag{17}$$

here I *is the Fisher information matrix and the exact expression of* I *is provided in Appendix* B.

In the next section, we provide the method of finding the MLEs.

4. Maximum likelihood estimators

In this section we are mainly discussing about the computation of the MLEs of the unknown parameters $\alpha, \lambda_0, \lambda_1$ and λ_2 of the BBBW distribution, when we have a random sample $\{(y_{11}, y_{21}), \ldots, (y_{1n}, y_{2n})\}$. We use the following notation: $I_1 = \{i : y_{1i} < y_{2i}\}$ and $I_2 = \{i : y_{1i} > y_{2i}\}$, $|I_1| = n_1$ and $|I_2| = n_2$. Based on the above notation, the log-likelihood function can be written as

$$\begin{aligned} l(\alpha, \lambda_0, \lambda_1, \lambda_2) &= n \ln(\lambda_0 + \lambda_1 + \lambda_2) - n \ln(\lambda_1 + \lambda_2) + \sum_{i \in I_1} \ln f_{\mathrm{WE}}(y_{1i}; \alpha, \lambda_1) \\ &\quad + \sum_{i \in I_1} \ln f_{\mathrm{WE}}(y_{2i}; \alpha, \lambda_0 + \lambda_2) + \sum_{i \in I_2} \ln f_{\mathrm{WE}}(y_{1i}; \alpha, \lambda_0 + \lambda_1) \\ &\quad + \sum_{i \in I_2} \ln f_{\mathrm{WE}}(y_{1i}; \alpha, \lambda_2) \\ &= n \ln(\lambda_0 + \lambda_1 + \lambda_2) - n \ln(\lambda_1 + \lambda_2) + (\alpha - 1) \sum_{I} [\ln y_{1i} + \ln y_{2i}] \\ &\quad + 2n \ln \alpha + n_1 (\ln \lambda_1 + \ln(\lambda_0 + \lambda_2)) + n_2 (\ln \lambda_2 + \ln(\lambda_0 + \lambda_1)) \\ &\quad - \lambda_1 \sum_{i \in I} y_{1i}^{\alpha} - \lambda_2 \sum_{i \in I} y_{2i}^{\alpha} - \lambda_0 \left(\sum_{i \in I_1} y_{1i}^{\alpha} + \sum_{i \in I_2} y_{2i}^{\alpha} \right). \end{aligned} \tag{18}$$

Clearly, the MLEs of α, λ_0, λ_1 and λ_2 cannot be obtained in explicit forms. They can be obtained only by solving four equations in four unknowns. We treat this problem as a missing value problem. Observe that if all the U_0, U_1 and U_2 are known, the maximum likelihood estimates of the unknown parameters can be obtained by solving a one dimensional optimization problem. We try to exploit this property to treat this problem as a missing value problem through EM algorithm.

First let us look at the complete observations and then we provide the missing observations. Let us define a pair of random variables (Δ_1, Δ_2) associated with each (X_1, X_2) as follows;

$$(\Delta_1, \Delta_2) = \begin{cases} (0, 0) & \text{if } X_1 = U_0,\ X_2 = U_0 \\ (0, 2) & \text{if } X_1 = U_0,\ X_2 = U_2 \\ (1, 0) & \text{if } X_1 = U_1,\ X_2 = U_0 \\ (1, 2) & \text{if } X_1 = U_1,\ X_2 = U_2. \end{cases} \tag{19}$$

If a sample is obtained from $(X_1, X_2, \Delta_1, \Delta_2)$, we call it as the complete observation. Note that if we have a random sample of size m from $(X_1, X_2, \Delta_1, \Delta_2)$, then the MLEs of the unknown parameters can be obtained as a one dimensional optimization problem as follows. Suppose we have the following observations; $\{(x_{1i}, x_{2i}, \delta_{1i}, \delta_{2i}),\ i = 1, \ldots m\}$. We use the following notation;

$$\begin{aligned} &I_0 = \{i; x_{1i} = x_{2i} = x_i, \delta_{1i} = \delta_{2i} = 0\}, \qquad I_{02} = \{i; x_{1i} > x_{2i}, \delta_{1i} = 0, \delta_{2i} = 2\} \\ &I_{10} = \{i; x_{1i} < x_{2i}, \delta_{1i} = 1, \delta_{2i} = 0\}, \qquad I_{121} = \{i; x_{1i} < x_{2i}, \delta_{1i} = 1, \delta_{2i} = 2\} \\ &I_{122} = \{i; x_{1i} > x_{2i}, \delta_{1i} = 1, \delta_{2i} = 2\}. \end{aligned}$$

The likelihood contributions of the observations from the set $I_0, I_{02}, I_{10}, I_{121}, I_{122}$ are $\alpha\lambda_0 x_i^{\alpha-1} \mathrm{e}^{-(\lambda_0+\lambda_1+\lambda_2)x_i^{\alpha}}$, $\alpha\lambda_2 x_{2i}^{\alpha-1} \mathrm{e}^{-\lambda_2 x_{2i}^{\alpha}} \alpha\lambda_0 x_{1i}^{\alpha-1} \mathrm{e}^{-(\lambda_0+\lambda_1)x_{1i}^{\alpha}}$, $\alpha\lambda_1 x_{1i}^{\alpha-1} \mathrm{e}^{-\lambda_1 x_{1i}^{\alpha}} \alpha\lambda_0 x_{2i}^{\alpha-1} \mathrm{e}^{-(\lambda_0+\lambda_2)x_{2i}^{\alpha}}$, $\alpha\lambda_1 x_{1i}^{\alpha-1} \mathrm{e}^{-\lambda_1 x_{1i}^{\alpha}} \alpha\lambda_2 x_{2i}^{\alpha-1} \mathrm{e}^{-(\lambda_0+\lambda_2)x_{2i}^{\alpha}}$, $\alpha\lambda_1 x_{1i}^{\alpha-1} \mathrm{e}^{-(\lambda_0+\lambda_1)x_{1i}^{\alpha}} \alpha\lambda_2 x_{2i}^{\alpha-1} \mathrm{e}^{-\lambda_2 x_{2i}^{\alpha}}$ respectively. It is clear from the above likelihood contributions, that given the complete observations, the MLEs of the unknown parameters can be obtained by solving one non-linear equation only.

In our case the missing observations are the whole set I_0. We do not observe I_{10} and I_{121} separately, but we observe $I_1 = I_{10} \cup I_{121}$. Therefore, in I_1, δ_{1i} is known, but δ_{2i} is missing. Similarly, we do not observe I_{02} and I_{122} separately, we observe $I_2 = I_{02} \cup I_{122}$. In this case δ_{1i} is missing but δ_{2i} is known. The observations in Table 1 will be useful for constructing E-step of the EM algorithm.
We use the following notation;

$$\gamma = (\lambda_0, \lambda_1, \lambda_2, \alpha), \qquad a_0 = E(U_0 | U_0 < \min\{U_1, U_2\}) = \frac{1}{(\lambda_0 + \lambda_1 + \lambda_2)^{1/\alpha}} \Gamma\left(\frac{1}{\alpha} + 1\right),$$

Table 1
All possible cases of U_0, U_1, U_2, corresponding probabilities and (Δ_1, Δ_2).

Different cases	Probability	(Δ_1, Δ_2)	X_1 & X_2	Set
$U_0 < U_1 < U_2$	$\frac{\lambda_0\lambda_1}{(\lambda_1+\lambda_2)(\lambda_0+\lambda_1+\lambda_2)}$	$(0, 0)$	$X_1 = X_2$	I_0
$U_0 < U_2 < U_1$	$\frac{\lambda_0\lambda_2}{(\lambda_1+\lambda_2)(\lambda_0+\lambda_1+\lambda_2)}$	$(0, 0)$	$X_1 = X_2$	I_0
$U_1 < U_0 < U_2$	$\frac{\lambda_0\lambda_1}{(\lambda_0+\lambda_2)(\lambda_0+\lambda_1+\lambda_2)}$	$(1, 0)$	$X_1 < X_2$	I_{10}
$U_1 < U_2 < U_0$	$\frac{\lambda_1\lambda_2}{(\lambda_0+\lambda_2)(\lambda_0+\lambda_1+\lambda_2)}$	$(1, 2)$	$X_1 < X_2$	I_{121}
$U_2 < U_0 < U_1$	$\frac{\lambda_0\lambda_2}{(\lambda_0+\lambda_1)(\lambda_0+\lambda_1+\lambda_2)}$	$(0, 2)$	$X_1 > X_2$	I_{02}
$U_2 < U_1 < U_0$	$\frac{\lambda_1\lambda_2}{(\lambda_0+\lambda_1)(\lambda_0+\lambda_1+\lambda_2)}$	$(1, 2)$	$X_1 > X_2$	I_{122}

$$u_1 = \frac{\lambda_0}{\lambda_0+\lambda_2} = P(\Delta_2 = 0|X_1 < X_2), \qquad u_2 = \frac{\lambda_2}{\lambda_0+\lambda_2} = P(\Delta_2 = 2|X_1 < X_2),$$
$$v_1 = \frac{\lambda_0}{\lambda_0+\lambda_1} = P(\Delta_1 = 0|X_1 > X_2), \qquad v_2 = \frac{\lambda_1}{\lambda_0+\lambda_1} = P(\Delta_1 = 1|X_1 > X_2),$$

and $n_0 = |I_0| = m - n_1 - n_2$ and $m \geq n$ is a random number which has negative binomial distribution with parameters $\frac{\lambda_1+\lambda_2}{\lambda_0+\lambda_1+\lambda_2}$ and $n_1 + n_2$. We further use,

$$\widetilde{n}_0 = E(n_0|n_1, n_2) = (n_1+n_2)\frac{\lambda_0}{\lambda_1+\lambda_2}.$$

Therefore, based on the observations $\{(y_{1i}, y_{2i}),\ i = 1, \ldots, n\}$, we write the pseudo-log-likelihood function by replacing the missing observation by its expected value. It becomes

$$\begin{aligned}
l_{\text{pseudo}}(\gamma) &= \widetilde{n}_0(\ln\alpha + \ln\lambda_0) + \widetilde{n}_0(\alpha-1)\ln a_0 - \widetilde{n}_0(\lambda_0+\lambda_1+\lambda_2)a_0^{\alpha} \\
&\quad + u_1\Bigg(n_1(\ln\alpha+\ln\lambda_1) - \lambda_1\sum_{i\in I_1} y_{1i}^{\alpha} + (\alpha-1)\sum_{i\in I_1}\ln y_{1i} + n_1(\ln\alpha+\ln\lambda_0) \\
&\quad - (\lambda_0+\lambda_2)\sum_{i\in I_1} y_{2i}^{\alpha} + (\alpha-1)\sum_{i\in I_1}\ln y_{2i}\Bigg) \\
&\quad + u_2\Bigg(n_1(\ln\alpha+\ln\lambda_1) - \lambda_1\sum_{i\in I_1} y_{1i}^{\alpha} + (\alpha-1)\sum_{i\in I_1}\ln y_{1i} + n_1(\ln\alpha+\ln\lambda_2) \\
&\quad - (\lambda_0+\lambda_2)\sum_{i\in I_1} y_{2i}^{\alpha} + (\alpha-1)\sum_{i\in I_1}\ln y_{2i}\Bigg) \\
&\quad + v_1\Bigg(n_2(\ln\alpha+\ln\lambda_2) - \lambda_2\sum_{i\in I_2} y_{2i}^{\alpha} + (\alpha-1)\sum_{i\in I_2}\ln y_{2i} + n_2(\ln\alpha+\ln\lambda_0) \\
&\quad - (\lambda_0+\lambda_1)\sum_{i\in I_2} y_{1i}^{\alpha} + (\alpha-1)\sum_{i\in I_2}\ln y_{1i}\Bigg) \\
&\quad + v_2\Bigg(n_2(\ln\alpha+\ln\lambda_2) - \lambda_2\sum_{i\in I_2} y_{2i}^{\alpha} + (\alpha-1)\sum_{i\in I_2}\ln y_{2i} + n_2(\ln\alpha+\ln\lambda_1) \\
&\quad - (\lambda_0+\lambda_1)\sum_{i\in I_2} y_{1i}^{\alpha} + (\alpha-1)\sum_{i\in I_2}\ln y_{1i}\Bigg) \\
&= (\widetilde{n}_0 + 2n_1 + 2n_2)\ln\alpha + (\alpha-1)\Bigg[\widetilde{n}_0\ln a_0 + \sum_{i\in I_1\cup I_2}(\ln y_{1i} + \ln y_{2i})\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&- \lambda_0 \left[\widetilde{n}_0 a_0^{\alpha} + \sum_{i \in I_1} y_{2i}^{\alpha} + \sum_{i \in I_2} y_{1i}^{\alpha} \right] + (\widetilde{n}_0 + u_1 n_1 + v_1 n_2) \ln \lambda_0 \\
&- \lambda_1 \left[\widetilde{n}_0 a_0^{\alpha} + \sum_{i \in I_1 \cup I_2} y_{1i}^{\alpha} \right] + \ln \lambda_1 (n_1 + v_2 n_2) \\
&- \lambda_2 \left[\widetilde{n}_0 a_0^{\alpha} + \sum_{i \in I_1 \cup I_2} y_{2i}^{\alpha} \right] + \ln \lambda_2 (u_2 n_1 + n_2).
\end{aligned}
$$

Therefore, for fixed α, at each step, the pseudo-maximum likelihood estimates of λ_0, λ_1 and λ_2 can be obtained as

$$
\begin{aligned}
&\widehat{\lambda}_0(\alpha) = \frac{\widetilde{n}_0 + u_1 n_1 + v_1 n_2}{\widetilde{n}_0 a_0^{\alpha} + \sum_{i \in I_2} y_{1i}^{\alpha} + \sum_{i \in I_1} y_{2i}^{\alpha}}, \qquad \widehat{\lambda}_1(\alpha) = \frac{n_1 + v_2 n_2}{\widetilde{n}_0 a_0^{\alpha} + \sum_{i \in I} y_{1i}^{\alpha}}, \\
&\widehat{\lambda}_2(\alpha) = \frac{n_2 + u_2 n_1}{\widetilde{n}_0 a_0^{\alpha} + \sum_{i \in I} y_{2i}^{\alpha}}.
\end{aligned}
\tag{20}
$$

The pseudo-maximum likelihood estimate of α at each step can be obtained by maximizing the pseudo-profile log-likelihood function $l_{\text{pseudo}}(\alpha, \widehat{\lambda}_0(\alpha), \widehat{\lambda}_1(\alpha), \widehat{\lambda}_2(\alpha))$. It has been shown in Appendix A (see Lemma 1) that the pseudo-log-likelihood function is a unimodal function and therefore it has the unique maximum.

The maximization of $l_{\text{pseudo}}(\alpha, \widehat{\lambda}_0(\alpha), \widehat{\lambda}_1(\alpha), \widehat{\lambda}_2(\alpha))$ with respect to α can be performed by solving the following fixed point equation

$$
g(\alpha) = \alpha, \tag{21}
$$

where $g(\alpha) = (\widetilde{n}_0 + 2n_1 + 2n_2)[h(\alpha)]^{-1}$, and

$$
\begin{aligned}
h(\alpha) = {} & \widehat{\lambda}_0(\alpha) \left[\widetilde{n}_0 a_0^{\alpha} \ln a_0 + \sum_{i \in I_1} y_{2i}^{\alpha} \ln y_{2i} + \sum_{i \in I_2} y_{1i}^{\alpha} \ln y_{2i} \right] + \widehat{\lambda}_1(\alpha) \left[\widetilde{n}_0 a_0^{\alpha} \ln a_0 + \sum_{i \in I} y_{1i}^{\alpha} \ln y_{1i} \right] \\
& + \widehat{\lambda}_2(\alpha) \left[\widetilde{n}_0 a_0^{\alpha} \ln a_0 + \sum_{i \in I} y_{2i}^{\alpha} \ln y_{2i} \right] - \left[\widetilde{n}_0 \ln a_0 + \sum_{i \in I} \ln y_{1i} + \sum_{i \in I} \ln y_{2i} \right].
\end{aligned}
$$

Note that solving (21) is quite simple. We suggest to use a similar algorithm as of Kundu and Gupta [8]. Start with an initial guess $\alpha^{(0)}$, then $\alpha^{(1)} = g(\alpha^{(0)})$, similarly, $\alpha^{(2)} = g(\alpha^{(1)})$ and so on. Continue the process until the convergence is obtained. Alternatively, the maximization of the pseudo-log-likelihood function can be performed by using Newton–Raphson or by bisection method also, but it is observed that they take larger number of iterations than the proposed one.

Now we can provide the EM algorithm as follows. Start with an initial guess of α, λ_0, λ_1 and λ_2 as $\alpha^{(0)}, \lambda_0^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}$ respectively.

- Obtain $u_1, u_2, v_1, v_2, \widetilde{n}_0$ and a_0 from $\alpha^{(0)}, \lambda_0^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}$.
- Maximize $l_{\text{pseudo}}(\alpha, \widehat{\lambda}_0(\alpha), \widehat{\lambda}_1(\alpha), \widehat{\lambda}_2(\alpha))$ and obtain $\alpha^{(1)}$.
- Obtain $\lambda_0^{(1)} = \widehat{\lambda}_0(\alpha^{(1)})$, $\lambda_1^{(1)} = \widehat{\lambda}_1(\alpha^{(1)})$, $\lambda_2^{(1)} = \widehat{\lambda}_2(\alpha^{(1)})$, using (20).
- Update $u_1, u_2, v_1, v_2, \widetilde{n}_0$ and a_0 from $\alpha^{(1)}, \lambda_0^{(1)}, \lambda_1^{(1)}, \lambda_2^{(1)}$.
- Continue the process until the convergence is met.

Comment: Note that the above EM algorithm can be easily used for the BBBE model. In that case since $\alpha = 1$, therefore Step 2 is not needed. From the ith step $(i + 1)$th step can be easily obtained from (20) using $\alpha = 1$.

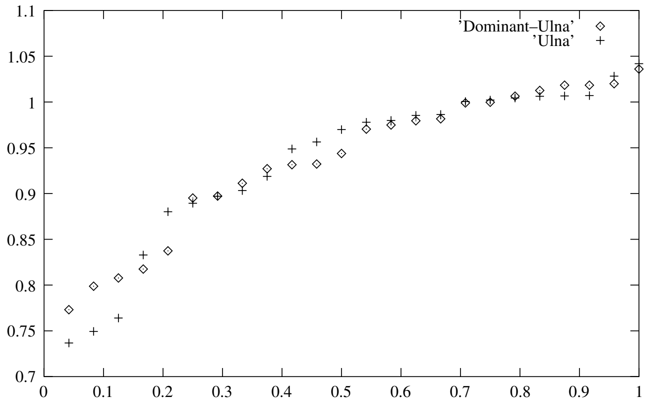

Fig. 3. The scaled TTT transform of the BMDs of Dominant Ulna and Ulna bones.

5. Data analysis

In this section we analyze one data set for illustrative purposes. This data set has been obtained from [6] (page 374). It represents the bone mineral density (BMD) measured in g/cm^2 for 24 children after one year of birth. These bivariate data represent the BMD for Dominant Ulna and Ulna bones and they are as follows;

Data set: (0.869 0.964), (0.602 0.689), (0.765 0.738), (0.761 0.698), (0.551 0.619), (0.753 0.515), (0.708 0.787), (0.687 0.715), (0.844 0.656), (0.869 0.789), (0.654 0.726), (0.692 0.526), (0.670 0.580), (0.823 0.773), (0.746 0.729), (0.656 0.506), (0.693 0.740), (0.883 0.785), (0.577 0.627), (0.802 0.769), (0.540 0.498), (0.804 0.779), (0.570 0.634), (0.585 0.640).

Just to get an idea about the hazard functions of the marginals we provide the scaled TTT plot as suggested by Aarset [1], which provides an idea of the shape of the hazard function of a distribution. For a family with survival function $S(y) = 1 - F(y)$, the scaled TTT transform with $H^{-1}(u) = \int_0^{F^{-1}(u)} S(y)\mathrm{d}y$ defined for $0 < u < 1$ is $g(u) = H^{-1}(u)/H^{-1}(1)$. The corresponding empirical version of the scaled TTT transform is given by $g_n(r/n) = H_n^{-1}(r/n)/H_n^{-1}(1) = \left[\sum_{i=1}^{r} y_{i:n} + (n-r)y_{r:n}\right]$, where $r = 1, \ldots, n$ and $y_{i:n},\ i = 1, \ldots, n$ represent the order statistics of the sample. It has been shown by Aarset [1] that the scaled TTT transform is convex (concave) if the hazard rate is decreasing (increasing) and for bathtub (unimodal) shaped hazard rate, the scaled TTT transform is first convex (concave) and then concave (convex). In this case the scale TTT transform of BMDs of Dominant radius and Radius bones are given in Fig. 3. It clearly indicates that both the marginals have increasing hazard rates. The correlation coefficient between the two variables is 0.628 and that shows they are positively correlated. We have fitted the Weibull distribution to the minimum of the two variables. The MLEs of the shape and scale parameters are 7.0795 and 11.8922 respectively. The Kolmogorov–Smirnov (K–S) distance between the empirical distribution function and the fitted distribution function is 0.141 and the corresponding p-value is 0.731. It implies that the Weibull distribution fits very well to the minimum of the two variables. Therefore it is not unreasonable to fit the BBBW model in this case.

We have used the proposed EM algorithm to compute the MLEs of the unknown parameters. We start the EM algorithm with the initial estimate of α as 7.00. We do not have immediate initial estimates of λ_0, λ_1 and λ_2, but we have an initial estimate of $\lambda_0 + \lambda_1 + \lambda_2$, *i.e.* 11.892. So we start the iterative process with the initial estimates of λ_0, λ_1 and λ_2 as 3.00 each. The EM algorithm provides the estimates of α, λ_0, λ_1 and λ_2 as 7.2547, 8.0262, 3.6399, 5.1833 respectively. Using the method proposed by Louis [10], the 95% confidence intervals of α, λ_0, λ_1 and λ_2 become (6.3381, 8.1712), (5.9827, 10.0698), (2.0628, 5.2170) and (3.2047, 7.1619) respectively.

Now we would like to see how good is the fit. We do not have any proper bivariate goodness of fit test like the univariate case. We examine the marginals and definitely it provide some indications about the goodness of fit. The empirical survival functions and the corresponding fitted survival functions for Dominant–Ulna and Ulna are reported in Fig. 4. The K–S distances between empirical distribution functions and the corresponding fitted distribution functions of the marginals are 0.141 and 0.109 and the corresponding p values are 0.730 and 0.936 respectively. Moreover, it has also been

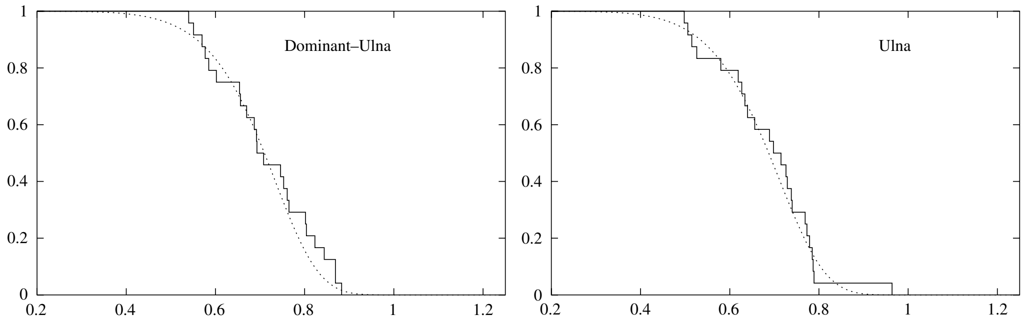

Fig. 4. The empirical survival function and fitted survival function for Dominant–Ulna and Ulna.

observed that the Weibull distribution fits the minimum also quite well. Considering all these points, we can say that BBBW may be used in this case for analyzing the data.

6. Conclusions

In this paper we have introduced a new absolutely continuous bivariate model following the approach of Block and Basu [3]. Block and Basu [3] obtained the BBBE model from the Marshall–Olkin bivariate exponential model by removing the singular component. Exactly in a similar manner we have obtained the BBBW model from the Marshall–Olkin bivariate Weibull model by removing the singular component. It has an absolutely continuous probability density function and we have studied several properties of this new distribution. This model has four unknown parameters and we have suggested an EM algorithm to compute the MLEs. It is observed that the proposed model work quite well for data analysis purposes.

Now we discuss some of the open problems. Note that although we have defined the model for the bivariate case but it is possible to define even for the multivariate case. It will be interesting to see how the estimation procedure can be generalized in this case. Moreover, Bayesian analysis of the proposed should be possible along the same line as Pena and Gupta [12]. More work is needed in these directions.

Acknowledgements

The authors would like to thank the referee and the associate editor for carefully reading the manuscript and made some valuable comments which had greatly improved the earlier draft of the manuscript.

The second author's work has been partly supported by a discovery grant from NSERC, CANADA.

Appendix A

Lemma 1. *The pseudo-log-likelihood function* $l_{\text{pseudo}}(\alpha, \widehat{\lambda}_0(\alpha), \widehat{\lambda}_1(\alpha), \widehat{\lambda}_2(\alpha)) = l_{\text{pseudo}}(\alpha)$ *(say), is unimodal.*

Proof. The pseudo-log-likelihood function without the additive constant can be written as

$$l_{\text{pseudo}}(\alpha) = (\widetilde{n}_0 + 2n)\ln\alpha + (\alpha - 1)\left[\widetilde{n}_0 \ln a_0 + \sum_{i\in I}\ln y_{1i} + \sum_{i\in I}\ln y_{2i}\right]$$
$$- (u_2 n_1 + v_1 n_2)\ln\left(\widetilde{n}_0 a_0^{\alpha} + \sum_{i\in I_2} y_{1i}^{\alpha} + \sum_{i\in I_1} y_{2i}^{\alpha}\right)$$
$$- (n_1 + v_1 n_2)\ln\left(\widetilde{n}_0 a_0^{\alpha} + \sum_{i\in I} y_{1i}^{\alpha}\right) - (n_2 + u_2 n_1)\ln\left(\widetilde{n}_0 a_0^{\alpha} + \sum_{i\in I} y_{2i}^{\alpha}\right). \quad (22)$$

First we prove that if $g(\alpha) = \sum_{i\in I} y_{1i}^{\alpha}$, then $\ln g(\alpha)$ is concave. From

$$g'(\alpha) = \sum_{i\in I} y_{1i}^{\alpha} \ln y_{1i}, \qquad g''(\alpha) = \sum_{i\in I} y_{1i}^{\alpha} (\ln y_{1i})^2,$$

and

$$\left(\sum_{i\in I} y_{1i}^{\alpha}(\ln y_{1i})^2\right) \times \left(\sum_{i\in I} y_{1i}^{\alpha}\right) - \left(\sum_{i\in I} y_{1i}^{\alpha}(\ln y_{1i})\right)^2 = \sum_{i<j} y_{1i}^{\alpha} y_{1j}^{\alpha} \left(y_{1i} - y_{1j}\right)^2 \geq 0,$$

it implies

$$g''(\alpha) g(\alpha) \geq (g'(\alpha))^2.$$

Therefore, $\ln g(\alpha)$ is concave. Similarly, it follows that

$$\ln\left(\widetilde{n}_0 a_0^{\alpha} + \sum_{i\in I_2} y_{1i}^{\alpha} + \sum_{i\in I_1} y_{2i}^{\alpha}\right), \ln\left(\widetilde{n}_0 a_0^{\alpha} + \sum_{i\in I} y_{ii}^{\alpha}\right) \ln\left(\widetilde{n}_0 a_0^{\alpha} + \sum_{i\in I} y_{2i}^{\alpha}\right)$$

are also concave functions. It implies $l_{\text{pseudo}}(\alpha)$ is a concave function. Now unimodality follows by observing that as α tends to 0 or ∞, $l_{\text{pseudo}}(\alpha)$ tends to $-\infty$. ■

Appendix B

Expected Fisher information matrix

Let the Fisher information matrix be

$$I = -E \begin{bmatrix} \frac{\partial^2 l}{\partial\alpha^2} & \frac{\partial^2 l}{\partial\alpha\partial\lambda_0} & \frac{\partial^2 l}{\partial\alpha\partial\lambda_1} & \frac{\partial^2 l}{\partial\alpha\partial\lambda_2} \\ \frac{\partial^2 l}{\partial\lambda_0\partial\alpha} & \frac{\partial^2 l}{\partial\lambda_0^2} & \frac{\partial^2 l}{\partial\lambda_0\partial\lambda_1} & \frac{\partial^2 l}{\partial\lambda_0\partial\lambda_2} \\ \frac{\partial^2 l}{\partial\lambda_1\partial\alpha} & \frac{\partial^2 l}{\partial\lambda_1\partial\lambda_0} & \frac{\partial^2 l}{\partial\lambda_1^2} & \frac{\partial^2 l}{\partial\lambda_1\partial\lambda_2} \\ \frac{\partial^2 l}{\partial\lambda_2\partial\alpha} & \frac{\partial^2 l}{\partial\lambda_2\partial\lambda_0} & \frac{\partial^2 l}{\partial\lambda_2\partial\lambda_1} & \frac{\partial^2 l}{\partial\lambda_2^2} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}. \tag{23}$$

Before, providing all the elements explicitly we introduce the following notations. If $Z \sim \text{WE}(\alpha, \lambda)$, then

$$E(Z^{\alpha} \ln Z^{\alpha}) = \frac{1}{\lambda}\left(\psi(2) - \ln(\lambda)\right) = \xi(\lambda) \quad \text{(Let)} \tag{24}$$

$$E(Z^{\alpha} (\ln Z^{\alpha})^2) = \frac{1}{\lambda}\left(\psi'(2) + (\psi(2) - \ln(\lambda))^2\right) = \eta(\lambda), \tag{25}$$

here $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and polygamma functions. We also need the following results:

$$E(n_1) = \frac{n\lambda_1}{\lambda_1 + \lambda_2}, \qquad E(n_2) = \frac{n\lambda_2}{\lambda_1 + \lambda_2}.$$

Then

$$a_{11} = \frac{n}{\alpha^2}\left[2 + c\lambda_1 \eta(\lambda_0 + \lambda_1) + c\lambda_2 \eta(\lambda_0 + \lambda_2)\right],$$

$$a_{22} = n \times \left[\frac{1}{(\lambda_0 + \lambda_1 + \lambda_2)^2} + \frac{\lambda_1}{(\lambda_1 + \lambda_2)(\lambda_0 + \lambda_2)^2} + \frac{\lambda_2}{(\lambda_1 + \lambda_2)(\lambda_0 + \lambda_1)^2}\right],$$

$$a_{33} = n \times \left[\frac{1}{(\lambda_0+\lambda_1+\lambda_2)^2} - \frac{\lambda_1}{(\lambda_1+\lambda_2)^3} + \frac{1}{\lambda_1(\lambda_1+\lambda_2)} + \frac{\lambda_2}{(\lambda_1+\lambda_2)(\lambda_0+\lambda_1)^2} \right],$$

$$a_{44} = n \times \left[\frac{1}{(\lambda_0+\lambda_1+\lambda_2)^2} - \frac{\lambda_2}{(\lambda_1+\lambda_2)^3} + \frac{\lambda_1}{(\lambda_1+\lambda_2)(\lambda_0+\lambda_2)^2} + \frac{1}{\lambda_2(\lambda_1+\lambda_2)} \right],$$

$$a_{12} = a_{21} = \frac{n}{\alpha}\xi(\lambda_0+\lambda_1+\lambda_2),$$

$$a_{13} = a_{31} = \frac{cn}{\alpha}\left[\xi(\lambda_0+\lambda_1) - \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}\xi(\lambda_0+\lambda_1+\lambda_2)\right],$$

$$a_{14} = a_{41} = \frac{cn}{\alpha}\left[\xi(\lambda_0+\lambda_2) - \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}\xi(\lambda_0+\lambda_1+\lambda_2)\right],$$

$$a_{23} = a_{32} = n \times \left[\frac{1}{(\lambda_0+\lambda_1+\lambda_2)^2} + \frac{\lambda_2}{(\lambda_1+\lambda_2)(\lambda_0+\lambda_1)^2} \right],$$

$$a_{24} = a_{42} = n \times \left[\frac{1}{(\lambda_0+\lambda_1+\lambda_2)^2} + \frac{\lambda_1}{(\lambda_1+\lambda_2)(\lambda_0+\lambda_2)^2} \right],$$

$$a_{34} = a_{43} = \frac{n}{(\lambda_1+\lambda_2)^2} - \frac{n}{(\lambda_0+\lambda_1+\lambda_2)^2}.$$

Observed Fisher information matrix

To compute the observed information matrix, we use the same notation as of Louis [10]. 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 $\mathbf{U}$.

$$\begin{aligned} S_{11} \;=\; & \frac{\widetilde{n}_0 + 2n_1 + 2n_2}{\widehat{\alpha}^2} + \widehat{\lambda}_0 \left[\sum_{i\in I_2} y_{1i}^{\widehat{\alpha}} (\ln y_{1i})^2 + \sum_{i \in I_1} y_{2i}^{\widehat{\alpha}} (\ln y_{2i})^2 + \widetilde{n}_0 a_0^{\widehat{\alpha}} (\ln a_0)^2 \right], \\ & + \widehat{\lambda}_1 \left[\sum_{i\in I_1\cup I_2} y_{1i}^{\widehat{\alpha}} (\ln y_{1i})^2 + \widetilde{n}_0 a_0^{\widehat{\alpha}} (\ln a_0)^2 \right] + \widehat{\lambda}_2 \left[\sum_{i\in I_1\cup I_2} y_{2i}^{\widehat{\alpha}} (\ln y_{2i})^2 + \widetilde{n}_0 a_0^{\widehat{\alpha}} (\ln a_0)^2 \right], \end{aligned}$$

$$S_{12} = S_{21} = \sum_{i\in I_2} y_{1i}^{\widehat{\alpha}} \ln y_{1i} + \sum_{i\in I_1} y_{2i}^{\widehat{\alpha}} \ln y_{2i} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \ln a_0,$$

$$S_{13} = S_{31} = \sum_{i\in I_1\cup I_2} y_{1i}^{\widehat{\alpha}} \ln y_{1i} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \ln a_0,$$

$$S_{14} = S_{41} = \sum_{i\in I_1\cup I_2} y_{2i}^{\widehat{\alpha}} \ln y_{2i} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \ln a_0,$$

$$S_{22} = \frac{\widetilde{n}_0 + u_1 n_1 + v_1 n_2}{\widehat{\lambda}_0^2}, \qquad S_{23} = S_{32} = 0,$$

$$S_{33} = \frac{n_1 + v_2 n_2}{\widehat{\lambda}_1^2}, \qquad S_{34} = S_{43} = 0, \qquad S_{44} = \frac{u_2 n_1 + n_2}{\widehat{\lambda}_2^2}.$$

$$\begin{aligned} U_1 \;=\; & \frac{\widetilde{n}_0 + 2n_1 + 2n_2}{\widehat{\alpha}} + \left[\sum_{i\in I_1\cup I_2} \ln y_{1i} + \sum_{i\in I_1\cup I_2} \ln y_{2i} + \widetilde{n}_0 \ln a_0 \right] \\ & - \widehat{\lambda}_1 \left[\sum_{i\in I_1\cup I_2} y_{1i}^{\widehat{\alpha}} \ln y_{1i} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \ln a_0 \right] - \widehat{\lambda}_2 \left[\sum_{i\in I_1\cup I_2} y_{2i}^{\widehat{\alpha}} \ln y_{2i} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \ln a_0 \right], \\ & - \widehat{\lambda}_0 \left[\sum_{i\in I_2} y_{1i}^{\widehat{\alpha}} \ln y_{1i} + \sum_{i\in I_1} y_{2i}^{\widehat{\alpha}} \ln y_{2i} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \ln a_0 \right], \end{aligned}$$

$$U_2 = \frac{\widetilde{n}_0 + u_1 n_1 + v_1 n_2}{\widehat{\lambda}_0} - \left[\sum_{i \in I_2} y_{1i}^{\widehat{\alpha}} + \sum_{i \in I_1} y_{2i}^{\widehat{\alpha}} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \right],$$

$$U_3 = \frac{n_1 + v_2 n_2}{\widehat{\lambda}_1} - \left[\sum_{i \in I_1 \cup I_2} y_{1i}^{\widehat{\alpha}} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \right],$$

$$U_4 = \frac{u_2 n_1 + n_2}{\widehat{\lambda}_2} - \left[\sum_{i \in I_1 \cup I_2} y_{2i}^{\widehat{\alpha}} + \widetilde{n}_0 a_0^{\widehat{\alpha}} \right].$$

References

[1] M.V. Aarset, How to identify a bathtub shaped hazard rate? IEEE Transactions on Reliability 36 (1987) 106–108.
[2] A.P. Basu, Bivariate hazard rate, Journal of the American Statistical Association 66 (1971) 103–104.
[3] H. Block, A.P. Basu, A continuous bivariate exponential extension, Journal of the American Statistical Association 69 (1974) 1031–1037.
[4] R. Harris, A multivariate defintion for increasing hazard rate distribution functions, The Annals of Mathematical Statistics 41 (1970) 713–717.
[5] N.L. Johnson, S. Kotz, A vector multivariate hazard rate, Journal of Multivariate Analysis 5 (1975) 53–66.
[6] R.A. Johnson, D.W. Wichern, Applied Multivariate Analysis, fourth ed., Prentice Hall, New Jersey, 1999.
[7] D. Karlis, ML estimation for multivarite shock models via an EM algorithm, Annals of the Institute of Statistical Mathematics 55 (2003) 817–830.
[8] D. Kundu, R.D. Gupta, Estimation of $P(Y < X)$ for Weibull distribution, IEEE Transactions on Reliability 55 (2006) 270–280.
[9] E.L. Lehmann, Some concepts of dependence, The Annals of Mathematical Statistics 37 (1966) 1137–1153.
[10] T.A. Louis, Finding the observed information matrix when using the EM algorithm, Journal of the Royal Statistical Society. Series B 44 (1982) 226–233.
[11] A.W. Marshall, I. Olkin, A multivariate exponential distribution, Journal of the American Statistical Association 62 (1967) 30–44.
[12] E.A. Peña, A.K. Gupta, Bayes estimation for the Marshall–Olkin exponential distribution, Journal of the Royal Statistical Society. Series B 52 (1990) 379–389.