Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 2, 275–302
DOI: 10.1214/16-BJPS313

On bivariate inverse Weibull distribution

Debasis Kundu[a] **and Arjun K. Gupta**[b]

[a]*Indian Institute of Technology Kanpur*
[b]*Bowling Green State University*

Abstract. Inverse Weibull distribution has been used quite successfully to analyze lifetime data which has non-monotone hazard function. The main aim of this paper is to introduce bivariate inverse Weibull distribution along the same line as the Marshall–Olkin bivariate exponential distribution, so that the marginals have inverse Weibull distributions. The proposed bivariate inverse Weibull distribution has four parameters and it has a singular component. Therefore, it can be used quite successfully if there are ties in the data. The joint probability density function, the joint cumulative distribution function and the joint survival function are all in closed forms. Several properties of this distribution have been discussed. It is observed that the proposed distribution can be obtained from the Marshall–Olkin copula. The maximum likelihood estimators of the unknown parameters cannot be obtained in closed form, and we propose to use EM algorithm to compute the maximum likelihood estimators. We propose to use parametric bootstrap method for construction of confidence intervals of the different parameters. We present some simulation experiments results to show the performances of the EM algorithm and they are quite satisfactory. We provide the Bayesian analysis of the unknown parameters based on very flexible priors. We analyze one bivariate American Football League data set for illustrative purposes, and it is observed that this model provides a slightly better fit than some of the existing models. Finally, we present some generalization to the multivariate case.

1 Introduction

Two-parameter Weibull distribution has been used quite successfully to analyze lifetime data. Due to the presence of two parameters, Weibull distribution is a very flexible lifetime distribution. It can have a decreasing or an unimodal probability density function (PDF). Moreover, depending on the shape parameter, it can have increasing, decreasing or constant hazard functions. Extensive work has been done on the Weibull distribution both from the frequentist and Bayesian points of view. See, for example, an excellent review by Johnson et al. (1994) or Kundu (2008) for some related references. Marshall and Olkin (1967) proposed a bivariate extension of the exponential distribution, whose marginals are Weibull distributions. From now on, we call it as the Marshall–Olkin bivariate Weibull (MOBW) distribution.

Key words and phrases. Marshall–Olkin bivariate exponential distribution, maximum likelihood estimator, failure rate, EM algorithm, Fisher information matrix.

Received April 2014; accepted February 2016.

MOBW is more flexible than the Marshall–Olkin bivariate exponential (MOBE) model, see Kundu and Gupta (2013), and it also can be used as a shock model similarly as the MOBE model.

Although, univariate Weibull distribution has been used quite extensively to analyze lifetime data, it may not be proper to use if the data indicate a non-monotone, for example, a unimodal hazard function. In many practical situations, it is known apriori that the hazard function cannot be monotone. For example, in mortality study, often it is known that the mortality reaches a peak after some finite period, and then declines slowly. Similarly, in a breast cancer study it is observed that the peak mortality occurs usually after three years of surgery, and then it gradually decreases. If the empirical study indicates that the hazard function might be unimodal, then the inverse Weibull (IW) distribution may be used to analyze such data set, see, for example, Nelson (1982). It can also be used as a heavy tail distribution.

The main aim of this paper is to introduce bivariate inverse Weibull (BIW) distribution, so that it has IW marginals. The proposed BIW distribution has four parameters. Due to the presence of four parameters, it becomes a very flexible model. The joint PDF can take different shapes. The joint PDF, joint CDF and joint survival function all are in closed forms, which make it very convenient to use it in practice for analyzing censored data also. There are several reasons to consider this specific bivariate distribution. It may be mentioned that several absolute continuous bivariate distributions are available in the literature, see, for example, Balakishnan and Lai (2009) for a detailed account of such distributions till that time, and see Aleem (2012), Myrhaug and Leira (2011), Teugels (2014), Yang et al. (2009) for some recent references. But other than the MOBE, MOBW or bivariate generalized exponential distribution of Kundu and Gupta (2009), not too many bivariate distributions are available in the literature, at least not known to the authors, with a singular component. In many practical applications ties between two components may occur quite naturally. Therefore, it may not be reasonable to analyze those data sets using any absolute continuous bivariate distribution. Moreover, all the existing bivariate distributions with a singular component have the marginals either with constant or with monotone hazard functions. The proposed bivariate distribution has the marginals with non-monotone hazard functions. Therefore, it will give the practitioner one more choice from the class of possible bivariate distributions with a singular component to analyze a bivariate data set with ties. Moreover, the proposed bivariate distribution has some interesting physical interpretations also.

The generation of random samples from the BIW distribution can be performed very easily, hence simulation experiments can be performed quite conveniently. The joint CDF of BIW has a singular component and an absolute continuous component. Due to the presence of the singular component, this distribution can be used quite naturally when there are ties in the data. We further study different properties of the BIW distribution. It is observed that the proposed BIW distribution can be obtained from the Marshall–Olkin copula. Therefore, several dependency measures and dependency properties can be easily established for this model using the copula structure.

The maximum likelihood estimators (MLEs) of the unknown parameters cannot be obtained in closed form. They have to obtained by solving four non-linear equations simultaneously. Standard methods like Newton–Raphson or downhill simplex may be used, but it usually takes a long time to converge. We propose to use EM algorithm as originally suggested by Dempster et al. (1977), to compute the MLEs of the unknown parameters. At each E-step of the EM algorithm the corresponding M-step can be performed by solving only a one dimensional optimization problem. Hence, the implementation of the EM algorithm is quite straight forward in practice. A FORTRAN code has been provided for this purpose. Parametric bootstrap method has been used for constructing confidence intervals of the unknown parameters. We present some simulation results to show the performances of the proposed EM algorithm, and they are quite satisfactory. One bivariate American Football League data set has been analyzed for illustrative purposes. It is observed that the proposed BIW model provides a slightly better fit than some of the existing models.

We further consider the Bayesian inference of the unknown parameters. For the fixed shape parameter, it is assumed that the scale parameters have Dirichlet-Gamma prior, and for the shape parameter no fixed prior distribution is assumed. It is assumed that the support of the shape parameter is on the entire positive real line and its probability density function (PDF) is log-concave. It may be mentioned that several well known life time distributions have log-concave PDFs. Based on the above priors, the posterior distribution of the unknown parameters are obtained. The Bayes estimates cannot be obtained in closed form. Although, the Lindley's approximation may be used to compute the approximate Bayes estimates, it is not followed here. Instead, we use importance sampling technique to compute the approximate Bayes estimates based on the squared error loss function, and also obtain the associated highest posterior density (HPD) credible intervals. Simulation results indicate that the performances of the Bayes estimates are quite satisfactory. Finally, we provide a multivariate generalization of the proposed model, and discuss some of its properties.

Rest of the paper is organized as follows. In Section 2, we introduce BIW model, and discuss its properties in Section 3. The EM algorithm is provided in Section 4. In Section 5, we provide the Bayesian inference of the unknown parameters. Simulation results and the analysis of a data set have been provided in Section 6. We provide a generalization to the multivariate case in Section 7. Finally, conclusions and some open problems appear in Section 8. All the proofs are provided in the Appendices.

2 Bivariate inverse Weibull distribution

In this section, we introduce the BIW distribution and provide some physical interpretations of the proposed model. We further provide the explicit expressions of the joint PDF of the absolute continuous part and the singular part. Moreover, we also provide the shapes of the absolute continuous part of the joint PDF for different parameter values.

If the random variable Y has a Weibull distribution with the PDF

$$f_{WE}(y;\alpha,\lambda) = \alpha\lambda y^{\alpha-1}e^{-\lambda y^{\alpha}}; \qquad y > 0,$$

then the random variable $X = 1/Y$ has an inverse Weibull (IW) distribution with the PDF

$$f_{\mathrm{IW}}(x;\alpha,\lambda) = \alpha\lambda x^{-(\alpha+1)}e^{-\lambda x^{-\alpha}}; \qquad x > 0. \tag{1}$$

Here $\alpha > 0$ and $\lambda > 0$, are the shape and scale parameters, respectively. A random variable with the PDF (1) will be denoted by IW(α,λ). If X follows $(\sim)$ IW(α,λ), then the CDF of X becomes

$$P(X \leq x) = F_X(x;\alpha,\lambda) = e^{-\lambda x^{-\alpha}}; \qquad x > 0.$$

From now on unless otherwise mentioned, it is assumed that $\alpha > 0$, $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_3 > 0$ and $\Theta = (\alpha, \lambda_1, \lambda_2, \lambda_3)$.

Suppose $U_1 \sim$ IW(α,λ_1), $U_2 \sim$ IW(α,λ_2), $U_3 \sim$ IW(α,λ_3), and they are independently distributed. If $X_1 = \max\{U_1, U_3\}$ and $X_2 = \max\{U_2, U_3\}$, then (X_1, X_2) is said to have a bivariate inverse Weibull distribution with parameters α, λ_1, λ_2 and λ_3, and it will be denoted by BIW$(\alpha,\lambda_1,\lambda_2,\lambda_3)$. When $\alpha = 1$, it will be called the bivariate inverse exponential distribution with parameters λ_1, λ_2 and λ_3. BIW can be used as a stress model or as a maintenance model, as follows.

Stress Model. Suppose a system has two components, and each component is subjected to individual independent stress say U_1 and U_2, respectively. The system has an overall stress U_3 which has been transmitted to both the components equally, and it is independent of the individual stresses. Therefore, the observed stresses at the two components are $X_1 = \max\{U_1, U_3\}$ and $X_2 = \max\{U_2, U_3\}$, respectively.

Maintenance. Suppose a system has two components, and each component has been maintained independently and there is an overall maintenance also. Due to individual component maintenance, suppose the lifetime of the individual component is increased by the amount U_i for $i = 1, 2$, and for the overall maintenance, the lifetime of each item is increased by the amount U_3. Therefore, the increased lifetimes of the two components are $X_1 = \max\{U_1, U_3\}$ and $X_2 = \max\{U_2, U_3\}$, respectively.

The following result will provide the joint CDF of X_1 and X_2.

Theorem 2.1. *If* $(X_1, X_2) \sim \mathrm{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, *then the joint CDF of* (X_1, X_2) *for* $x_1 > 0$ *and* $x_2 > 0$ *is*

$$F_{X_1,X_2}(x_1, x_2) = \begin{cases} e^{-(\lambda_1+\lambda_3)x_1^{-\alpha}-\lambda_2 x_2^{-\alpha}} & \text{if } x_1 < x_2 \\ e^{-\lambda_1 x_1^{-\alpha}-(\lambda_2+\lambda_3)x_2^{-\alpha}} & \text{if } x_1 > x_2 \\ e^{-(\lambda_1+\lambda_2+\lambda_3)x^{-\alpha}} & \text{if } x_1 = x_2 = x \end{cases}$$

Proof. It is trivial and hence it is omitted. □

We have the following unique decomposition of the joint CDF of (X_1, X_2).

Theorem 2.2. *If* $(X_1, X_2) \sim \mathrm{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, *then*

$$F_{X_1,X_2}(x_1, x_2) = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} F_a(x_1, x_2) + \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} F_s(x_1, x_2),$$

where for $z = x_1 \wedge x_2 = \min\{x_1, x_2\}$.

$$F_s(x_1, x_2) = e^{-(\lambda_1+\lambda_2+\lambda_3)z^{-\alpha}},$$

and

$$F_a(x_1, x_2) = \frac{\lambda_1 + \lambda_2 + \lambda_3}{\lambda_1 + \lambda_2} e^{-\lambda_1 x_1^{-\alpha}-\lambda_2 x_2^{-\alpha}-\lambda_3 z^{-\alpha}} - \frac{\lambda_3}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda-2+\lambda_3)z^{-\alpha}}.$$

Here $F_s(\cdot, \cdot)$ *and* $F_a(\cdot, \cdot)$ *are the singular and absolute continuous parts, respectively.*

Proof. See in the Appendix A. □

It is immediate that $P(X_1 \le x_1, X_2 \le x_2|A)$ is the singular part, as its mixed partial second derivative is 0, and $P(X_1 \le x_1, X_2 \le x_2|A^c)$ is the absolute continuous part, as its mixed second partial derivative is a density function. Hence, the joint PDF of (X_1, X_2) can be written in the following form.

$$f_{X_1,X_2}(x_1, x_2) = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} f_a(x_1, x_2) + \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} f_s(z),$$

where

$$f_a(x_1, x_2) = \frac{\lambda_1 + \lambda_2 + \lambda_3}{\lambda_1 + \lambda_2} \times \begin{cases} f_{\mathrm{IW}}(x_1; \alpha, \lambda_1 + \lambda_3) f_{\mathrm{IW}}(x_2; \alpha, \lambda_2) & \text{if } x_1 < x_2, \\ f_{\mathrm{IW}}(x_1; \alpha, \lambda_1) f_{\mathrm{IW}}(x_2; \alpha, \lambda_2 + \lambda_3) & \text{if } x_1 > x_2 \end{cases}$$

and

$$f_s(x) = f_{\mathrm{IW}}(x; \lambda_1 + \lambda_2 + \lambda_3).$$

In this case $f_a(x_1, x_2)$ and $f_s(x)$ are the absolute continuous part and singular part, respectively.

It should be noted that when the function $f_{X_1,X_2}(x_1, x_2)$ is mentioned to be a joint PDF of (X_1, X_2), it is understood that the first term $f_a(x_1, x_2)$ is the joint PDF with respect to two dimensional Lebesgue measure and the second term is a PDF with respect to one dimensional Lebesgue measure.

The following result will provide the shape of $f_a(x_1, x_2)$.

Theorem 2.3. *Let* $(X_1, X_2) \sim \text{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$.

(a) *If* $\lambda_1 = \lambda_2 = \lambda$, *then* $f_a(x_1, x_2)$ *is continuous for* $0 < x_1, x_2 < \infty$. $f_a(x_1, x_2)$ *is unimodal, and the mode is at* (x_m, x_m), *where* $x_m = \{\alpha(2\lambda + \lambda_3)/(2(\alpha + 1))\}^{1/\alpha}$.
(b) *If* $\lambda_1 + \lambda_3 < \lambda_2$, *then* $f_a(x_1, x_2)$ *is not continuous on* $x_1 = x_2$. $f_a(x_1, x_2)$ *is unimodal, and the mode occurs at* (x_{1m}, x_{2m}), *where* $x_{1m} = \{\alpha(\lambda_1 + \lambda_3)/(\alpha + 1)\}^{1/\alpha}$ *and* $x_{2m} = \{\alpha\lambda_2/(\alpha + 1)\}^{1/\alpha}$.
(c) *If* $\lambda_2 + \lambda_3 < \lambda_1$, $f_a(x_1, x_2)$, *then* $f_a(x_1, x_2)$ *is not continuous on* $x_1 = x_2$. $f_a(x_1, x_2)$ *is unimodal, and the mode occurs at* (x_{1m}, x_{2m}), *where* $x_{1m} = \{\alpha\lambda_1/(\alpha + 1)\}^{1/\alpha}$ *and* $x_{2m} = \{\alpha(\lambda_2 + \lambda_3)(\alpha + 1)\}^{1/\alpha}$.

Proof. See in the Appendix A. □

In Figure 1, we provide the surface plots of the absolute continuous part of the BIW distribution function for different choices of α, λ_1, λ_2 and λ_3. It shows the unimodality of the PDF for different choices of the parameter values. Note that it is very simple to generate samples from a BIW distribution. The following simple procedure can be used to generate sample (x_1, x_2) from a BIW$(\alpha, \lambda_1, \lambda_2, \lambda_3)$. Step 1: Generate v_1, v_2 and v_3 independently from a uniform (0, 1). Step 2: $u_1 = (-\ln v/\lambda_1)^{-1/\alpha}$, $u_2 = (-\ln v/\lambda_2)^{-1/\alpha}$, $u_3 = (-\ln v/\lambda_3)^{-1/\alpha}$. Step 3: $x_1 = \max\{u_1, u_3\}$ and $x_2 = \max\{u_2, u_3\}$.

3 Different properties

The main purpose of this section is to provide some basic properties of the BIW distribution. We provide some dependency properties of the bivariate distribution and provide the copula structure, which can be used to provide different dependency measures of the two components. It has its own theoretical interest or it can be used for other purposes also.

3.1 Marginals, conditionals and dependence

The following result provides the distributions of the marginals and the maximum and the stress-strength measure of the two components of a BIW distribution.

Theorem 3.1. *Let* $(X_1, X_2) \sim \text{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, *then*

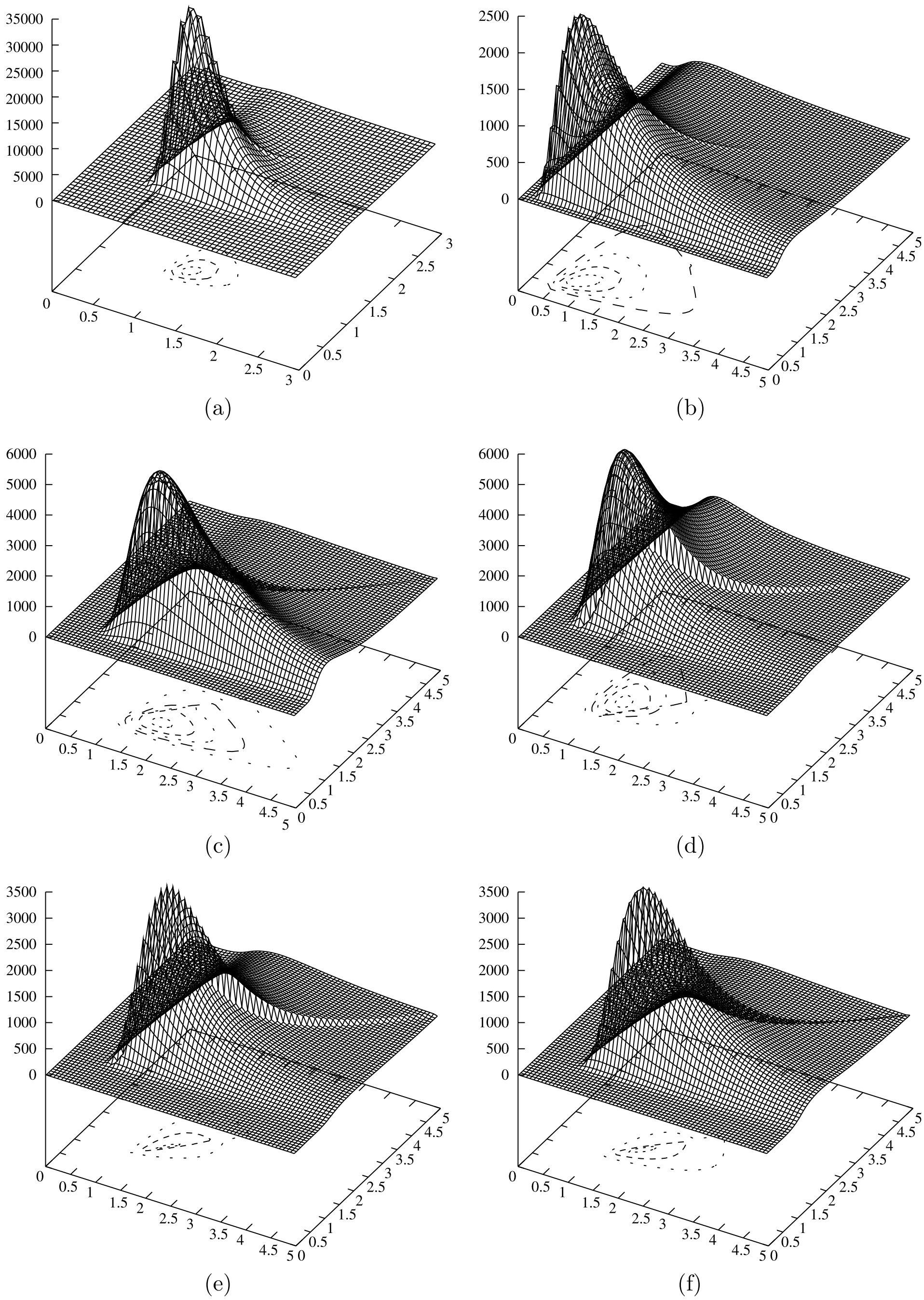

Figure 1 *The PDF of the absolute continuous part of BIW distribution for different parameter values of* α, λ_1, λ_2 *and* λ_3: (a) (4, 1, 1, 1); (b) (1, 1, 1, 1); (c) (2, 4, 1, 1); (d) (2, 1, 4, 1); (e) (2, 1, 2, 4); (f) (2, 2, 1, 4).

(a) $X_1 \sim \mathrm{IW}(\alpha, \lambda_1 + \lambda_3)$ *and* $X_2 \sim \mathrm{IW}(\alpha, \lambda_2 + \lambda_3)$
(b) $\max\{X_1, X_2\} \sim \mathrm{IW}(\alpha, \lambda_1 + \lambda_2 + \lambda_3)$.
(c) $P(X_1 < X_2) = \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}$.

Proof. See in the Appendix B. □

The following result provides the conditional results of BIW distribution.

Theorem 3.2. *Let* $(X_1, X_2) \sim \mathrm{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, *then*

(a) *the conditional distribution of* X_1 *given* $X_2 = x_2$, *say* $F_{X_1|X_2=x_2}(x_1)$ *is a convex combination of an absolute continuous distribution function and a degenerate distribution function as follows.*

$$F_{X_1|X_2=x_2}(x_1) = pG(x_1) + (1-p)H(x_1),$$

where

$$G(x_1) = \frac{1}{p} \times \begin{cases} \dfrac{\lambda_2}{\lambda_2+\lambda_3} e^{-(\lambda_1+\lambda_3)x_1^{-\alpha}+\lambda_3 x_2^{-\alpha}} & \text{if } x_1 < x_2, \\ e^{-\lambda_1 x_1^{-\alpha}} - \dfrac{\lambda_3}{\lambda_2+\lambda_3} e^{-\lambda_1 x_2^{-\alpha}} & \text{if } x_1 > x_2, \end{cases}$$

$$H(x_1) = \begin{cases} 0 & \text{if } x_1 < x_2, \\ 1 & \text{if } x_1 \geq x_2 \end{cases}$$

and

$$p = 1 - \frac{\lambda_3}{\lambda_2+\lambda_3} e^{-\lambda_1 x_2^{-\alpha}}.$$

(b) *the conditional distribution function of* X_1 *given* $X_2 \leq x_2$, *say* $F_{X_1|X_2 \leq x_2}(x_1)$, *is an absolute continuous distribution function as follows*;

$$\begin{aligned} P(X_1 \leq x_1 | X_2 \leq x_2) &= F_{X_1|X_2 \leq x_2}(x_1) \\ &= \begin{cases} e^{-(\lambda_1+\lambda_3)x_1^{-\alpha}+\lambda_3 x_2^{-\alpha}} & \text{if } x_1 \leq x_2, \\ e^{-\lambda_1 x_1^{-\alpha}} & \text{if } x_1 > x_2. \end{cases} \end{aligned}$$

Proof. The proofs can be obtained in a routine manner, hence they are avoided. □

Theorem 3.3. *Let* $(X_1, X_2) \sim \mathrm{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, *then* (X_1, X_2) *is*

(a) *PLOD, positively lower orthant dependent.*
(b) *LTD, left tail decreasing.*
(c) *LCSD, left corner set decreasing.*

Proof. See in the Appendix B. □

3.2 Copula representation

Every bivariate distribution, $F_{X_1,X_2}(x_1, x_2)$, with continuous marginals distribution functions $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$, corresponds a unique function $C : [0, 1]^2 \to [0, 1]$ called a copula such that for $(x_1, x_2) \in (-\infty, \infty) \times (-\infty, \infty)$,

$$F_{X_1,X_2}(x_1, x_2) = C\big(F_{X_1}(x_1), F_{X_2}(x_2)\big),$$

see Nelsen (2006) for more details. If $(X_1, X_2) \sim \text{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, then the corresponding copula function for $0 < u_1, u_2 < 1$ and for $u = \min\{u_1, u_2\}$, becomes

$$C(u_1, u_2) = \begin{cases} u_1^{1-\beta_1} u_2 & \text{if } u_1^{\beta_1} \geq u_2^{\beta_2}, \\ u_1 u_2^{1-\beta_2} & \text{if } u_1^{\beta_1} < u_2^{\beta_2}, \end{cases} \tag{2}$$

where $\beta_1 = \frac{\lambda_3}{\lambda_1+\lambda_3}$ and $\beta_2 = \frac{\lambda_3}{\lambda_2+\lambda_3}$. The copula (2) is the well-known Marshall–Olkin copula, see, for example, Nelsen (2006). Therefore, it easily follows that for a $\text{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$ distribution, Kendall's τ and Spearman's ρ become $\frac{\beta_1\beta_2}{\beta_1-\beta_1\beta_2+\beta_2}$ and $\frac{3\beta_1\beta_2}{2\beta_1-\beta_1\beta_2+2\beta_2}$, respectively. Using the copula structure, different other dependence properties and dependence measures of the $\text{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$ can be easily obtained.

4 Maximum likelihood estimation

In this subsection, we discuss the maximum likelihood estimation procedures of the unknown parameters of BIW distribution based on a random sample of size n. It is observed that to compute the MLEs of the unknown parameters, one needs to solve a four dimensional optimization problem. To avoid that, we propose to use EM algorithm which involves solving only a one-dimensional problem at each 'E-step', hence it can be implemented very conveniently.

The problem can be formulated as follows. Suppose $\mathcal{D} = \{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\}$ is a random sample from $\text{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, the problem is to find the MLEs of the unknown parameters. We use the following notations

$$I_1 = \{i : x_{1i} < x_{2i}\}, \qquad I_2 = \{i : x_{2i} > x_{2i}\},$$
$$I_0 = \{i : x_{1i} = x_{2i} = x_i\}, \qquad I = I_1 \cup I_2 \cup I_3$$
$$|I_1| = n_1, \qquad |I_2| = n_2, \qquad |I_0| = n_0, \quad \text{and} \quad n = n_0 + n_1 + n_2.$$

Based on the observations, the log-likelihood function can be written as follows:

$$\begin{aligned} &l(\alpha, \lambda_1, \lambda_2, \lambda_3|\mathcal{D}) \\ &\quad = (2n_1 + 2n_2 + n_0)\ln\alpha + n_1\ln(\lambda_1 + \lambda_3) + n_1\ln\lambda_2 + n_2\ln\lambda_1 \\ &\qquad + n_2\ln(\lambda_2 + \lambda_3) + n_0\ln\lambda_3 - \lambda_1\Bigg(\sum_{i\in I_1\cup I_2} x_{1i}^{-\alpha} + \sum_{i\in I_0} x_i^{-\alpha}\Bigg) \end{aligned} \tag{3}$$

$$-\lambda_2\Big(\sum_{i\in I_1\cup I_2} x_{2i}^{-\alpha} + \sum_{i\in I_0} x_i^{-\alpha}\Big)$$
$$-\lambda_3\Big(\sum_{i\in I_1} x_{1i}^{-\alpha} + \sum_{i\in I_2} x_{2i}^{-\alpha} + \sum_{i\in I_0} x_i^{-\alpha}\Big).$$

It is clear from (3) that the MLEs of α, λ_1, λ_2 and λ_3 can be obtained by solving four non-linear equations. We propose to use EM algorithm to avoid solving four dimensional optimization problem. It is observed that to implement the EM algorithm, at each E-step, the corresponding M-step can be performed by solving one one-dimensional optimization problem. Hence, it saves computational burden significantly.

We treat this as a missing value problem. It is assumed that for a bivariate random vector (X_1, X_2), there is an associated random vector (Δ_1, Δ_2), defined as follows

$$\Delta_1 = \begin{cases} 1 & \text{if } U_1 > U_3, \\ 3 & \text{if } U_1 < U_3, \end{cases} \quad \text{and} \quad \Delta_2 = \begin{cases} 2 & \text{if } U_2 > U_3, \\ 3 & \text{if } U_2 < U_3. \end{cases}$$

It can be easily seen that if we had a sample of size n from $(X_1, X_2, \Delta_1, \Delta_2)$, then the MLEs of unknown parameters can be obtained by solving a one non-linear equation. That is the main motivation of the proposed EM algorithm. It is immediate that when $X_1 = X_2$, then $\Delta_1 = \Delta_2 = 3$, but if $X_1 < X_2$ or $X_1 > X_2$, the corresponding (Δ_1, Δ_2) is missing. If $(x_1, x_2) \in I_1$, then the possible values of (Δ_1, Δ_2) are $(1, 2)$ or $(3, 2)$, respectively. Similarly, if $(x_1, x_2) \in I_2$, then the possible values of (Δ_1, Δ_2) $(1, 3)$ or $(1, 2)$. We need the following result for further developments, and they can be obtained very easily. If U_1, U_2 and U_3 are three random variables same as defined in Section 2, then

$$\{X_1 < X_2\} = \{U_1 < U_3 < U_2\} \cup \{U_3 < U_1 < U_2\},$$
$$\{X_2 < X_1\} = \{U_2 < U_3 < U_1\} \cup \{U_3 < U_2 < U_1\},$$
$$P(U_3 < U_1 < U_2) = \frac{\lambda_1\lambda_2}{(\lambda_1+\lambda_3)(\lambda_1+\lambda_2+\lambda_3)},$$
$$P(U_1 < U_3 < U_2) = \frac{\lambda_2\lambda_3}{(\lambda_1+\lambda_3)(\lambda_1+\lambda_2+\lambda_3)},$$
$$P(U_3 < U_2 < U_1) = \frac{\lambda_1\lambda_2}{(\lambda_2+\lambda_3)(\lambda_1+\lambda_2+\lambda_3)},$$
$$P(U_2 < U_3 < U_1) = \frac{\lambda_1\lambda_3}{(\lambda_2+\lambda_3)(\lambda_1+\lambda_2+\lambda_3)},$$
$$P(U_3 < U_1 | X_1 < X_2) = \frac{\lambda_1}{\lambda_1+\lambda_3},$$

$$P(U_1 < U_3 | X_1 < X_2) = \frac{\lambda_3}{\lambda_1 + \lambda_3},$$
$$P(U_3 < U_2 | X_2 < X_1) = \frac{\lambda_2}{\lambda_2 + \lambda_3},$$
$$P(U_2 < U_3 | X_2 < X_1) = \frac{\lambda_3}{\lambda_2 + \lambda_3}.$$

Now we provide the EM algorithm. In the E-Step, we treat the observations belonging to I_0 as the complete observations. An observation (x_1, x_2) is treated as missing if $(x_1, x_2) \in I_1 \cup I_2$. If the observation $(x_1, x_2) \in I_1$, we form the 'pseudo observation' by fractioning (x_1, x_2) to two partially complete 'pseudo observation' of the form $(x_1, x_2, u_1(\Theta))$ and $(x_1, x_2, u_2(\Theta))$, similarly as in Dinse (1982). Here

$$u_1(\Theta) = P(\Delta_1 = 1, \Delta_2 = 2 | X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_3},$$
$$u_2(\Theta) = P(\Delta_1 = 3, \Delta_2 = 2 | X_1 < X_2) = \frac{\lambda_3}{\lambda_1 + \lambda_3}.$$

Similarly, if $(x_1, x_2) \in I_2$, we form the 'pseudo observation' $(x_1, x_2, v_1(\Theta))$ and $(x_1, x_2, v_2(\Theta))$. Here

$$v_1(\Theta) = P(\Delta_1 = 2, \Delta_2 = 1 | X_2 < X_1) = \frac{\lambda_2}{\lambda_2 + \lambda_3},$$
$$v_2(\Theta) = P(\Delta_1 = 3, \Delta_2 = 1 | X_2 < X_1) = \frac{\lambda_3}{\lambda_2 + \lambda_3}.$$

From now on, for brevity we write $u_1(\Theta)$, $u_2(\Theta)$, $v_1(\Theta)$ and $v_2(\Theta)$ as u_1, u_2, v_1, v_2, respectively. Based on the above notations, the log-likelihood function of the 'pseudo data' is

$$\begin{aligned} l_{\text{pseudo}}(\Theta|\mathcal{D}) = {} & (n_0 + 2n_1 + 2n_2) \ln \alpha \\ & - (\alpha + 1)\Bigg(\sum_{i \in I_0} \ln x_i + \sum_{i \in I_1 \cup I_2} \ln x_{1i} + \sum_{i \in I_1 \cup I_2} \ln x_{2i}\Bigg) \\ & + (u_1 n_1 + n_2) \ln \lambda_1 - \lambda_1 \Bigg(\sum_{i \in I_0} x_i^{-\alpha} + \sum_{i \in I_1 \cup I_2} x_{1i}^{-\alpha}\Bigg) \\ & + (n_1 + v_1 n_2) \ln \lambda_2 \\ & - \lambda_2 \Bigg(\sum_{i \in I_0} x_i^{-\alpha} + \sum_{i \in I_1 \cup I_2} x_{2i}^{-\alpha}\Bigg) + (n_0 + u_2 n_1 + v_2 n_2) \ln \lambda_3 \\ & - \lambda_3 \Bigg(\sum_{i \in I_0} x_i^{-\alpha} + \sum_{i \in I_1} x_{1i}^{-\alpha} + \sum_{i \in I_2} x_{2i}^{-\alpha}\Bigg). \end{aligned} \tag{4}$$

Now the M-step involves maximizing (4) with respect to α, λ_1, λ_2 and λ_3. For fixed α the maximum with respect to λ_1, λ_2 and λ_3 occur at

$$\begin{aligned}\widehat{\lambda}_1(\alpha) &= \frac{u_1 n_1 + n_2}{\sum_{i\in I_0} x_i^{-\alpha} + \sum_{i\in I_1\cup I_2} x_{1i}^{-\alpha}},\\ \widehat{\lambda}_2(\alpha) &= \frac{n_1 + v_1 n_2}{\sum_{i\in I_0} x_i^{-\alpha} + \sum_{i\in I_1\cup I_2} x_{2i}^{-\alpha}},\end{aligned} \tag{5}$$

$$\widehat{\lambda}_3(\alpha) = \frac{n_0 + u_2 n_1 + v_2 n_2}{\sum_{i\in I_0} x_i^{-\alpha} + \sum_{i\in I_1} x_{1i}^{-\alpha} + \sum_{i\in I_2} x_{2i}^{-\alpha}}. \tag{6}$$

If $\widehat{\alpha}$ maximizes $l_{\text{pseudo}}(\Theta)$, then $\widehat{\alpha}$ can be obtained by maximizing the profile 'pseudo' log-likelihood function $l_{\text{pseudo}}(\alpha, \widehat{\lambda}_1(\alpha), \widehat{\lambda}_2(\alpha), \widehat{\lambda}_3(\alpha)) = c + g(\alpha)$, where

$$\begin{aligned}g(\alpha) = {} & (n_0 + 2n_1 + 2n_2)\ln\alpha - (\alpha+1)\Bigg(\sum_{i\in I_0} \ln x_i + \sum_{i\in I_1\cup I_2} \ln x_{1i} + \sum_{i\in I_1\cup I_2} \ln x_{2i}\Bigg)\\ & - (u_1 n_1 + n_2)\ln\Bigg(\sum_{i\in I_0} x_i^{-\alpha} + \sum_{i\in I_1\cup I_2} x_{1i}^{-\alpha}\Bigg)\\ & - (n_1 + v_2 n_2)\ln\Bigg(\sum_{i\in I_0} x_i^{-\alpha} + \sum_{i\in I_1\cup I_2} x_{2i}^{-\alpha}\Bigg)\\ & - (n_0 + u_2 n_1 + v_2 n_2)\ln\Bigg(\sum_{i\in I_0} x_i^{-\alpha} + \sum_{i\in I_1} x_{1i}^{-\alpha} + \sum_{i\in I_2} x_{2i}^{-\alpha}\Bigg),\end{aligned} \tag{7}$$

and c is independent of α. The following result indicates that $g(\alpha)$ has a unique maximum.

Theorem 4.1. *$g(\alpha)$ is a unimodal function.*

Proof. See in the Appendix C. □

Since, $g(\alpha)$ is a unimodal function, it is very easy to obtain $\widehat{\alpha}$, which maximizes (7), by using by-section or Newton–Raphson method. We propose the following algorithm to compute $(k+1)$th step from the kth step of the EM algorithm. At the kth step the estimates of α, λ_1, λ_2 and λ_3 will be denoted by $\alpha^{(k)}$, $\lambda_1^{(k)}$, $\lambda_2^{(k)}$ and $\lambda_3^{(k)}$.

EM Algorithm.

- Step 1: Compute u_1, u_2, v_1, v_2 using $\alpha^{(k)}$, $\lambda_1^{(k)}$, $\lambda_2^{(k)}$ and $\lambda_3^{(k)}$.
- Step 2: Maximize (7), and obtain $\alpha^{(k+1)}$.
- Step 3: Once $\alpha^{(k+1)}$ is obtained, compute $\lambda_1^{(k+1)} = \widehat{\lambda}_1(\alpha^{(k+1)})$ $\lambda_2^{(k+1)} = \widehat{\lambda}_1(\alpha^{(k+1)})$, $\lambda_3^{(k+1)} = \widehat{\lambda}_1(\alpha^{(k+1)})$.

Now we will discuss how to choose the initial values of the unknown parameters, namely $\alpha^{(0)}$, $\lambda_1^{(0)}$, $\lambda_2^{(0)}$ and $\lambda_3^{(0)}$. We estimate α and $\lambda_1 + \lambda_3$ from $\{x_{11}, \ldots, x_{1n}\}$. Similarly, we can obtain estimates of α and $\lambda_2 + \lambda_3$ from $\{x_{21}, \ldots, x_{2n}\}$, and the estimates of α and $\lambda_1 + \lambda_2 + \lambda_3$ from $\{z_1, \ldots, z_n\}$, where $z_i = \max\{x_{1i}, x_{2i}\}$, for $i = 1, \ldots, n$. We take the average of the three estimates of α to get the initial estimate of α, and using the estimates of $\lambda_1 + \lambda_3$, $\lambda_2 + \lambda_3$ and $\lambda_1 + \lambda_2 + \lambda_3$, we get initial estimate of λ_1, λ_2 and λ_3.

5 Bayesian inference

In this section, we discuss the Bayesian inference of the unknown parameters of the BIW distribution based on a random sample of size n. We assume a very flexible prior on the scale parameters $(\lambda_1, \lambda_2, \lambda_3)$ and on the shape parameter α. It is observed that the Bayes estimator under the squared error loss function cannot be obtained in explicit form, and we propose to use importance sampling procedure to compute the Bayes estimate and the associated credible interval. It is assumed that we have a random sample $\{(x_{11}, x_{21}), \ldots, (x_{1n}, x_{2n})\}$ from $\text{BIW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, and we are using the same notations as in the previous section.

5.1 Prior assumption

When the common shape parameter α is known, we assume the conjugate prior on $(\lambda_1, \lambda_2, \lambda_3)$ as follows. If we denote $\lambda = \lambda_1 + \lambda_2 + \lambda_3$, then it is assumed that for $a > 0$ and $b > 0$, λ has a Gamma(a, b) prior distribution, say $\pi_0(a, b)$. Here the PDF of a Gamma(a, b) for $\lambda > 0$ is

$$\pi_0(\lambda|a, b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda};$$

and 0, otherwise. Given λ, $(\frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda})$ has a Dirichlet prior, say $\pi_1(a_1, a_2, a_3)$, i.e.

$$\pi_1\left(\frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}\middle|\lambda, a_1, a_2, a_3\right) = \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)}\left(\frac{\lambda_1}{\lambda}\right)^{a_1-1}\left(\frac{\lambda_2}{\lambda}\right)^{a_2-1}\left(\frac{\lambda_3}{\lambda}\right)^{a_3-1},$$

for $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_3 > 0$ and $\lambda_3 = \lambda - \lambda_1 - \lambda_2$. Here all the hyper parameters a, b, a_1, a_2, a_3 are greater than 0. For known α, it happens to be the conjugate prior also. If we denote $\bar{a} = a_1 + a_2 + a_3$, then after simplification the joint prior of λ_1, λ_2, λ_3 becomes

$$\pi_1(\lambda_1, \lambda_2, \lambda_3|a, b, a_1, a_2, a_3) = \frac{\Gamma(\bar{a})}{\Gamma(a)}(b\lambda)^{a-\bar{a}} \times \prod_{i=1}^{3} \frac{b^{a_i}}{\Gamma(a_i)}\lambda_i^{a_i-1}e^{-b\lambda_i}. \quad (8)$$

The joint PDF (8) is known as the Gamma–Dirichlet (GD) distribution with parameters a, b, a_1, a_2, a_3, and from now on we will denote it by $\text{GD}(a, b, a_1, a_2, a_3)$. It may be mentioned that the GD distribution is a very flexible distribution. The joint PDF of a GD distribution can take variety of shapes depending on the parameters, and the correlation between the marginals can be both positive and negative depending. Moreover, the marginals become independent if $\bar{a} = a$.

At this moment, we do not assume any specific prior on α. It is simply assumed that the prior on α has a positive support on $(0, \infty)$, and the PDF of prior α, say $\pi_2(\alpha)$ is log-concave. It is further assumed that $\pi_2(\alpha)$ is independent of $\pi_1(\lambda_1, \lambda_2, \lambda_3)$. From now on, the joint prior of α, λ_1, λ_2, λ_3 will be denoted by $\pi(\alpha, \lambda_1, \lambda_2, \lambda_3) = \pi_2(\alpha)\pi_1(\lambda_1, \lambda_2, \lambda_3)$.

5.2 Posterior analysis

In this section, we provide the Bayes estimates of the unknown parameters based on the squared error loss function and the associated HPD credible intervals. Based on the observations, the joint likelihood function of the observed data can be written as

$$\begin{aligned} &l(\mathcal{D}|\alpha, \lambda_1, \lambda_2, \lambda_3) \\ &\quad = \alpha^{2n_1+2n_2+n_0} \lambda_1^{n_2} \lambda_2^{n_1} \lambda_3^{n_0} (\lambda_1 + \lambda_3)^{n_1} (\lambda_2 + \lambda_3)^{n_2} e^{-\lambda_1 T_1(\alpha) - \lambda_2 T_2(\alpha) - \lambda_3 T_3(\alpha)} \\ &\qquad \times \left\{ \prod_{i \in I_0} x_i^{-(\alpha+1)} \right\} \left\{ \prod_{i \in I_1 \cup I_2} x_{1i}^{-(\alpha+1)} x_{2i}^{-(\alpha+1)} \right\}, \end{aligned}$$

where

$$T_1(\alpha) = \sum_{i \in I_1 \cup I_2} x_{1i}^{-\alpha} + \sum_{i \in I_0} x_i^{-\alpha}, \qquad T_2(\alpha) = \sum_{i \in I_1 \cup I_2} x_{2i}^{-\alpha} + \sum_{i \in I_0} x_i^{-\alpha},$$

and

$$T_3(\alpha) = \sum_{i \in I_1} x_{1i}^{-\alpha} + \sum_{i \in I_2} x_{2i}^{-\alpha} + \sum_{i \in I_0} x_i^{-\alpha}.$$

The joint posterior density function of α, λ_1, λ_2, λ_3 can be written as follows

$$l(\lambda_1, \lambda_2, \lambda_3, \alpha|\mathcal{D}) = l(\lambda_1, \lambda_2, \lambda_3|\alpha, \mathcal{D}) \times l(\alpha|\mathcal{D}).$$

In this case, $l(\lambda_1, \lambda_2, \lambda_3|\alpha, \mathcal{D})$ can be written as

$$\begin{aligned} l(\lambda_1, \lambda_2, \lambda_3|\alpha, \mathcal{D}) \propto h(\lambda_1, \lambda_2, \lambda_3) &\times \mathrm{Gamma}\big(\lambda_1; a_1 + n_1, T_1(\alpha) + b\big) \\ &\times \mathrm{Gamma}\big(\lambda_2; a_2 + n_2, T_2(\alpha) + b\big) \\ &\times \mathrm{Gamma}\big(\lambda_3; a_3 + n_0, T_3(\alpha) + b\big), \end{aligned}$$

where

$$h(\lambda_1, \lambda_2, \lambda_3) = \lambda^{a-\bar{a}} (\lambda_1 + \lambda_3)^{n_2} (\lambda_2 + \lambda_3)^{n_1}$$

and $l(\alpha|\mathcal{D})$ can be written as

$$l(\alpha|\mathcal{D}) \propto \frac{\pi_2(\alpha) \times \alpha^{n_0+2n_1+2n_2} \{\prod_{i \in I_0} x_i^{-(\alpha+1)}\} \{\prod_{i \in I_1 \cup I_2} x_{1i}^{-(\alpha+1)} x_{2i}^{-(\alpha+1)}\}}{(T_1(\alpha) + b)^{a_1+n_1} \times (T_2(\alpha) + b)^{a_2+n_2} \times (T_3(\alpha) + b)^{a_3+n_0}}.$$

Therefore, the Bayes estimate of any function of $\alpha, \lambda_1, \lambda_2, \lambda_3$, say $\theta(\alpha, \lambda_1, \lambda_2, \lambda_3)$ under squared error loss function can be obtained as

$$\widehat{\theta}_B = \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \theta(\alpha, \lambda_1, \lambda_2, \lambda_3) l(\alpha, \lambda_1, \lambda_2, \lambda_3 | \mathcal{D})\, d\alpha\, d\lambda_1\, d\lambda_2\, d\lambda_3. \quad (9)$$

Clearly (9) cannot be obtained in explicit form for a general $\theta(\alpha, \lambda_1, \lambda_2, \lambda_3)$. If we denote

$$\begin{aligned} &l_N(\alpha, \lambda_1, \lambda_2, \lambda_3 | \mathcal{D}) \\ &\quad = h(\lambda_1, \lambda_2, \lambda_3) \times \mathrm{Gamma}\big(\lambda_1; a_1 + n_1, T_1(\alpha) + b\big) \\ &\qquad \times \mathrm{Gamma}\big(\lambda_2; a_2 + n_2, T_2(\alpha) + b\big) \\ &\qquad \times \mathrm{Gamma}\big(\lambda_3; a_3 + n_0, T_3(\alpha) + b\big) \\ &\qquad \times \frac{\pi_2(\alpha) \times \alpha^{n_0 + 2n_1 + 2n_2} \{\prod_{i \in I_0} x_i^{-(\alpha+1)}\}\{\prod_{i \in I_1 \cup I_2} x_{1i}^{-(\alpha+1)} x_{2i}^{-(\alpha+1)}\}}{(T_1(\alpha) + b)^{a_1 + n_1} \times (T_2(\alpha) + b)^{a_2 + n_2} \times (T_3(\alpha) + b)^{a_3 + n_0}}, \end{aligned}$$

then (9) can be written as

$$\widehat{\theta}_B = \frac{\int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \theta(\alpha, \lambda_1, \lambda_2, \lambda_3) l_N(\alpha, \lambda_1, \lambda_2, \lambda_3 | \mathcal{D})\, d\alpha\, d\lambda_1\, d\lambda_2\, d\lambda_3}{\int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty l_N(\alpha, \lambda_1, \lambda_2, \lambda_3 | \mathcal{D})\, d\alpha\, d\lambda_1\, d\lambda_2\, d\lambda_3}. \quad (10)$$

We will compute (10) using importance sampling technique, and that can be used to compute the associated HPD credible interval of θ also. The following result will be useful for further development.

Theorem 5.1. *$l(\alpha|\mathcal{D})$ is log-concave.*

Proof. The proof can be obtained following similar approaches as the proof of the Theorem 2 of Kundu (2008) and the transformation used in the proof of the Theorem 4.1. □

Now we suggest an importance sampling technique which will produce simulation consistent estimator of $\widehat{\theta}_B$, and it can be used to construct HPD credible interval of θ also.

Algorithm.

Step 1: Generate α_1 from $l(\alpha|\mathcal{D})$ using the method suggested by Devroye (1984) or Kundu (2008).

Step 2: Generate

$$\begin{aligned} \lambda_{11} | \alpha, \mathcal{D} &\sim \mathrm{Gamma}\big(\lambda_1; a_1 + n_1, T_1(\alpha) + b\big), \\ \lambda_{21} | \alpha, \mathcal{D} &\sim \mathrm{Gamma}\big(\lambda_2; a_2 + n_2, T_2(\alpha) + b\big), \\ \lambda_{31} | \alpha, \mathcal{D} &\sim \mathrm{Gamma}\big(\lambda_3; a_3 + n_3, T_3(\alpha) + b\big). \end{aligned}$$

Step 3: Repeat Step 1 and Step 2, N times and obtained $\{(\alpha_{1i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i});\ i = 1, \ldots, N\}$.

Step 4: A simulation consistent estimator of $\widehat{\theta}_B$ can be obtained as

$$\frac{\sum_{i=1}^{N} \theta_i h(\lambda_{1i}, \lambda_{2i}, \lambda_{3i})}{\sum_{i=1}^{N} h(\lambda_{1i}, \lambda_{2i}, \lambda_{3i})},$$

here $\theta_i = \theta(\alpha_i, \lambda_{1i}, \lambda_{2i}, \lambda_{3i})$.

Now to construct the HPD credible interval of θ, we propose the following steps.

Step 5: Compute w_i for $i = 1, \ldots, N$ as follows

$$w_i = \frac{h(\lambda_{1i}, \lambda_{2i}, \lambda_{3i})}{\sum_{i=1}^{N} h(\lambda_{1i}, \lambda_{2i}, \lambda_{3i})}$$

Step 6: Rearrange $\{(\theta_1, w_1), \ldots, (\theta_N, w_N)\}$ as $\{(\theta_{(1)}, w_{(1)}), \ldots, (\theta_{(N)}, w_{(N)})\}$, where $\theta_{(1)} < \cdots < \theta_{(N)}$ and $w_{(i)}$'s are not ordered, they are just associated with $\theta_{(i)}$. Then a simulation consistent $100(1-\gamma)\%$ credible interval of θ can be obtained as $(\widehat{\theta}_\delta, \widehat{\theta}_{\delta+1-\gamma})$, for $\delta = w_{(1)}, w_{(1)} + w_{(2)}, \ldots, \sum_{i=1}^{N_{1-\gamma}} w_{(i)}$. Here $\widehat{\theta}_p = \theta_{(N_p)}$ and N_p is the integer satisfying

$$\sum_{i=1}^{N_p} \leq p < \sum_{i=1}^{N_p+1} w_{(i)}.$$

Step 7: A $100(1-\gamma)\%$ HPD credible interval of θ can be obtained as $(\widehat{\theta}_{\delta^*}, \widehat{\theta}_{\delta^*+1-\gamma})$, where δ^* satisfies

$$\widehat{\theta}_{\delta^*+1-\gamma} - \widehat{\theta}_{\delta^*} \leq \widehat{\theta}_{\delta+1-\gamma} - \widehat{\theta}_\delta,$$

$$\text{for } \delta = w_{(1)}, w_{(1)} + w_{(2)}, \ldots, \sum_{i=1}^{N_{1-\gamma}} w_{(i)}.$$

6 Simulation results and data analysis

6.1 Simulation results

In this section, we present some simulation results for different samples sizes and for different parameter values, mainly to see how the MLEs computed using the EM algorithm and proposed Bayes estimators work in practice. We mainly consider three different sets of parameter values namely: (i) $\alpha = \lambda_1 = \lambda_2 = \lambda_3 = 1.0$, (ii) $\alpha = \lambda_1 = \lambda_2 = \lambda_3 = 1.5$, (iii) $\alpha = \lambda_1 = \lambda_2 = \lambda_3 = 2.0$ and different $n = 25, 50, 75$ and 100. In each case, we compute the MLEs of the unknown parameters by using the EM algorithm. We start the EM algorithm with the initial guesses as suggested in Section 4, and stop the iteration when the sum of the absolute differences of the estimates at the two consecutive iterates is less than $\varepsilon = 10^{-4}$. We replicate the process 1000 times, and report in Tables 1–3 the average estimates, the associated mean squared errors (MSEs) and the median number of iterations (MNI) required for the convergence of the EM algorithm. We also compute the Bayes estimators of the unknown parameters as suggested in the previous section. To compute the Bayes estimates of the unknown parameters, we need to specify $\pi_2(\alpha)$, the prior on α, and all the hyper parameters of the priors $\pi_1(\alpha)$ and $\pi_2(\lambda_1, \lambda_2, \lambda_3)$. We have assumed $\pi_2(\alpha) \sim \text{Gamma}(c, d)$, and $a = b = a_1 = a_2 = a_3 = c = d = 0.0001$, as suggested by Congdon (2014). We have used $N = 10{,}000$. In this case also we replicate the process 1000 times, and report the average estimates and the MSEs in each case. All the results are also reported in Tables 1–3.

Some of the points are quite clear from the simulation experiments. The Bayes estimates with respect to non-informative priors and MLEs behave very similarly. In all the cases it is observed that as sample size increases the biases and MSEs decrease. It verifies the consistency properties of the MLEs and the Bayes estimators. It is observed that as the parameter values increase the corresponding MSEs increase. The performance of the EM algorithm is quite satisfactory, and it converges within a reasonable number of iterations. The Bayes estimates obtained using importance sampling technique are also as expected.

Table 1 *The average estimates, the associated MSEs (reported within braces below) and MNI for the model* BIW(1.0, 1.0, 1.0, 1.0)

$n \downarrow$	Method	α	λ_1	λ_2	λ_3	MNI
25	MLE	1.0446	1.0562	1.0532	1.0769	12
		(0.0205)	(0.1348)	(0.1324)	(0.1086)	
	Bayes	1.0397	1.0661	1.0481	1.0881	–
		(0.0275)	(0.1211)	(0.1378)	(0.0951)	
50	MLE	1.0179	1.0224	1.0203	1.0260	11
		(0.0084)	(0.0533)	(0.0539)	(0.0438)	
	Bayes	1.0256	1.0119	1.1113	1.0219	–
		(0.0079)	(0.0498)	(0.0610)	(0.0497)	
75	MLE	1.0118	1.0080	1.0118	1.0162	11
		(0.0054)	(0.0365)	(0.0354)	(0.0282)	
	Bayes	1.0154	1.0167	1.0078	1.0218	–
		(0.0049)	(0.0315)	(0.0389)	(0.0310)	
100	MLE	1.0084	1.0051	1.0074	1.0142	11
		(0.0040)	(0.0256)	(0.0254)	(0.0221)	
	Bayes	1.0043	1.0023	1.0089	1.0079	–
		(0.0029)	(0.0227)	(0.0238)	(0.0248)	

Table 2 *The average estimates, the associated MSEs (reported within braces below) and MNI for the model* BIW(1.5, 1.5, 1.5, 1.5)

$n \downarrow$	Method	α	λ_1	λ_2	λ_3	MNI
25	MLE	1.5691	1.6244	1.6121	1.6576	13
		(0.0475)	(0.3650)	(0.3347)	(0.3119)	
	Bayes	1.5976	1.6456	1.5567	1.6213	–
		(0.0427)	(0.4016)	(0.3289)	(0.3216)	
50	MLE	1.5275	1.5490	1.5423	1.5538	13
		(0.0191)	(0.1347)	(0.1265)	(0.1119)	
	Bayes	1.5311	1.5519	1.5523	1.5568	–
		(0.0187)	(0.1468)	(0.1267)	(0.1189)	
75	MLE	1.5181	1.5212	1.5253	1.5337	12
		(0.0123)	(0.0873)	(0.0817)	(0.0706)	
	Bayes	1.5212	1.5318	1.5310	1.5277	–
		(0.0165)	(0.0798)	(0.0799)	(0.0765)	
100	MLE	1.5132	1.5147	1.5168	1.5282	12
		(0.0091)	(0.0621)	(0.0588)	(0.0541)	
	Bayes	1.5127	1.5125	1.5189	1.5178	–
		(0.0078)	(0.0598)	(0.0595)	(0.0588)	

Table 3 *The average estimates, the associated MSEs (reported within braces below) and MNI for the model* BIW(2.0, 2.0, 2, 0, 2, 0)

$n \downarrow$	Method	α	λ_1	λ_2	λ_3	MNI
25	MLE	2.0933	2.2093	2.1817	2.2588	15
		(0.0849)	(0.7671)	(0.6541)	(0.6961)	
	Bayes	2.1019	2.1789	2.1899	2.2123	–
		(0.0823)	(0.8123)	(0.7167)	(0.7221)	
50	MLE	2.0373	2.0823	2.0682	2.0894	14
		(0.0343)	(0.2692)	(0.2342)	(0.2316)	
	Method	2.0512	2.0687	2.0689	2.0699	–
		(0.0318)	(0.2531)	(0.2401)	(0.2405)	
75	MLE	2.0246	2.0384	2.0414	2.0564	14
		(0.0222)	(0.1676)	(0.1489)	(0.1442)	
	Bayes	2.0289	2.0297	2.0311	2.0321	–
		(0.0198)	(0.1705)	(0.1651)	(0.1523)	
100	MLE	2.0182	2.0274	2.0284	2.0461	13
		(0.0163)	(0.1203)	(0.1074)	(0.1089)	
	Bayes	2.0178	2.0198	2.0267	2.0337	–
		(0.0159)	(0.1176)	(0.1123)	(0.1112)	

6.2 Data analysis

In this section, we present the analysis of a data set mainly for illustrative purpose. The main aim of this section is to show how the proposed method can be used in practice. Moreover, it has been shown here that the proposed model works better (in terms of better fitting) to this particular data set than some of the existing bivariate models.

We have analyzed one data set which represents the American Football (National Football League) League data and they are obtained from the matches played on three consecutive weekends in 1986. This is a bivariate date set (X_1, X_2), where X_1 represents the 'game time' to the first points scored by kicking the ball between goal posts, and X_2 represents the 'game time' to the first points scored by moving the ball into the end zone.

The data (scoring times in minutes and seconds) are represented in Table 4. The data set was first analyzed by Csorgo and Welsh (1989), by converting the seconds to the decimal minutes, that is, 2:03 has been converted to 2.05, 3:59 to 3.98 and so on. We have also adopted the same procedure. These times are of interest to a casual spectator who wants to know how long one has to wait to watch a touchdown or to a spectator who is interested only at the beginning stages of a game.

The variables X_1 and X_2 have the following structure: (i) $X_1 < X_2$ means that the first score is a field goal, (ii) $X_1 = X_2$ means the first score is a converted touchdown, (iii) $X_1 > X_2$ means the first score is an unconverted touchdown or safety. In this case, the ties are exact because no 'game time' elapses between a touchdown and a point-after conversion attempt. Therefore, here ties occur quite

Table 4 *American Football League (NFL) data*

Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
2:03	3:59	5:47	25:59	10:24	14:15
9:03	9:03	13:48	49:45	2:59	2:59
0:51	0:51	7:15	7:15	3:53	6:26
3:26	3:26	4:15	4:15	0:45	0:45
7:47	7:47	1:39	1:39	11:38	17:22
10:34	14:17	6:25	15:05	1:23	1:23
7:03	7:03	4:13	9:29	10:21	10:21
2:35	2:35	15:32	15:32	12:08	12:08
7:14	9:41	2:54	2:54	14:35	14:35
6:51	34:35	7:01	7:01	11:49	11:49
32:27	42:21	6:25	6:25	5:31	11:16
8:32	14:34	8:59	8:59	19:39	10:42
31:08	49:53	10:09	10:09	17:50	17:50
14:35	20:34	8:52	8:52	10:51	38:04

naturally and they can not be ignored. Csorgo and Welsh (1989) analyzed the data using the MOBE model but concluded that it does not work well.

Before progressing further first we have fitted IW(α, λ) model to the marginals and the maximum of the two marginals. The MLEs of the unknown parameters, the Kolmogorov–Smirnov (KS) distances between the empirical distribution function (EDF) and the fitted distribution function and the associated p values are reported in Table 5. Based on the p values, it is observed that IW distribution may be used to fit X_1, X_2 and $\max\{X_1, X_2\}$.

Hence, we have used the BIW model to analyze the bivariate data set. To compute the MLEs, we have used the EM algorithm as it has been proposed. We have used the initial guesses as suggested in Section 4. We start the EM algorithm with the above initial guesses, and stop the iteration when the sum of the absolute differences of the estimates at the two consecutive iterates is less than $\varepsilon = 10^{-4}$. In this case, the EM algorithm stops after 11 iterations. The progress of the EM algorithm is provided in Table 6. The final estimates of the unknown parameters are $\widehat{\alpha} = 0.9199$, $\widehat{\lambda}_1 = 0.1605$, $\widehat{\lambda}_2 = 1.9037$, $\widehat{\lambda}_3 = 3.9318$, and the associated 95% parametric bootstrap confidence intervals are (0.7689, 1.1757), (0.0000, 0.5656), (0.9439, 3.5807), (2.8999, 6.5094), respectively. The programs are written in FORTRAN and they are available in the supplementary section.

Table 5 *MLEs, K-S test statistics and the associated p values*

Variable	$\widehat{\alpha}$	$\widehat{\lambda}$	K-S	p
X_1	1.0419	4.6222	0.1855	0.1511
X_2	0.9123	4.6148	0.1961	0.1317
$\max\{X_1, X_2\}$	0.9199	4.6394	0.1941	0.1342

Table 6 *Progress of the EM algorithm*

$\alpha^{(k)}$	$\lambda_1^{(k)}$	$\lambda_2^{(k)}$	$\lambda_3^{(k)}$	LL	k
0.9633	0.0247	0.0172	4.5976	−31.46828	0
0.9254	0.1057	1.8795	4.0418	−24.93369	1
0.9202	0.1381	1.9032	3.9559	−24.92566	2
0.9199	0.1516	1.9036	3.9409	−24.92469	3
0.9199	0.1569	1.9036	3.9354	−24.92455	4
0.9197	0.1590	1.9031	3.9321	−24.92453	5
0.9197	0.1599	1.9031	3.9312	−24.92453	6
0.9197	0.1603	1.9031	3.9309	−24.92453	7
0.9199	0.1604	1.9037	3.9319	−24.92453	8
0.9199	0.1605	1.9037	3.9318	−24.92453	9
0.9196	0.1605	1.9029	3.9302	−24.92453	10
0.9199	0.1605	1.9037	3.9318	−24.92453	11

Now we test whether BIW fits the data or not. We have used the multivariate Kolmogorov–Smirnov test of goodness of fit as proposed by Justel et al. (1997). We obtain the value of the test statistic as 0.2865. Based on 10,000 replications, we obtain the 10% critical value as 0.3112. Hence, $p > 0.1$. Therefore, based on the p value, we cannot reject the null hypotheses that the data are from a BIW distribution.

We compute the Bayes estimates of the unknown parameters based on the same prior assumptions as mentioned in the previous sub-section. In this case, the Bayes estimates and the associated HPD credible intervals are as follows: $\widetilde{\alpha} = 0.9733$, $\widetilde{\lambda}_1 = 0.1495$, $\widetilde{\lambda}_2 = 1.8689$, $\widetilde{\lambda}_3 = 3.9278$, and the associated 95% parametric bootstrap confidence intervals are $(0.7823, 1.1927)$, $(0.0005, 0.4687)$, $(0.8998, 3.5328)$, $(2.7524, 6.1978)$, respectively.

For comparison purposes, we have fitted four-parameter bivariate generalized exponential (BGE) distribution as proposed by Kundu and Gupta (2009) and bivariate generalized Rayleigh (BGR) distribution. Both the distributions have three shape parameters and one scale parameter. We present the MLEs of the unknown parameters and the associated log-likelihood (LL) values in Table 7. Based on the log-likelihood values in this case it is observed the proposed BIW model provides a better fit than BGE or BGR models for this data set.

7 Multivariate inverse Weibull distribution

In this section, we introduce multivariate inverse Weibull (MIW) distribution along the same line, and discuss some of its properties. It can be used as a multivariate heavy tail distribution. Suppose $U_1 \sim \text{IW}(\alpha, \lambda_1), \ldots, U_{p+1} \sim \text{IW}(\alpha, \lambda_{p+1})$, and they are independently distributed. If

$$X_1 = \max\{U_1, U_{p+1}\}, \qquad \ldots, \qquad X_p = \max\{U_p, U_{p+1}\},$$

then $\boldsymbol{X} = (X_1, \ldots, X_p)^T$, is called MIW distribution of order p, with parameters $\alpha, \lambda_1, \ldots, \lambda_{p+1}$. From now on, it will be denoted by $\text{MIW}_p(\alpha, \lambda_1, \ldots, \lambda_{p+1})$. We have the following result regarding MIW distribution.

Theorem 7.1. *Let* $\boldsymbol{X} \sim \text{MIW}_p(\alpha, \lambda_1, \ldots, \lambda_{p+1})$.

Table 7 *MLEs and the associated LL values for BGE and BGR models*

Model	λ	α_1	α_2	α_3	LL
BGE	9.5634	0.0481	0.5959	0.1706	−38.25
BGR	18.0844	0.0152	0.1880	0.3705	−36.53

(a) *The joint CDF of* $\boldsymbol{X}$ *is*

$$P(X_1 \le x_1, \ldots, X_p \le x_p) = e^{-\lambda_1 x_1^{-\alpha} - \cdots - \lambda_p x_p^{-\alpha} - \lambda_{p+1} z^{-\alpha}},$$

where $z = \min\{x_1, \ldots, x_p\}$.

(b) $X_1 \sim \mathrm{IW}(\alpha, \lambda_1 + \lambda_{p+1}), \ldots, X_p \sim \mathrm{IW}(\alpha, \lambda_p + \lambda_{p+1})$.

(c) *For any non-empty subset* $I_q = \{i_1, \ldots, i_q\} \subset \{1, \ldots, p\}$, *the* q*-dimensional marginal*

$$\boldsymbol{X}_{I_q} = (X_{i_1}, \ldots, X_{i_q})^T \sim \mathrm{MIW}_q(\alpha, \lambda_{i_1} + \lambda_{p+1}, \ldots, \lambda_{i_q} + \lambda_{p+1}).$$

(d) *The conditional distribution of* $\boldsymbol{X}_B$ *given* $\{\boldsymbol{X}_A \le \boldsymbol{x}_A\}$, *where the non-empty subsets A and B are disjoint partition of* $\{1, \ldots, p\}$, *is an absolute continuous distribution function as follows*;

$$P(\boldsymbol{X}_B \le \boldsymbol{x}_B | \boldsymbol{X}_A \le \boldsymbol{x}_A) = \begin{cases} e^{-\sum_{i \in B} \lambda_i x_i^{-\alpha}} & \text{if } z = v, \\ e^{-\sum_{i \in B} \lambda_i x_i^{-\alpha} - \lambda_{p+1}(z^{-\alpha} - v^{-\alpha})} & \text{if } z < v, \end{cases}$$

where $z = \min\{x_i; i \in A \cup B\}$ *and* $v = \min\{x_i; i \in A\}$.

(e) *If* $T_n = \max\{X_1, \ldots, X_p\}$, *then* $T_n \sim \mathrm{IW}(\alpha, \lambda_1 + \cdots + \lambda_{p+1})$.

(f) *If* $T_1 = \min\{X_1, \ldots, X_p\}$, *then*

$$P(T_1 \le t) = e^{-\lambda_{p+1} t^{-\alpha}} \times \left(1 - \prod_{i=1}^{p} (1 - e^{-\lambda_i t^{-\alpha}})\right).$$

Proof. (a) Follows from the definition of MIW. (b) and (c) follow from (a). (d) and (e) also follow from the definition. (f) Note that

$$F_{T_1}(t) = \sum_{i=1}^{p} (-1)^{k-1} \sum_{I_k \in S_k} F_{I_k}(t, \ldots, t).$$

Here $I_k = (i_1, \ldots, i_k)$, $1 \le i_1 \ne i_2 \ne \ldots \ne i_k \le n$, is a k-dimensional subset, and S_k is the set of all ordered k-dimensional subsets of $\{1, \ldots, n\}$. Since

$$F_{I_k}(t, \ldots, t) = P(X_{i_1} \le t, \ldots, X_{i_k} \le t) = e^{-\lambda_{p+1} t^{-\alpha} - \sum_{j=1}^{k} \lambda_{i_j} t^{-\alpha}},$$

$$F_{T_1}(t) = e^{-\lambda_{p+1} t^{-\alpha}} \times \sum_{i=1}^{p} (-1)^{k-1} \sum_{I_k \in S_k} e^{-\sum_{j=1}^{k} \lambda_{i_j} t^{-\alpha}}.$$

Now the result follows by observing the fact

$$\sum_{i=1}^{p} (-1)^{k-1} \sum_{I_k \in S_k} e^{-\sum_{j=1}^{k} \lambda_{i_j} t^{-\alpha}} = \left(1 - \prod_{i=1}^{p} (1 - e^{-\lambda_i t^{-\alpha}})\right).$$

□

In Theorem 7.1, we provided the joint CDF of a MIW distribution. It is immediate that the CDF of a MIW is not an absolute continuous distribution except when

$p = 1$. For $p > 1$, it has an absolute continuous part and a singular part. The MIW distribution can be written as

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \alpha F_a(\boldsymbol{x}) + (1-\alpha) F_s(\boldsymbol{x}).$$

Here $0 < \alpha < 1$, $F_a(\boldsymbol{x})$ and $F_s(\boldsymbol{x})$, denote the absolute continuous and singular part of $F_{\boldsymbol{X}}(\boldsymbol{x})$, respectively. Further the corresponding PDF can be written as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \alpha f_a(\boldsymbol{x}) + (1-\alpha) f_s(\boldsymbol{x}).$$

The absolute continuous part of $f_a(\boldsymbol{x})$ and α can be obtained from $\frac{\partial^p F_{\boldsymbol{X}}(x_1,\ldots,x_p)}{\partial x_1 \ldots \partial x_p}$. It is clear that $\boldsymbol{x} = (x_1, \ldots, x_p)^T$ belongs to the set where $F_{\boldsymbol{X}}(\boldsymbol{x})$ is absolutely continuous, if and only if x_i's are different. For a given $\boldsymbol{x}$, so that all the x_i's are different, there exists a permutation $\mathcal{P} = \{i_1, \ldots, i_p\}$, so that $x_{i_1} < \cdots < x_{i_p}$. We define for $x_{i_1} < \cdots < x_{i_p}$

$$\begin{aligned} f_{\mathcal{P}}(x_{i_1}, \ldots, x_{i_p}) = {} & f_{\mathrm{IW}}(x_{i_1}; \alpha, \lambda_{i_1} + \lambda_{p+1}) \times f_{\mathrm{IW}}(x_{i_2}; \alpha, \lambda_{i_2}) \times \cdots \\ & \times f_{\mathrm{IW}}(x_{i_p}; \alpha, \lambda_{i_p}). \end{aligned}$$

Then for $x_{i_1} < \cdots < x_{i_p}$,

$$\frac{\partial^p F_{\boldsymbol{X}}(x_1, \ldots, x_p)}{\partial x_1 \ldots \partial x_p} = \alpha f_a(x_1, \ldots, x_p) = f_{\mathcal{P}}(x_{i_1}, \ldots, x_{i_p}).$$

Further,

$$\begin{aligned} \alpha &= \alpha \int_{\mathbb{R}^p} f_a(x_1, \ldots, x_p)\, dx_1 \ldots dx_p \\ &= \sum_{\mathcal{P}} \int_{x_{i_p}=0}^{\infty} \int_{x_{i_{p-1}}=0}^{x_{i_p}} \cdots \int_{x_{i_1}=0}^{x_{i_2}} f_{\mathcal{P}}(x_1, \ldots, x_p)\, dx_{i_1} \ldots d_{x_{i_p}} = \sum_{\mathcal{P}} J_{\mathcal{P}} \qquad \text{(say)}. \end{aligned}$$

Since

$$\int_{x_{i_1}=0}^{x_{i_2}} f_{\mathcal{P}}(x_1, \ldots, x_p)\, dx_{i_1} = F_{\mathrm{IW}}(x_{i_2}; \alpha, \lambda_{i_1} + \lambda_{p+1}) \prod_{j=2}^{p} f_{\mathrm{IW}}(x_{i_j}; \alpha, \lambda_{i_j}),$$

and

$$\begin{aligned} & \int_{x_{i_2}=0}^{x_{i_3}} \int_{x_{i_1}=0}^{x_{i_2}} f_{\mathcal{P}}(x_1, \ldots, x_p)\, dx_{i_1}\, dx_{i_2} \\ & \qquad = \frac{\lambda_{i_2}}{\lambda_{i_1} + \lambda_{i_2} + \lambda_{p+1}} \times F_{\mathrm{IW}}(x_{i_3}; \alpha, \lambda_{i_1} + \lambda_{i_2} + \lambda_{i_3}) \times \prod_{j=3}^{p} f_{\mathrm{IW}}(x_{i_j}; \alpha, \lambda_{i_j}) \\ & \vdots \\ & J_{\mathcal{P}} = \frac{\lambda_{i_2}}{\lambda_{i_1} + \lambda_{i_2} + \lambda_{p+1}} \times \cdots \times \frac{\lambda_{i_p}}{\lambda_{i_1} + \cdots + \lambda_{i_p} + \lambda_{p+1}}. \end{aligned}$$

Therefore,

$$\alpha = \sum_{\mathcal{P}} \frac{\lambda_{i_2}}{\lambda_{i_1} + \lambda_{i_2} + \lambda_{p+1}} \times \cdots \times \frac{\lambda_{i_p}}{\lambda_{i_1} + \cdots + \lambda_{i_p} + \lambda_{p+1}},$$

and for all $x_{i_1} < \cdots < x_{i_p}$,

$$f_a(\boldsymbol{x}) = \frac{1}{\alpha} f_{\mathcal{P}}(\boldsymbol{x}).$$

Now we provide different components of $f_s(\boldsymbol{x})$, taking into account that $f_{\boldsymbol{X}}(\boldsymbol{x})$ can be written as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \alpha f_a(\boldsymbol{x}) + \sum_{k=2}^{p} \sum_{I_k \subset I} \alpha_k f_{I_k}(\boldsymbol{x}),$$

where $I_k = \{i_1, \ldots, i_k\} \subset I = \{1, \ldots, p\}$, such that $i_1 < \cdots < i_k$. Here, it is understood that each $f_{I_k}(\boldsymbol{x})$ is a PDF with respect to $(p - k + 1)$ dimensional Lebesgue measure on the hyperplane $A_{I_k} = \{\boldsymbol{x} \in \mathbb{R}^p : x_{i_1} = \cdots = x_{i_k}\}$. The exact meaning of $f_{\boldsymbol{X}}(\boldsymbol{x})$ is as follows. For any Borel measurable set $B \in \mathbb{R}^p$,

$$P(\boldsymbol{X} \in B) = \alpha \int_B f_a(\boldsymbol{x}) + \sum_{k=2}^{p} \sum_{I_k \subset I} \alpha_{I_k} \int_{B_{I_k}} f_{I_k}(\boldsymbol{x}),$$

where $B_{I_k} = B \cap A_{I_k}$ is the projection of the set B onto the $(p-k+1)$-dimensional hyperplane A_{I_k}. Now we provide α_{I_k} and $f_{I_k}(\boldsymbol{x})$. Note that if $\boldsymbol{x} \in A_{I_k}$, then $\boldsymbol{x}$ has the following form

$$\boldsymbol{x} = \big(x_1, \ldots, x_{i_1-1}, x^*, x_{i_1+1}, \ldots, x_{i_2-1}, x^*, x_{i_2+1}, \ldots, x_{i_k-1}, x^*, x_{i_k+1}, \ldots, x_p\big).$$

For a given $\boldsymbol{x} \in \mathbb{R}^p$, we define a function g_{I_k} from the $(p - k + 1)$-dimensional hyperplane A_{I_k} to $\mathbb{R}$ as follows

$$g_{I_k}(\boldsymbol{x}) = f_{\mathrm{IW}}\big(x^*; \alpha, \lambda_{p+1}\big) F_{\mathrm{IW}}\Big(x^*; \alpha, \sum_{i \in I_k} \lambda_i\Big) \prod_{i \in I - I_k} f_{\mathrm{IW}}(x_i; \alpha, \lambda_i),$$

if $x_i > x^*$ for $i \in I - I_k$, and zero otherwise. We used the notation $\prod_{i \in I - I_k} = 1$, when $k = p$. Now it follows along the same line as before that

$$\begin{aligned}
\int_{A_{I_k}} g_{I_k}(\boldsymbol{x})\, d\boldsymbol{x} &= \sum_{\mathcal{P}_{I-I_k}} \int_{x_{j_{p-k}}=0}^{\infty} \int_{x_{j_{p-k-1}}=0}^{x_{j_{p-k}}} \cdots \int_{x_{j_1}=0}^{x_{j_2}} g_{I_k}(\boldsymbol{x})\, dx^*\, dx_{j_1} \ldots dx_{j_{p-k}} \\
&= \sum_{\mathcal{P}_{I-I_k}} \frac{\lambda_{p+1}}{\sum_{i \in I_k} \lambda_i + \lambda_{p+1}} \times \frac{\lambda_{j_1}}{\sum_{i \in I_k} \lambda_i + \lambda_{j_1} + \lambda_{p+1}} \times \cdots \\
&\quad \times \frac{\lambda_{j_{p-k}}}{\sum_{i \in I} \lambda_i + \lambda_{p+1}},
\end{aligned}$$

and

$$f_{I_k}(\boldsymbol{x}) = \frac{1}{\alpha_{I_k}} g_{I_k}(\boldsymbol{x}).$$

8 Conclusions

In this paper, we have introduced BIW distribution along the same line as the MOBE distribution. The proposed BIW distribution has four parameters and it has an absolute continuous part and a singular part. The joint PDF of the absolute continuous part can take different shapes depending on the parameter values but it is always unimodal. The MLEs of the unknown parameters cannot be obtained in closed form. A very convenient EM algorithm has been proposed, and in this case at each E-step the corresponding M-step can be performed by solving only one non-linear equation. One data set has been analyzed and it is observed that the performance of the proposed EM algorithm is quite satisfactory. Bayesian inference of the unknown parameters have also been developed based on a very flexible prior. Finally, a multivariate generalization has been proposed and several properties have been developed. It will be interesting to develop inferential issues for the multivariate case. More work is needed in these directions.

Comment. At the final acceptance stage of this article, the referee pointed out the manuscript by Muhammad (2016), where the author also introduced the same bivariate inverse Weibull model. Although the model is same, but our treatments are much more intensive. We have provided some physical interpretations of the model, and provided several properties of the model. We have considered both the frequentist and Bayesian inference of the model parameters and finally we have provided the multivariate generalization of the model. Hopefully it will generate further interest along that direction.

Appendix A

Proof of Theorem 2.2. Suppose A is the following event $A = \{U_1 < U_3\} \cap \{U_2 < U_3\}$, then $P(A) = \lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$. Therefore,

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2|A)P(A) + P\big(X_1 \le x_1, X_2 \le x_2|A^c\big)P\big(A^c\big).$$

if $z = x_1 \wedge x_2$, then

$$P(X_1 \le x_1, X_2 \le x_2|A) = e^{-(\lambda_1+\lambda_2+\lambda_3)z^{-\alpha}},$$

and we obtain $P(X_1 \le x_1, X_2 \le x_2|A^c)$ by subtraction. □

Proof of Theorem 2.3. (a) We will use the following notations: $S_0 = \{(x_1, x_2) : x_1 = x_2 \ge 0\}$, $S_1 = \{(x_1, x_2) : 0 \le x_1 < x_2 < \infty\}$, $S_2 = \{(x_1, x_2) : 0 \le x_2 < x_1 < \infty\}$. It is clear that $f_a(x_1, x_2)$ is continuous in $S_1 \cup S_2$. Since $f_a(x, x) =$

$\lim_{x_1,x_2\to x} f_a(x_1,x_2)$, it follows that $f_a(x_1,x_2)$ is continuous in $S_1 \cup S_2$. Since for all $0 < x_1, x_2 < \infty$,

$$f_a(0,0) = f_a(\infty,\infty) = f_a(x_1,0) = f_a(x_1,\infty) = f_a(0,x_2) = f_a(\infty,x_2) = 0,$$

$f_a(x_1,x_2)$ has a local maximum. It can be easily checked by taking derivatives of $\ln f_a(x_1,x_2)$ that $f_a(x_1,x_2)$ does not have any critical point in the region $S_1 \cup S_2$, hence $f_a(x_1,x_2)$ cannot have any local maximum in $S_1 \cup S_2$. Therefore, in this case the local maximum will be at S_0. Note that

$$f_a(x,x) \propto x^{-2(\alpha+1)} e^{-(2\lambda+\lambda_3)x^{-\alpha}},$$

hence x_m can be easily obtained as the solution of $\frac{d}{dx} f_a(x,x) = 0$. Since the solution is unique, it provides the unique maximum. Proofs of (b) and (c) can be obtained by solving the two equations

$$\frac{\partial \ln f_a(x_1,x_2)}{\partial x_1} = 0 \quad \text{and} \quad \frac{\partial \ln f_a(x_1,x_2)}{\partial x_2} = 0,$$

and by observing the fact that under the restrictions the critical points cannot occur simultaneously at in S_1 and S_2 both. □

Appendix B

Proof of Theorem 3.1. (a) can be obtained easily from the joint CDF of X_1 and X_2. (b) can be obtained as for $x \geq 0$,

$$\begin{aligned} P\big(\max\{X_1, X_2\} \leq x\big) &= P(X_1 \leq x, X_2 \leq x) \\ &= P(U_1 \leq x, U_2 \leq x, U_3 \leq x) \\ &= e^{-(\lambda_1+\lambda_2+\lambda_3)x^{-\alpha}}. \end{aligned}$$

(c) can be obtained as follows:

$$\begin{aligned} P(X_1 < X_2) &= \int_0^\infty \int_0^{x_2} f_{\mathrm{IW}}(x_1;\alpha,\lambda_1+\lambda_3) f_{\mathrm{IW}}(x_2;\alpha,\lambda_2)\,dx_1\,dx_2 \\ &= \int_0^\infty \alpha\lambda_2 x_2^{-(\alpha+1)} e^{-(\lambda_1+\lambda_2+\lambda_3)x_2^{-\alpha}}\,dx_2 \\ &= \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}. \end{aligned}$$

□

Proof of Theorem 3.3. (a) Note that the random vector (X_1, X_2) is PLOD, if and only if for all $0 < x_1, x_2 < \infty$,

$$F_{X_1,X_2}(x_1,x_2) \geq F_{X_1}(x_1) F_{X_2}(x_2). \tag{11}$$

In case of BIW model, it immediately follows that it satisfies (11). To prove (b), we need to prove that $P(X_1 \le x_1|X_2 \le x_2)$ is a non-increasing function of x_2 for all $0 < x_1 < \infty$, and also $P(X_2 \le x_2|X_1 \le x_1)$ is a non-increasing function of x_1 for all $0 < x_2 < \infty$. Now the result immediately follows from part (b) of Theorem 3.2. In order to prove (c), we need to show that

$$P\big(X_1 \le x_1, X_2 \le x_2|X_1 \le x_1', X_2 \le x_2'\big), \tag{12}$$

is a non-increasing in x_1', x_2', for all choices of $0 < x_1, x_2 < \infty$. Note that in case of BIW model, (12) can be written as

$$e^{-\lambda_1((x_1\wedge x_1')^{-\alpha}-x_1'^{-\alpha})-\lambda_2((x_2\wedge x_2')^{-\alpha}-x_2'^{-\alpha})-\lambda_3((x_1\wedge x_1'\wedge x_2\wedge x_2')^{-\alpha}-(x_1'\wedge x_2')^{-\alpha})}. \tag{13}$$

Now proof can be established by considering twenty four possible cases namely (i) $x_1 < x_1' < x_2 < x_2'$, (ii) $x_1' < x_1 < x_2 < x_2'$ etc. □

Appendix C

Proof of Theorem 4.1. First, we will prove the following result. Suppose $y_i > 0$, for $i = 1, \dots, n$, then $w(\alpha) = -\ln(\sum_{i=1}^n y_i^\alpha)$ is a concave function. To prove that, note that

$$\frac{d^2 w(\alpha)}{d\alpha^2} = -\frac{\sum_{i\neq j} y_i^\alpha, y_j^\alpha (\ln y_i - \ln y_j)^2}{(\sum_{i=1}^n y_i^\alpha)^2} < 0.$$

If we take $y_i = x_i^{-1}$ in (7), it easily follows that $g(\alpha)$ is a concave function, since $\frac{d^2 g(\alpha)}{d\alpha^2} < 0$. Now the result follows as $\lim_{\alpha\downarrow 0} g(\alpha) = -\infty$ and $\lim_{\alpha\to\infty} g(\alpha) = -\infty$ □

Acknowledgments

The authors would like to thank three anonymous referees for their constructive suggestions.

References

Aleem, M. (2012). The bivariate inverse Weibull distribution its characteristics and properties. *Proceedings of the World Congress on Engineering* **1**, 348.

Balakrishnan, N. and Lai, C.-D. (2009). *Continuous Bivariate Distributions*. New York: Springer.

Congdon, P. (2014). *Applied Bayesian Modeling*, 2nd. ed. New York: Wiley. MR3221807

Csorgo, S. and Welsh, A. H. (1989). Testing for exponential and Marshall–Olkin distribution. *Journal of Statistical Planning and Inference* **23**, 287–300. MR1032686

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm, (with discussion). *Journal of the Royal Statistical Society, Ser. B*. **39**, 1–38. MR0501537

Devroye, L. (1984). A simple algorithm for generating random variables with log-concave density. *Computing* **33**, 247–257. MR0773927

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

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distribution, Vol. 1*. New York: Wiley. MR1299979

Justel, A., Pena, D. and Zamar, R. (1997). A multivariate Kolmogorov–Smirnov test of goodness of fit. *Statistics and Probability Letters* **35**, 251–291. MR1484961

Kundu, D. (2008). Bayesian inference and life testing plan for the Weibull distribution in presence of progressive censoring. *Technometrics* **50**, 144–154. MR2439875

Kundu, D. and Gupta, R. D. (2009). Bivariate generalized exponential distribution. *Journal of Multivariate Analysis* **100**, 581–593. MR2478183

Kundu, D. and Gupta, A. K. (2013). Bayes estimation for the Marshall–Olkin bivariate Weibull distribution. *Computational Statistics and Data Analysis* **57**, 271–281. MR2981087

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association* **62**, 30–44. MR0215400

Muhammed, H. (2016). Bivariate inverse Weibull distribution. *Journal of Statistical Computation and Simulation*. MR3502165

Myrhaug, D. and Leira, B. J. (2011). A bivariate Frechet distribution and its application to the statistics of two successive surf parameters. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment* **225**, 67–74.

Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd edn. New York: Springer. MR2197664

Nelson, W. (1982). *Applied Lifetime Data Analysis*. New York: Wiley. MR0646615

Teugels, J. L. (2014). Bivariate extreme value distribution. *Wiley StatsRef: Statistics Reference Online*.

Yang, J., Qi, Y. and Wang, R. (2009). A class of multivariate copulas with bivariate Frechet marginal copulas. *Insurance: Mathematics and Economics* **45**, 139–147.

Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Pin 208016
India
E-mail: kundu@iitk.ac.in

Department of Mathematics and Statistics
Bowling Green State University
Bowling Green
OH 43403
USA
E-mail: gupta@bgsu.edu