Chilean Journal of Statistics
Vol. 3, No. 1, April 2012, 93–110

Statistical Distributions
Research Paper

Discriminating between the bivariate generalized exponential and bivariate Weibull distributions

Arabin Kumar Dey[1] and Debasis Kundu[2,*]

[1]Department of Mathematics, IIT Gauhati, Gauhati, India
[2]Department of Mathematics and Statistics, IIT Kanpur, Kanpur, India
(Received: 24 July 2011 · Accepted in final form: 03 January 2012)

Abstract

Recently Kundu and Gupta (2009) introduced a bivariate generalized exponential distribution, whose marginals are generalized exponential distributions. The bivariate generalized exponential distribution is a singular distribution, similarly as the well known bivariate Weibull distribution. The corresponding two singular bivariate distributions functions have very similar joint probability density functions. In this paper, we consider the discrimination between these two bivariate distribution functions. The difference of the maximized log-likelihood functions is used in discriminating between the two distribution functions. The asymptotic distribution of the test statistic has been obtained and it can be used to compute the asymptotic probability of correct selection. Monte Carlo simulations are performed to study the effectiveness of the proposed method. One data set has been analyzed for illustrative purposes.

Keywords: Asymptotic distribution · EM algorithm · Likelihood ratio test · Maximum likelihood · Monte Carlo simulations · Probability of correct selection.

Mathematics Subject Classification: Primary 62H30 · Secondary 62E20.

1. Introduction

Recently, the two-parameter generalized exponential (GE) distribution proposed by Gupta and Kundu (1999) has received some attention. The two-parameter GE model, which has one shape parameter and one scale parameter, is a positively skewed distribution. This model has several desirable properties and many of them are very similar to the corresponding properties of the well known Weibull distribution. For example, the probability density functions (PDFs) and the hazard functions (HFs) of the GE and Weibull distributions are very similar. In addition, both distributions have compact cumulative distribution functions (CDFs). These distributions contain the exponential distribution as a special case. Therefore, they are extensions of the exponential distribution but in different manners. It is further observed that the GE distribution can also be used quite successfully in analyzing positively skewed data sets in place of the Weibull distribution. Moreover, often it is very difficult to distinguish between these two distributions. For some recent developments on the GE distribution, and for its different applications, the readers are referred to the review article by Gupta and Kundu (2007).

*Corresponding author. Debasis Kundu. Department of Mathematics, Indian Institute of Technology Kanpur, Kanpur 208016, India. Email: kundu@iitk.ac.in

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

The problem of testing whether some given observations follow one of two (or more) distributions is quite an old statistical problem. Cox (1961) (see also Cox, 1962) was the pioneer in considering this problem. He also discussed the effect of choosing a wrong model. Since then extensive work has been done in discriminating between two or more distributions; see, e.g., Atkinson (1969, 1970), Bain and Englehardt (1980), Marshall et al. (2001), Dey and Kundu (2009, 2010) and the references cited therein.

In recent times, it has been observed (see Gupta and Kundu, 2003, 2006) that, due to the closeness between the Weibull and GE distributions, it is extremely difficult to discriminate between their two corresponding CDFs. Note that if the shape parameter is one, the two CDFs are not distinguishable. For small sample sizes, the probability of correct selection (PCS) can be quite small, even if the shape parameter is not very close to one. Interestingly, although extensive work has been done in discriminating between two or more univariate distributions, but no work has been found in discriminating between two bivariate distributions.

Recently, Kundu and Gupta (2009) introduced a singular bivariate distribution whose marginals follow GE distributions, which is named as the bivariate generalized exponential (BGE) distribution. The four-parameter BGE distribution has several desirable properties and it can be used quite effectively to analyze bivariate data when there are ties. Another well known four-parameter bivariate singular distribution is the bivariate Marshall-Olkin Weibull (BMOW) distribution, which has been used quite effectively to analyze bivariate data when there are ties; see, e.g., Kotz et al. (2000). The BMOW distribution has Weibull marginals. Therefore, it is clear that for certain range of parameter values, the marginals of the BGE and BMOW distributions are very similar. In fact, it is observed that the shapes of the joint PDFs of the BGE and BMOW distributions can also be very similar in nature.

In this paper, we consider discriminating between BGE and BMOW distributions. We use the difference of the values for maximized log-likelihood functions in discriminating between the two CDFs. The exact distribution of the proposed test statistic is difficult to obtain, and hence we obtain its asymptotic distribution. It is observed that the asymptotic distribution of the test statistic is normally distributed and it is used to compute the PCS. In computing the PCS, one needs to compute the misspecified parameters. Computation of the misspecified parameters involves solving a four dimensional optimization problem. We suggest an approximation, which involves solving an one dimensional optimization problem only, which it computationally becomes very efficient. Monte Carlo simulations are performed to study the effectiveness of the proposed method, and it is observed that, even for moderate sample sizes, the asymptotic results match very well with the simulated results.

Rest of the paper is organized as follows. In Section 2, we briefly discuss about the BGE and BMOW distributions. In Section 3, we present the discrimination procedure. In Section 4, we provide the asymptotic distribution of the test statistics for both cases. In Section 5, we discuss the calculation of the misspecified parameters. In Section 6, we conduct Monte Carlo simulation study. In Section 7, we analyze a data set for illustrative purposes. Finally, in Section 8 we conclude the paper.

2. BMOW and BGE Distributions

In this section, we briefly discuss about the BMOW and BGE distributions. We use the following notations throughout the paper. It is assumed that the univariate Weibull distribution with the shape parameter $\alpha > 0$ and the scale parameter $\lambda > 0$ has PDF, CDF and survival function (SF) given by

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

respectively. From now on a Weibull distribution with the PDF as given in Equation (1) is denoted by $\mathrm{WE}(\alpha, \lambda)$. The GE distribution, with the shape parameter $\alpha > 0$ and the scale parameter $\lambda > 0$, has PDF given by

$$f_{\mathrm{GE}}(x; \alpha, \lambda) = \alpha\lambda e^{-\lambda x}\left(1 - e^{-\lambda x}\right)^{\alpha-1}, \; x > 0. \tag{2}$$

The corresponding CDF and SF are

$$F_{\mathrm{GE}}(x; \alpha, \lambda) = \left(1 - e^{-\lambda x}\right)^{\alpha}, \quad \text{and} \quad S_{\mathrm{GE}}(x; \alpha, \lambda) = 1 - \left(1 - e^{-\lambda x}\right)^{\alpha},$$

respectively. A GE distribution with the PDF given in Equation (2) is denoted by $\mathrm{GE}(\alpha, \lambda)$.

2.1 The BMOW distribution

Suppose $U_0 \sim \mathrm{WE}(\alpha, \lambda_0)$, $U_1 \sim \mathrm{WE}(\alpha, \lambda_1)$ and $U_2 \sim \mathrm{WE}(\alpha, \lambda_2)$ and they are independently distributed. Define $X_1 = \min\{U_0, U_1\}$ and $X_2 = \min\{U_0, U_2\}$. Then, the bivariate vector (X_1, X_2) has the BMOW distribution with parameters $\alpha, \lambda_0, \lambda_1, \lambda_2$, and it is denoted by $\mathrm{BMOW}(\Gamma)$, where $\Gamma = (\alpha, \lambda_0, \lambda_1, \lambda_2)$. Then, the (X_1, X_2) has joint SF of the form

$$\begin{aligned} S_{\mathrm{BMOW}}(x_1, x_2; \Gamma) &= \mathrm{P}(X_1 > x_1, X_2 > x_2) = \mathrm{P}(U_1 > x_1, U_2 > x_2, U_0 > z) \\ &= S_{\mathrm{WE}}(x_1; \alpha, \lambda_1) S_{\mathrm{WE}}(x_2; \alpha, \lambda_2) S_{\mathrm{WE}}(z; \alpha, \lambda_0), \end{aligned}$$

where $z = \max\{x_1, x_2\}$. The joint PDF of (X_1, X_2) can be written as

$$f_{\mathrm{BMOW}}(x_1, x_2; \Gamma) = \begin{cases} f_{1W}(x_1, x_2; \Gamma), \text{ if } & 0 < x_1 < x_2; \\ f_{2\mathrm{W}}(x_1, x_2; \Gamma), \text{ if } & 0 < x_2 < x_1; \\ f_{0\mathrm{W}}(x; \Gamma), & \text{if } 0 < x_1 = x_2 = x; \end{cases}$$

where

$$\begin{aligned} f_{1\mathrm{W}}(x_1, x_2; \Gamma) &= f_{\mathrm{WE}}(x_1; \alpha, \lambda_1) f_{\mathrm{WE}}(x_2; \alpha, \lambda_0 + \lambda_2), \\ f_{2\mathrm{W}}(x_1, x_2; \Gamma) &= f_{\mathrm{WE}}(x_1; \alpha, \lambda_0 + \lambda_1) f_{\mathrm{WE}}(x_2; \alpha, \lambda_2), \\ f_{0\mathrm{W}}(x; \Gamma) &= \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} f_{\mathrm{WE}}(x; \alpha, \lambda_0 + \lambda_1 + \lambda_2). \end{aligned}$$

Note that the function $f_{\mathrm{BMOW}}(\cdot)$ may be considered to be a PDF for the BMOW distribution if it is understood that the first two terms are PDFs with respect to two dimensional Lebesgue measure, and the third term is a PDF with respect to a one dimensional Lebesgue measure; see, e.g., Bemis et al. (1972). It is clear that the BMOW distribution has an absolute continuous part on $\{(x_1, x_2); 0 < x_1 < \infty, 0 < x_2 < \infty, x_1 \neq x_2\}$, and a singular part on $\{(x_1, x_2); 0 < x_1 < \infty, 0 < x_2 < \infty, x_1 = x_2\}$. The surface plot of the absolute continuous part of the joint PDF has been provided in Figure 1 for different parameter values. It is immediate that the joint BMOW PDF can take variety of shapes, and, therefore, it can be used quite effectively in analyzing singular bivariate data.

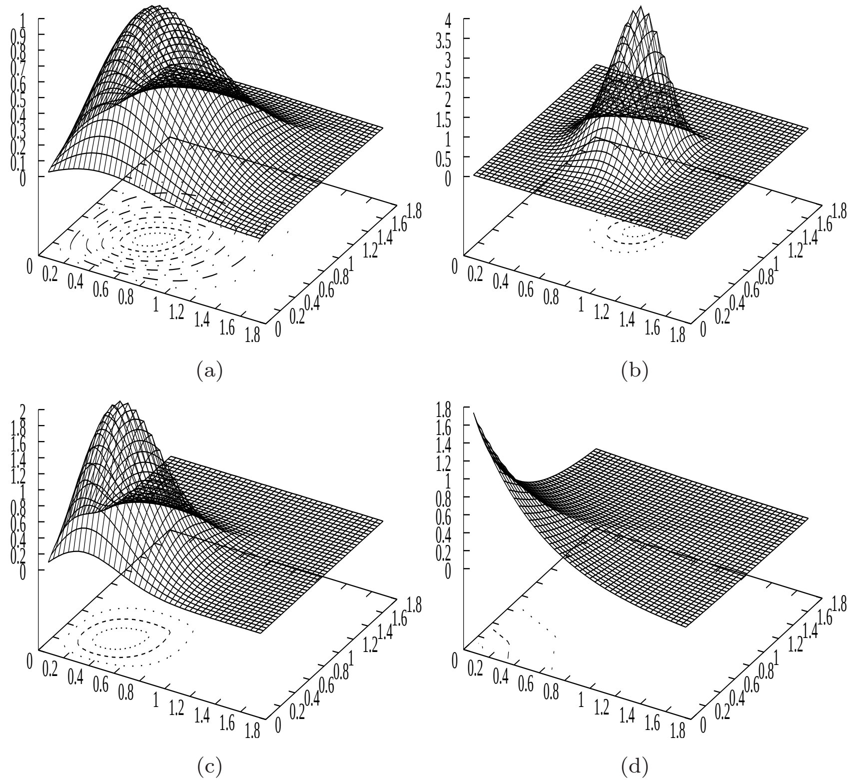

Figure 1. Surface plots of the absolute continuous part of the joint PDF of BMOW for $(\alpha, \lambda_1, \lambda_2, \lambda_3)$: (a) (2.0, 1.0, 1.0, 1.0) (b) (5.0, 1.0, 1.0, 1.0) (c) (2.0, 2.0, 2.0, 2.0) (d) (1.0, 1.0, 1.0, 1.0).

The following probabilities are used later in deriving the asymptotic PCS. If $(X_1, X_2) \sim$ BMOW(Γ), then

$$\begin{aligned}
p_{1\mathrm{W}} = \mathrm{P}(X_1 < X_2) &= \int_0^\infty \int_0^y f_{\mathrm{WE}}(x; \alpha, \lambda_1) f_{\mathrm{WE}}(y; \alpha, \lambda_0 + \lambda_2) dx dy \\
&= \frac{\lambda_1}{\lambda_0 + \lambda_1 + \lambda_2}, \\
p_{2\mathrm{W}} = \mathrm{P}(X_1 > X_2) &= \int_0^\infty \int_y^\infty f_{\mathrm{WE}}(x; \alpha, \lambda_0 + \lambda_1) f_{\mathrm{WE}}(y; \alpha, \lambda_2) dx dy \\
&= \frac{\lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}, \\
p_{0W} = \mathrm{P}(X_1 = X_2) &= \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \int_0^\infty f_{WE}(z; \alpha, \lambda_0 + \lambda_1 + \lambda_2) dz \\
&= \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}.
\end{aligned}$$

2.2 The BGE distribution

Suppose $V_0 \sim \mathrm{GE}(\alpha_0, \lambda)$, $V_1 \sim \mathrm{GE}(\alpha_1, \lambda)$ and $V_2 \sim \mathrm{GE}(\alpha_2, \lambda)$. Define $Y_1 = \max\{V_0, V_1\}$ and $Y_2 = \max\{V_0, V_2\}$. Then the bivariate random vector (Y_1, Y_2) is said to have the BGE distribution with parameters $\alpha_0, \alpha_1, \alpha_2, \lambda$, and it is denoted by BGE(Σ), where $\Sigma = (\alpha_0, \alpha_1, \alpha_2, \lambda)$. It is immediate that $Y_1 \sim$ GE $(\alpha_0 + \alpha_1, \lambda)$ and $Y_2 \sim \mathrm{GE}(\alpha_0 + \alpha_2, \lambda)$. The

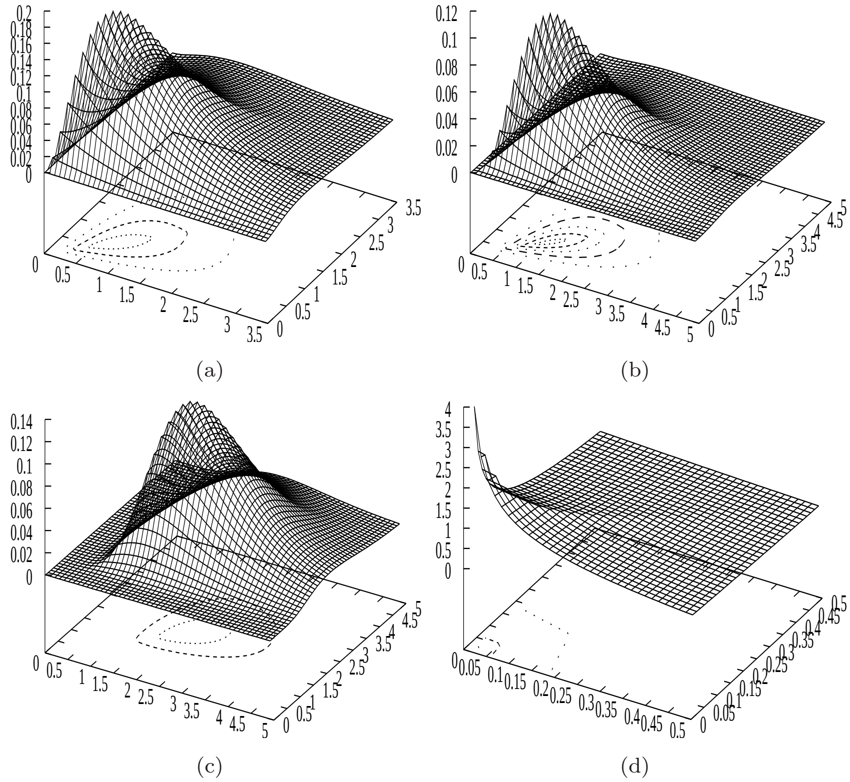

Figure 2. Surface plots of the absolute continuous part of the joint BGE PDF for $(\alpha_1, \alpha_2, \alpha_3, \lambda)$: (a) (1.0, 1.0, 2.0, 1.0) (b) (1.0, 1.0, 1.0, 4.0) (c) (5.0, 5.0, 5.0, 1.0) (d) (0.5, 0.5, 0.5, 1.0).

joint CDF of (Y_1, Y_2) can be expressed as

$$
\begin{aligned}
F_{\mathrm{BGE}}(y_1, y_2; \Sigma) &= \mathrm{P}(Y_1 \leq y_1, Y_2 \leq y_2) = \mathrm{P}(V_1 \leq y_1, V_2 \leq y_2, V_0 \leq v) \\
&= (1 - e^{-\lambda y_1})^{\alpha_1}(1 - e^{-\lambda y_2})^{\alpha_2}(1 - e^{-\lambda v})^{\alpha_0},
\end{aligned}
$$

where $v = \min\{y_1, y_2\}$. In this case, the joint CDF of Y_1 and Y_2 can be written as

$$
f_{\mathrm{BGE}}(y_1, y_2; \Sigma) = \begin{cases} f_{1\mathrm{G}}(y_1, y_2), \text{ if } & 0 < y_1 < y_2; \\ f_{2\mathrm{G}}(y_1, y_2), \text{ if } & 0 < y_2 < y_1; \\ \quad f_{0\mathrm{G}}(y), & \text{if } 0 < y_1 = y_2 = y, \end{cases}
$$

where

$$
\begin{aligned}
f_{1\mathrm{G}}(y_1, y_2; \Sigma) &= f_{\mathrm{GE}}(y_1; \alpha_0 + \alpha_1, \lambda) f_{\mathrm{GE}}(y_2; \alpha_2, \lambda), \\
f_{2\mathrm{G}}(y_1, y_2; \Sigma) &= f_{\mathrm{GE}}(y_1; \alpha_1, \lambda) f_{\mathrm{GE}}(y_2; \alpha_0 + \alpha_2, \lambda), \\
f_{0\mathrm{G}}(y; \Sigma) &= \frac{\alpha_0}{\alpha_0 + \alpha_1 + \alpha_2} f_{\mathrm{GE}}(y; \alpha_0 + \alpha_1 + \alpha_2, \lambda).
\end{aligned}
$$

It is clear that the BGE distribution has also a singular part and an absolute continuous part similarly as the BMOW distribution. The surface plot of the joint BGE PDF is provided in Figure 2 for different parameter values. It is clear that the shape of the joint BGE and BMOW PDFs are very similar.

The following probabilities are needed later. If $(Y_1, Y_2) \sim \mathrm{BGE}(\Sigma)$, then

$$\begin{aligned} p_{1\mathrm{G}} = \mathrm{P}(Y_1 < Y_2) &= \int_0^\infty \int_0^y f_{\mathrm{GE}}(x;\alpha_0+\alpha_1,\lambda) f_{\mathrm{GE}}(y;\alpha_2,\lambda)dxdy \\ &= \frac{\alpha_2}{\alpha_0+\alpha_1+\alpha_2}, \\ p_{2\mathrm{G}} = \mathrm{P}(Y_1 > Y_2) &= \int_0^\infty \int_y^\infty f_{\mathrm{GE}}(x;\alpha_1,\lambda) f_{\mathrm{GE}}(y;\alpha_0+\alpha_2,\lambda)dxdy \\ &= \frac{\alpha_1}{\alpha_0+\alpha_1+\alpha_2}, \\ p_{0\mathrm{G}} = \mathrm{P}(Y_1 = Y_2) &= \frac{\alpha_0}{\alpha_0+\alpha_1+\alpha_2} \int_0^\infty f_{\mathrm{WE}}(z;\alpha_0+\alpha_1+\alpha_2,\lambda)dz \\ &= \frac{\alpha_0}{\alpha_0+\alpha_1+\alpha_2}. \end{aligned}$$

3. Discrimination Procedure

In this section, we present the discrimination procedure between the distributions. Specifically, suppose $\{(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})\}$ is a random bivariate sample of size n generated either from a BGE(Σ) distribution or from a BMOW(Γ) distribution. Based on the above sample, we want to decide from which distribution the data set has been obtained. We use the following notations and sets for the rest of the paper; $I_0 = \{(x_{1i}, x_{2i}), x_{1i} = x_{2i} = x_i, i = 1, \ldots, n\}$, $I_1 = \{(x_{1i}, x_{2i}), x_{1i} < x_{2i}, i = 1, \ldots, n\}$, $I_2 = \{(x_{1i}, x_{2i}), x_{1i} > x_{2i}, i = 1, \ldots, n\}$, $I = I_0 \cup I_1 \cup I_2$, $n_0 = |I_0|$, $n_1 = |I_1|$ and $n_2 = |I_2|$, $n_0 + n_1 + n_2 = n$. It is assumed that $n_0 \neq 0$, $n_1 \neq 0$, and $n_2 \neq 0$. Let $\widehat{\Sigma} = (\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\lambda})$ be the maximum likelihood (ML) estimators of Σ, based on the assumption that the data have been obtained from the BGE(Σ) distribution. Similarly, let $\widehat{\Gamma} = (\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1, \widehat{\lambda}_2)$ be the ML estimator of Γ based on the assumption that the data have been obtained from the BMOW(Γ) distribution. Note that $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\lambda})$ and $(\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1, \widehat{\lambda}_2)$ are obtained by maximizing the corresponding log-likelihood function, say $L_1(\alpha_0, \alpha_1, \alpha_2, \lambda)$ and $L_2(\alpha, \lambda_0, \lambda_1, \lambda_2)$, respectively. Note that here the log-likelihood function of the BGE distribution can be written as

$$\begin{aligned} L_1(\Sigma) = {} & (n_0 + 2n_1 + 2n_2)\log(\lambda) + n_1\log(\alpha_0+\alpha_1) + n_1\log(\alpha_2) + n_2\log(\alpha_1) \\ & + n_2\log(\alpha_0+\alpha_2) + (\alpha_0+\alpha_1-1)\sum_{i\in I_1}\log(1-e^{-\lambda x_{1i}}) + (\alpha_2-1)\sum_{i\in I_1}\log(1-e^{-\lambda x_{2i}}) \\ & + (\alpha_1-1)\sum_{i\in I_2}\log(1-e^{-\lambda x_{1i}}) + (\alpha_0+\alpha_2-1)\sum_{i\in I_2}\log(1-e^{-\lambda x_{2i}}) + n_0\log(\alpha_0) \\ & + (\alpha_0+\alpha_1+\alpha_2-1)\sum_{i\in I_0}\log(1-e^{-\lambda x_i}) - \lambda\left(\sum_{i\in I_0} x_i + \sum_{i\in I_1\cup I_2} x_{1i} + \sum_{i\in I_1\cup I_2} x_{2i}\right), \quad (3) \end{aligned}$$

and the BMOW log-likelihood function can be written as

$$
\begin{aligned}
L_2(\Gamma) = {} & (n_0 + 2n_1 + 2n_2)\log(\alpha) + n_1\log(\lambda_1) + n_2\log(\lambda_2) + n_0\log(\lambda_0) + n_1\log(\lambda_0+\lambda_2) \\
& + n_2\log(\lambda_0+\lambda_1) + (\alpha-1)\left[\sum_{i\in I_0}\log(x_{1i}) + \sum_{i\in I_1\cup I_2}\log(x_{2i}) + \sum_{i\in I_0}\log(x_i)\right] \\
& - \lambda_1\left[\sum_{i\in I_1\cup I_2} x_{1i}^{\alpha} + \sum_{i\in I_0} x_i^{\alpha}\right] - \lambda_2\left[\sum_{i\in I_1\cup I_2} x_{2i}^{\alpha} + \sum_{i\in I_0} x_i^{\alpha}\right] \\
& - \lambda_0\left[\sum_{i\in I_2} x_{1i}^{\alpha} + \sum_{i\in I_1} x_{2i}^{\alpha} + \sum_{i\in I_0} x_i^{\alpha}\right].
\end{aligned}
$$

We use the following discrimination procedure. Consider the statistic

$$
T = L_2(\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1, \widehat{\lambda}_2) - L_1(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\lambda}). \tag{4}
$$

If $T > 0$, we choose the BMOW distribution, otherwise we prefer the BGE distribution. It may be mentioned that $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\lambda})$ and $(\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1, \widehat{\lambda}_2)$ are obtained by maximizing Equations (3) and (4) respectively. Computationally both are quite challenging problems. To maximize directly these problems one needs to solve a four dimensional optimization problem in each case. In both the cases the EM algorithm can be used quite effectively to compute the ML estimators of the unknown parameters; see e.g., Kundu and Gupta (2009) and Kundu and Dey (2009) for the BGE and BMOW distributions, respectively. In each case, it involves solving just a one-dimensional optimization problem at each "E" step, and both the methods work quite well. In the next section we provide the asymptotic distribution of T, which helps to compute the asymptotic PCS.

4. Asymptotic Distributions

In this section, we provide the asymptotic distributions of the test statistics for both cases and use the following notations. For any functions, $f_1(U)$ and $f_2(U)$, $\mathrm{E}_{\mathrm{BGE}}[f_1(U)]$, $\mathrm{V}_{\mathrm{BGE}}[f_1(U)]$ and $\mathrm{Cov}_{\mathrm{BGE}}(f_2(U), f_1(U))$ denote the mean of $f_1(U)$, the variance of $f_1(U)$, and the covariance of $f_1(U)$ and $f_2(U)$ respectively, under the assumption the $U \sim \mathrm{BGE}(\Sigma)$. Similarly, we define $\mathrm{E}_{\mathrm{BWE}}[f_1(U)]$, $\mathrm{V}_{\mathrm{BWE}}[f_1(U)]$ and $\mathrm{Cov}_{\mathrm{BWE}}(f_2(U), f_1(U))$ as the mean of $f_1(U)$, the variance of $f_1(U)$ and the covariance of $f_1(U)$ and $f_2(U)$ respectively, under the assumption that $U \sim \mathrm{BWE}(\Gamma)$ (bivariate Weibull). We have the following two main results.

Theorem 4.1 Under the assumption that data come from the $\mathrm{BMOW}(\alpha, \lambda_0, \lambda_1, \lambda_2)$ distribution, the distribution of T as defined in Equation (4) is approximately normally distributed with mean $\mathrm{E}_{\mathrm{BMOW}}[T]$ and variance $\mathrm{V}_{\mathrm{BMOW}}[T]$. The expressions of $\mathrm{E}_{\mathrm{BMOW}}[T]$ and $\mathrm{V}_{\mathrm{BMOW}}[T]$ are provided below.

Proof It is provided in Appendix. ■

Now we provide the expressions for $\mathrm{E}_{\mathrm{BMOW}}[T]$ and $\mathrm{V}_{\mathrm{BMOW}}[T]$. We denote

$$
\lim_{n\to\infty} \frac{\mathrm{E}_{\mathrm{BMOW}}[T]}{n} = AM_{\mathrm{BMOW}} \quad \text{and} \quad \lim_{n\to\infty} \frac{\mathrm{E}_{\mathrm{BMOW}}[T]}{n} = AV_{\mathrm{BMOW}}.
$$

Therefore,

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{E}_{\mathrm{BMOW}}[T] = AM_{\mathrm{BMOW}}$$
$$= \mathrm{E}_{\mathrm{BMOW}}[\log(f_{\mathrm{BMOW}}(X_1, X_2; \Gamma)) - \log(f_{\mathrm{BGE}}(X_1, X_2; \widetilde{\Sigma}))],$$
$$\lim_{n\to\infty}\frac{1}{n}\mathrm{V}_{\mathrm{BMOW}}[T] = AV_{\mathrm{BMOW}}$$
$$= \mathrm{V}_{\mathrm{BMOW}}[\log(f_{\mathrm{BMOW}}(X_1, X_2; \Gamma)) - \log(f_{\mathrm{BGE}}(X_1, X_2; \widetilde{\Sigma}))].$$

Note that both AM_{BMOW} and AV_{BMOW} cannot be obtained in explicit form. They have to be obtained numerically and they are functions of $p_{1\mathrm{W}}$, $p_{2\mathrm{W}}$, $p_{3\mathrm{W}}$, Γ and $\widetilde{\Sigma}$. Moreover, it should be mentioned that the misspecified parameter $\widetilde{\Sigma}$ as defined in Lemma 8.1 (see Appendix) also needs to be computed numerically.

Theorem 4.2 Under the assumption that data come from the BGE(Σ) distribution, the distribution of T as defined in Equation (4) is approximately normally distributed with mean $\mathrm{E}_{\mathrm{BGE}}[T]$ and variance $\mathrm{V}_{\mathrm{BGE}}[T]$. The expressions of $\mathrm{E}_{\mathrm{BGE}}[T]$ and $\mathrm{V}_{\mathrm{BGE}}[T]$ are provided below.

Proof It is provided in Appendix. ■

Now we provide the expressions for $\mathrm{E}_{\mathrm{BGE}}[T]$ and $\mathrm{V}_{\mathrm{BGE}}[T]$. In this case, we denote

$$\lim_{n\to\infty}\frac{\mathrm{E}_{\mathrm{BGE}}[T]}{n} = AM_{\mathrm{BGE}} \quad \text{and} \quad \lim_{n\to\infty}\frac{\mathrm{V}_{\mathrm{BGE}}[T]}{n} = AV_{\mathrm{BGE}}.$$

Therefore,

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{E}_{\mathrm{BGE}}[T] = AM_{\mathrm{BGE}} = \mathrm{E}_{\mathrm{BGE}}[\log(f_{\mathrm{BMOW}}(X_1, X_2; \widetilde{\Gamma})) - \log(f_{\mathrm{BGE}}(X_1, X_2; \Sigma))],$$
$$\lim_{n\to\infty}\frac{1}{n}\mathrm{V}_{\mathrm{BGE}}[T] = AV_{\mathrm{BGE}} = \mathrm{V}_{\mathrm{BGE}}[\log(f_{\mathrm{BMOW}}(X_1, X_2; \widetilde{\Gamma})) - \log(f_{\mathrm{BGE}}(X_1, X_2; \Sigma))].$$

As mentioned, here also both AM_{BGE} and AV_{BGE} cannot be obtained in explicit form. They have to be obtained numerically and they are also functions of $p_{1\mathrm{G}}$, $p_{2\mathrm{G}}$, $p_{3\mathrm{G}}$, $\widetilde{\Gamma}$ and Σ. The misspecified parameter $\widetilde{\Gamma}$ as defined in Lemma 8.2 (see Appendix) also needs to be computed numerically.

Then, based on the corresponding asymptotic distributions, it is possible to compute the PCS for both the cases.

5. Misspecified Parameter Estimates

In this section, we discuss the estimation of the misspecified parameters.

5.1 Estimation of $\widetilde{\Sigma}$

In this case, it is assumed that the data have been obtained from the BMOW(Γ) distribution and we would like to compute $\widetilde{\Sigma}$, the misspecified BGE parameters, as defined in Lemma 8.1. Suppose $(X_1, X_2) \sim$ BMOW(Γ). Consider the following events:

$A_1 = \{X_1 < X_2\}$, $A_2 = \{X_1 > X_2\}$ and $A_0 = \{X_1 = X_2\}$. Moreover, 1_A is the indicator function taking value 1 at the set A and 0 otherwise. Therefore, $\widetilde{\Sigma}$ can be obtained as the argument maximum of $\mathrm{E}_{\mathrm{BMOW}}[\log(f_{\mathrm{BGE}}(X_1, X_2; \Sigma))] = \Pi_1(\Sigma)$ (say), where

$$\begin{aligned}
\Pi_1(\Sigma) = {} & \log(\lambda) + p_{1\mathrm{W}}\log(\alpha_0 + \alpha_1) + p_{1\mathrm{W}}\log(\alpha_2) \\
& + (\alpha_0 + \alpha_1 - 1)\mathrm{E}_{\mathrm{BMOW}}[\log(1 - e^{-\lambda X_1}) \cdot 1_{A_1}] \\
& + (\alpha_2 - 1)\mathrm{E}_{\mathrm{BMOW}}[\log(1 - e^{-\lambda X_2}) \cdot 1_{A_1}] - \lambda \mathrm{E}_{\mathrm{BMOW}}[(X_1 + X_2) \cdot 1_{A_1}] \\
& + p_{2\mathrm{W}}\log(\alpha_1) + (\alpha_1 - 1)\mathrm{E}_{\mathrm{BMOW}}[\log(1 - e^{-\lambda X_1}) \cdot 1_{A_2}] + p_{2\mathrm{W}}\log(\alpha_0 + \alpha_2) \\
& + (\alpha_0 + \alpha_2 - 1)\mathrm{E}_{\mathrm{BMOW}}[\log(1 - e^{-\lambda X_2}) \cdot 1_{A_2}] - \lambda \mathrm{E}_{\mathrm{BMOW}}[(X_1 + X_2) \cdot I_{A_2}] \\
& + (\alpha_0 + \alpha_1 + \alpha_2 - 1)\mathrm{E}_{\mathrm{BMOW}}[\log(1 - e^{-\lambda X}) \cdot 1_{A_0}] - \lambda \mathrm{E}_{\mathrm{BMOW}}[X \cdot 1_{A_0}] + p_{0\mathrm{W}}\log\alpha_0.
\end{aligned}$$

We need to maximize $\Pi_1(\Sigma)$ with respect to Σ for fixed Γ, to compute $\widetilde{\Sigma}$, numerically. Clearly, $\widetilde{\Sigma}$ is a function of Γ, but we do not make it explicit for brevity. Since maximizing $\Pi_1(\Sigma)$ involves a four dimensional optimization process, we suggest to use an approximate version of it, which can be performed very easily, and works quite well in practice. The idea basically came from the missing value principle, and it has been used by Kundu and Gupta (2009) in developing the EM algorithm. We suggest to use the following $\Pi_1^*(\Sigma)$, the "pseudo" version of $\Pi_1(\Sigma)$

$$\begin{aligned}
\Pi_1^*(\Sigma) = {} & (p_{0\mathrm{W}} + u_2 p_{1\mathrm{W}} + w_2 p_{2\mathrm{W}})\log(\alpha_0) + (p_{0\mathrm{W}} + 2p_{1\mathrm{W}} + 2p_{2\mathrm{W}})\log(\lambda) \\
& + (\alpha_0 + \alpha_1 + \alpha_2 - 1)\mathrm{E}\left[\log(1 - e^{-\lambda X_1}) \cdot 1_{A_0}\right] \\
& - \lambda\left(\mathrm{E}[X_1 \cdot 1_{A_0}] + \mathrm{E}\left[(X_1 + X_2) \cdot 1_{A_1 \cup A_2}\right]\right) + (u_1 p_{1\mathrm{W}} + p_{2\mathrm{W}})\log(\alpha_1) \\
& + (w_1 p_{2\mathrm{W}} + p_{1\mathrm{W}})\log(\alpha_2) + (\alpha_0 + \alpha_1 - 1)\mathrm{E}\left[\log(1 - e^{-\lambda X_1}) \cdot 1_{A_1}\right] \\
& + (\alpha_0 + \alpha_2 - 1)\mathrm{E}\left[\log(1 - e^{-\lambda X_2}) \cdot 1_{A_2}\right] + (\alpha_2 - 1)\mathrm{E}\left[\log(1 - e^{-\lambda X_2}) \cdot 1_{A_1}\right] \\
& + (\alpha_1 - 1)\mathrm{E}\left[\log(1 - e^{-\lambda X_1}) \cdot 1_{A_2}\right].
\end{aligned}$$

Here,

$$u_1 = \frac{\lambda_0}{\lambda_0 + \lambda_2}, \quad u_2 = \frac{\lambda_2}{\lambda_0 + \lambda_2}, \quad w_1 = \frac{\lambda_0}{\lambda_0 + \lambda_1}, \quad w_2 = \frac{\lambda_1}{\lambda_0 + \lambda_1}, \tag{5}$$

and $p_{1\mathrm{W}}, p_{2\mathrm{W}}, p_{3\mathrm{W}}$ are same as defined before. The explicit expressions of the expected values are provided in Appendix. Note that $\Pi_1^*(\Sigma)$ is actually

$$\Pi_1^*(\Sigma) = \lim_{n \to \infty} \frac{1}{n}\mathrm{E}\left[l_{\mathrm{pseudo}}(\alpha_0, \alpha_1, \alpha_2, \lambda \mid (X_{1i}, X_{2i}; i = 1, \ldots, n)\right].$$

Here $l_{\mathrm{pseudo}}(\cdot)$ is the "pseudo" log-likelihood function of the complete data set, as described in Kundu and Gupta (2009). Moreover, it has the same form as in Kundu and Gupta (2009), but since here it is assumed that $(X_{1i}, X_{2i}) \sim \mathrm{BMOW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, therefore the expressions of u_1, u_2, w_1, w_2 are as Equation (5), and they are different than Kundu and Gupta (2009).

Now the maximization of $\Pi_1^*(\Sigma)$ can be performed as follows. Note that for a given λ, the maximization of $\Pi_1^*(\Sigma)$ with respect to α_0, α_1 and α_2 can occur at

$$\widetilde{\alpha}_0(\lambda) = \frac{p_{0\mathrm{W}} + u_2 p_{1\mathrm{W}} + w_2 p_{2\mathrm{W}}}{\mathrm{E}\left[\log\left(1-e^{-\lambda X_1}\right)\cdot 1_{A_0}\right] + \mathrm{E}\left[\log\left(1-e^{-\lambda X_1}\right)\cdot 1_{A_1}\right] + \mathrm{E}\left[\log\left(1-e^{-\lambda X_2}\right)\cdot 1_{A_2}\right]},$$
$$\widetilde{\alpha}_1(\lambda) = \frac{u_1 p_{1\mathrm{W}} + p_{2\mathrm{W}}}{\mathrm{E}\left[\log\left(1-e^{-\lambda X_1}\right)\cdot 1_{A_0}\right] + \mathrm{E}\left[\log\left(1-e^{-\lambda X_1}\right)\cdot 1_{A_1}\right] + \mathrm{E}\left[\log\left(1-e^{-\lambda X_1}\right)\cdot 1_{A_2}\right]},$$
$$\widetilde{\alpha}_2(\lambda) = \frac{p_{1\mathrm{W}} + w_1 p_{2\mathrm{W}}}{\mathrm{E}\left[\log\left(1-e^{-\lambda X_2}\right)\cdot 1_{A_0}\right] + \mathrm{E}\left[\log\left(1-e^{-\lambda X_2}\right)\cdot 1_{A_2}\right] + \mathrm{E}\left[\log\left(1-e^{-\lambda X_2}\right)\cdot 1_{A_1}\right]},$$

respectively, and finally maximization of $\Pi_1^*(\Sigma)$ can be obtained by maximizing profile function, namely, $\Pi_1^*(\widetilde{\alpha}_0(\lambda), \widetilde{\alpha}_1(\lambda), \widetilde{\alpha}_2(\lambda), \lambda)$ with respect to λ only. Therefore, it involves solving an one dimensional optimization problem only.

5.2 Estimation of $\widetilde{\Gamma}$

In this case, it is assumed that the data have been obtained from the BGE(Σ) distribution and we compute $\widetilde{\Gamma}$, the misspecified BMOW parameters, as defined in Lemma 8.2. In this case, $\widetilde{\Gamma}$ can be obtained as the argument maximum of $\mathrm{E}_{\mathrm{BGE}}[\log\left(f_{\mathrm{BMOW}}(X_1, X_2; \Gamma)\right)] = \Pi_2(\Gamma)$ (say), where

$$\begin{aligned}
\Pi_2(\Gamma) = {} & (p_{0\mathrm{G}} + 2p_{1\mathrm{G}} + 2p_{2\mathrm{G}})\log(\alpha) \\
& + p_{1\mathrm{G}}\log(\lambda_1) + p_{2\mathrm{G}}\log(\lambda_2) + p_{0\mathrm{G}}\log(\lambda_0) + p_{1\mathrm{G}}\log(\lambda_0+\lambda_2) \\
& + p_{2\mathrm{G}}\log(\lambda_0+\lambda_1) + (\alpha-1)\left(\mathrm{E}_{\mathrm{BGE}}[\log X_1\cdot 1_{A_1}] + \mathrm{E}_{\mathrm{BGE}}[\log X_1\cdot 1_{A_2}]\right) \\
& + (\alpha-1)\left(\mathrm{E}_{\mathrm{BGE}}[\log X_2\cdot 1_{A_1}] + \mathrm{E}_{\mathrm{BMOW}}[\log X_2\cdot 1_{A_2}] + \mathrm{E}_{\mathrm{BMOW}}[\log X_1\cdot 1_{A_0}]\right) \\
& - \lambda_1\left(\mathrm{E}_{\mathrm{BMOW}}[X_1^{\alpha}\cdot 1_{A_1}] + \mathrm{E}_{\mathrm{BMOW}}[X_1^{\alpha}\cdot 1_{A_2}] + \mathrm{E}_{\mathrm{BMOW}}[X_1^{\alpha}\cdot 1_{A_0}]\right) \\
& - \lambda_2\left(\mathrm{E}_{\mathrm{BMOW}}[X_2^{\alpha}\cdot 1_{A_1}] + \mathrm{E}_{\mathrm{BMOW}}[X_2^{\alpha}\cdot 1_{A_2}] + \mathrm{E}_{\mathrm{BMOW}}[X_1^{\alpha}\cdot 1_{A_0}]\right) \\
& - \lambda_0\left(\mathrm{E}_{\mathrm{BMOW}}[X_1^{\alpha}\cdot 1_{A_2}] + \mathrm{E}_{\mathrm{BMOW}}[X_2^{\alpha}\cdot 1_{A_1}] + \mathrm{E}_{\mathrm{BMOW}}[X_1^{\alpha}\cdot 1_{A_0}]\right).
\end{aligned}$$

In this case, we need to maximize $\Pi_2(\Gamma)$ with respect to Γ numerically to obtain $\widetilde{\Gamma}$, for a fixed Σ. Clearly, $\widetilde{\Gamma}$ depends on Σ, and we do not make it explicit for brevity.

Similarly, as before since maximization of $\Pi_2(\Gamma)$ involves a four dimensional optimization problem, we suggest to use the following approximation of $\Pi_2^*(\Gamma)$. We suggest to use

$$\begin{aligned}
\Pi_2^*(\Gamma) = {} & (p_{0\mathrm{G}} + 2p_{1\mathrm{G}} + 2p_{2\mathrm{G}})\log\alpha + (\alpha-1)\mathrm{E}\left[\log X_1\cdot 1_{A_0} + (\log X_1 + \log X_2)\cdot 1_{A_1\cup A_2}\right] \\
& - \lambda_0\mathrm{E}\left[X_1^{\alpha}\cdot 1_{A_0} + X_1^{\alpha}\cdot 1_{A_2} + X_2^{\alpha}\cdot 1_{A_1}\right] + (p_{0\mathrm{G}} + a_1 p_{1\mathrm{G}} + b_1 p_{2\mathrm{G}})\log(\lambda_0) \\
& - \lambda_1\mathrm{E}\left[X_1^{\alpha}\right] + (p_{1\mathrm{G}} + a_2 p_{2\mathrm{G}})\log(\lambda_1) - \lambda_2\mathrm{E}\left[X_2^{\alpha}\right] + (p_{2\mathrm{G}} + b_2 p_{1\mathrm{G}})\log(\lambda_1).
\end{aligned}$$

Here

$$a_1 = \frac{\alpha_1}{\alpha_0+\alpha_1}, \quad a_2 = \frac{\alpha_0}{\alpha_0+\alpha_2}, \quad b_1 = \frac{\alpha_2}{\alpha_0+\alpha_2}, \quad b_2 = \frac{\alpha_0}{\alpha_0+\alpha_2},$$

$p_{0\mathrm{G}}, p_{1\mathrm{G}}, p_{2\mathrm{G}}$ are same as defined before. The expressions of the different expectations are provided in Appendix.

It may be similarly observed as before that

$$\Pi_2^*(\Gamma) = \lim_{n\to\infty} \frac{1}{n} \mathrm{E}\left[l_{\text{pseudo}}(\alpha, \lambda_0, \lambda_1, \lambda_2 \mid (X_{1i}, X_{2i}); i = 1, \ldots, n)\right],$$

where $(X_{1i}, X_{2i}) \sim \mathrm{BGE}(\alpha_0, \alpha_1, \alpha_2, \lambda)$. The explicit expression of $l_{\text{pseudo}}(\cdot)$ is available in Kundu and Dey (2009).

The maximization of $\Pi_2^*(\Gamma)$ with respect to Γ can be performed quite easily. For fixed α, the maximization $\Pi_2^*(\Gamma)$ with respect to λ_1, λ_2 and λ_0 can be obtained for

$$\begin{aligned}
\widetilde{\lambda}_1 &= \frac{p_{1\mathrm{G}} + b_2 p_{2\mathrm{G}}}{\mathrm{E}\left[X_1^{\alpha}\right]}, \\
\widetilde{\lambda}_2 &= \frac{p_{2\mathrm{G}} + a_2 p_{1\mathrm{G}}}{\mathrm{E}\left[X_2^{\alpha}\right]}, \\
\widetilde{\lambda}_0 &= \frac{p_{0\mathrm{G}} + a_1 p_{1\mathrm{G}} + b_1 p_{2\mathrm{G}}}{\mathrm{E}\left[X_1^{\alpha} \cdot 1_{A_0}\right] + \mathrm{E}\left[X_1^{\alpha} \cdot 1_{A_2}\right] + \mathrm{E}\left[X_2^{\alpha} \cdot 1_{A_1}\right]},
\end{aligned}$$

respectively, and finally the maximization $\Pi_2^*(\Gamma)$ can be performed by maximizing the profile function $\Pi_2^*(\alpha, \widetilde{\lambda}_0(\alpha), \widetilde{\lambda}_1(\alpha), \widetilde{\lambda}_2(\alpha))$ with respect to α only.

6. Numerical Results

In this section, we perform some numerical experiments to observe how these asymptotic results work for different sample sizes, and for different parameter values. All these computations are performed at the Indian Institute of Technology Kanpur, using Intel(R) Core(TM)2 Quad CPU Q9550 2.83GHz, 3.23 GB RAM machines. The programs are written in R software (2.8.1), which can be obtained from the authors on request. We compute the PCS based on Monte Carlo (MC) simulation, and also based on the asymptotic results. We replicate the process 1000 times and compute the proportion of correct selection. For computing the PCS based on asymptotic results, first we compute the misspecified parameters and based on those misspecified parameters we compute the PCS.

6.1 Case 1: parent distribution is BMOW

In this case, we consider the following parameter sets:
Set 1: $\alpha = 2.0$, $\lambda_0 = 1.0$, $\lambda_1 = 1.0$, $\lambda_2 = 1.0$; Set 2: $\alpha = 1.5$, $\lambda_0 = 1.0$, $\lambda_1 = 1.0$, $\lambda_2 = 1.0$; Set 3: $\alpha = 1.5$, $\lambda_0 = 0.5$, $\lambda_1 = 0.5$, $\lambda_2 = 0.5$; Set 4: $\alpha = 1.5$, $\lambda_0 = 2.0$, $\lambda_1 = 1.0$, $\lambda_2 = 1.5$, and different sample sizes namely $n = 20$, 40, 60, 80, 100. For each parameter set and for each sample size, we have generated the sample from the BMOW distribution. Then, we compute the ML estimates of the unknown parameters and the values for the corresponding maximized log-likelihood functions, assuming that the data are coming from the BMOW or BGE distribution. In computing the ML estimates of the unknown parameters, we have used the EM algorithm as suggested in Kundu and Dey (2009) and Kundu and Gupta (2009), respectively. Finally, based on the values for the corresponding maximized log-likelihood functions, we decide whether we have made the correct decision or not. We replicate the process 1000 times, and compute the proportion of correct selection. The results are reported in the first rows of Tables 1 to 4.

Now, to compare these results with the corresponding asymptotic results, first we compute the misspecified parameters for each parameter set, and they are presented in the

Table 1. PCS based on MC simulations and based on asymptotic distribution (AD) for parameter Set 1.

n	20	40	60	80	100
MC	0.9255	0.9808	0.9953	0.9987	0.9997
AD	0.9346	0.9837	0.9956	0.9987	0.9996

Table 2. PCS based on MC simulations and based on AD for parameter Set 2.

n	20	40	60	80	100
MC	0.9255	0.9808	0.9953	0.9987	0.9997
AD	0.9212	0.9772	0.9928	0.9976	0.9992

Table 3. PCS based on MC simulations and based on AD for parameter Set 3.

n	20	40	60	80	100
MC	0.9073	0.9749	0.9914	0.9979	0.9989
AS	0.9204	0.9767	0.9926	0.9975	0.9992

Table 4. PCS selection based on MC simulations and based on AD for parameter Set 4.

n	20	40	60	80	100
MC	0.8834	0.9587	0.9843	0.9952	0.9973
AS	0.8996	0.9648	0.9866	0.9947	0.9979

following Table 5. In each case, we need to compute AM_{BMOW} and AV_{BMOW}, as defined in Theorem 4.1. Since the exact expressions of AM_{BMOW} and AV_{BMOW} are quite complicated, we have used simulation consistent estimates of AM_{BMOW} and AV_{BMOW}, which can be obtained very easily. The simulation consistent estimators of AM_{BMOW} and AV_{BMOW} are obtained using 10,000 replications, and they are reported in Table 6.

Table 5. Misspecified parameter values $\widetilde{\Sigma}$ for different parameter sets.

Set	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_0$	$\widetilde{\lambda}$
1	1.5098	1.5098	1.6228	2.81
2	0.8853	0.8853	0.9458	2.35
3	0.8908	0.8908	0.9600	1.49
4	1.3393	1.0782	0.7362	3.10

Table 6. AM_{BMOW} and AV_{BMOW} for different parameter sets.

Set	AM_{BMOW}	AV_{BMOW}
1	0.2346	0.4823
2	0.1982	0.3936
3	0.2297	0.4317
4	0.1762	0.4317

Now, using Theorem 4.1 and based on the asymptotic distribution of T and the discrimination statistic, we compute the PCS, i.e., $\text{P}(T > 0)$, for different sample sizes. The results are reported in the second rows of Tables 1 to 4 for all the parameter sets. It is very interesting to observe that for the bivariate case, even for small sample sizes the PCS are very high, and the asymptotic results match very well with the simulated results.

6.2 Case 2: parent distribution is BGE

In this case, we consider the parameter sets:
Set 5: $\alpha_0 = 1.5$, $\alpha_1 = 2.0$, $\alpha_2 = 1.0$, $\lambda = 1.0$; Set 6: $\alpha_0 = 1.0$, $\alpha_1 = 1.0$, $\alpha_2 = 1.0$, $\lambda = 1.0$;

Set 7: $\alpha_0 = 2.0$, $\alpha_1 = 2.0$, $\alpha_2 = 2.0$, $\lambda = 1.0$; Set 8: $\alpha_0 = 1.5$, $\alpha_1 = 1.5$, $\alpha_2 = 1.5$, $\lambda = 1.0$, and the same sample sizes as in Case 1. In this case, we generate the sample from the BGE distribution and using the same procedure as before we compute the proportion of correct selection. The results are reported in the first rows of Tables 7 to 10.

Table 7. PCS based on MC simulations and based on AD for parameter Set 5.

n	20	40	60	80	100
MC	0.9195	0.9797	0.9935	0.9986	0.9993
AS	0.9330	0.9830	0.9953	0.9986	0.9996

Table 8. PCS based on MC simulations and based on AD for parameter Set 6.

n	20	40	60	80	100
MC	0.9001	0.9701	0.9892	0.9962	0.9984
AS	0.9153	0.9741	0.9914	0.9970	0.9989

Table 9. PCS based on MC simulations and based on AD for parameter Set 7.

n	20	40	60	80	100
MC	0.9189	0.9811	0.9944	0.9987	0.9994
AS	0.9347	0.9837	0.9955	0.9987	0.9996

Table 10. PCS based on MC simulations and based on AD for parameter Set 8.

n	20	40	60	80	100
MC	0.9096	0.9768	0.9929	0.9975	0.9991
AS	0.9299	0.9816	0.9947	0.9984	0.9995

Now, to compute the asymptotic PCS, first we compute the misspecified parameters as suggested in Section 5, and they are reported in Table 11. We also report simulated consistent estimates of AM_{BGE} and AV_{BGE} in Table 12.

Table 11. Misspecified parameter values $\widetilde{\Gamma}$ for different parameter sets.

Set	$\widetilde{\alpha}$	$\widetilde{\lambda}_0$	$\widetilde{\lambda}_1$	$\widetilde{\lambda}_2$
5	1.6199	0.1732	0.1137	0.1992
6	1.4199	0.2575	0.2418	0.2418
3	1.8200	0.1123	0.1050	0.1050
8	1.6199	0.1665	0.1553	0.1553

Table 12. AM_{BMOW} and AV_{BMOW} for different parameter sets.

Set	AM_{BMOW}	AV_{BMOW}
5	0.2224	0.4406
6	0.1967	0.4095
3	0.2316	0.4692
8	0.2128	0.4157

Now, similarly as before, based on the asymptotic distribution of T, as provided in Theorem 4.2, we compute the PCS in this case, i.e., $\text{P}(T < 0)$, for different sample sizes. We report the results in the second rows of Tables 7 to 10 for all the parameter sets. In this case, it observed that the asymptotic results match extremely well with the simulated results.

7. Data Analysis

In this section, we present the analysis of a real data set for illustrative purposes. These data are from the National Football League (NFL), American Football, matches played on three consecutive weekends in 1986. It has been originally published in “Washington Post”. In this bivariate data set, the variables are the “game time” to the first points scored by kicking the ball between goal posts (X_1) and the “game time” to the first points scored by moving the ball into the end zone (X_2). 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 data (scoring times in minutes and seconds) are represented in Table 13. We have analyzed the data by converting the seconds to the decimal minutes, i.e., 2:03 has been converted to 2.05.

Table 13. American Football League (NFL) data.

X_1	X_2	X_1	X_2	X_1	X_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

The variables X_1 and X_2 have the 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, it is clear that, in this case, $X_1 = X_2$ occurs with positive probability, and some singular distribution should be used to analyze this data set.

If we define the random variables

$$U_1 = \text{time to first field goal},$$
$$U_2 = \text{time to first safety or unconverted touchdown},$$
$$U_0 = \text{time to first converted touchdown}.$$

Then, $X_1 = \min\{U_0, U_1\}$ and $X_2 = \min\{U_0, U_2\}$. Therefore, (X_1, X_2) has a similar structure as the bivariate Marshall-Olkin exponential model. Csorgo and Welsh (1989) analyzed the data using the bivariate Marshall-Olkin exponential model but concluded that it does not work well, because X_2 may be exponential but X_1 is not. In fact it is observed that the empirical HFs of both X_1 and X_2 are increasing functions.

Since both BMOW and BGE distributions can have increasing marginal HFs, we fit both the models to the data set. For the BMOW distribution, using EM algorithm as suggested in Kundu and Dey (2009), we compute the ML estimates of the unknown parameters as $\widehat{\alpha} =$

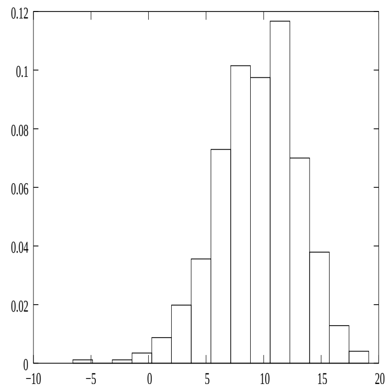

Figure 3. Histogram of the bootstrap sample of the discrimination statistic.

1.2889, $\widehat{\lambda}_0 = 11.2073$, $\widehat{\lambda}_1 = 8.3572$, $\widehat{\lambda}_2 = 0.4720$, and the associated 95% confidence intervals are (1.0372, 1.5406), (5.7213, 16.6932), (2.5312, 14.1831), (-0.4872, 1.4314) respectively. The value for the corresponding maximized likelihood function is 47.8041. In case of the BGE distribution using the EM algorithm as suggested in Kundu and Gupta (2009), we obtained the ML estimates of the unknown parameters as $\widehat{\alpha}_0 = 1.1628$, $\widehat{\alpha}_1 = 0.0558$, $\widehat{\alpha}_2 = 0.5961$, $\widehat{\lambda} = 9.5634$, and the associated 95% confidence intervals are (0.6991, 1.6266), (-0.0205, 0.1322), (0.2751, 0.9171) and (6.5298, 12.5970) respectively. The value for the corresponding maximized likelihood function is 38.0042. Therefore, based on the values for the corresponding maximized likelihood function, we prefer to use the BMOW model rather than the BGE model to analyze this data set.

Now, to compute the PCS in this case, we perform non-parametric bootstrap. The histogram of the bootstrap sample of the discrimination statistic is provided in Figure 3. Based on one thousand bootstrap replications, it is observed that the PCS is 0.98.

8. Conclusion

In this paper, we have considered discrimination between two singular bivariate models, namely the BMOW and BGE distributions. Both the distributions have singular part and absolute continuous part. The difference of the values for the corresponding maximized likelihood function has been used as the discrimination statistic. We have obtained the asymptotic distribution of the discrimination statistic, which can be used to compute the asymptotic PCS. MC simulations are performed to see the behavior of the proposed method. It is known that the discrimination between Weibull and generalized exponential distributions is quite difficult (see Gupta and Kundu, 2003), but in this paper it is observed that the discrimination between the BMOW and BGE distributions is relatively easier. Even with small sample sizes the PCS quite high. Moreover, the asymptotic PCS matches very well with the simulated PCS even for moderate sample sizes. We have performed the analysis of a data set and computed the PCS using non-parametric bootstrap method. Although we do not have any theoretical results, it seems non-parametric bootstrap method also can be used quite effectively in computing the PCS in this case. More work is needed in this direction.

Appendix

To prove Theorem 4.1, we need of Lemma 8.1. Here $\overset{\text{a.s.}}{\rightarrow}$ means converges almost surely.

Lemma 8.1 Under the assumption that data are from the BWE$(\alpha, \lambda_0, \lambda_1, \lambda_2)$ distribution, as $n \to \infty$, we have
(i) $\widehat{\alpha} \xrightarrow{\text{a.s.}} \alpha$, $\widehat{\lambda}_0 \xrightarrow{\text{a.s.}} \lambda_0$, $\widehat{\lambda}_1 \xrightarrow{\text{a.s.}} \lambda_1$ and $\widehat{\lambda}_2 \xrightarrow{\text{a.s.}} \lambda_2$ where for $\Gamma = (\alpha, \lambda_0, \lambda_1, \lambda_2)$,

$$\mathrm{E}_{\mathrm{BMOW}}[\log(f_{\mathrm{BMOW}}(X_1, X_2; \Gamma))] = \max_{\bar{\Gamma}} \mathrm{E}_{\mathrm{BMOW}}[\log(f_{\mathrm{BMOW}}(X_1, X_2; \bar{\Gamma}))];$$

(ii) $\widehat{\alpha}_0 \xrightarrow{\text{a.s.}} \widetilde{\alpha}_0$, $\widehat{\alpha}_1 \xrightarrow{\text{a.s.}} \widetilde{\alpha}_1$, $\widehat{\alpha}_2 \xrightarrow{\text{a.s.}} \widetilde{\alpha}_2$, $\widehat{\lambda} \xrightarrow{\text{a.s.}} \widetilde{\lambda}$, where for $\Sigma = (\alpha_0, \alpha_1, \alpha_2, \lambda)$,

$$\mathrm{E}_{\mathrm{BMOW}}[\log(f_{\mathrm{BGE}}(X_1, X_2; \widetilde{\Sigma}))] = \max_{\Sigma} \mathrm{E}_{\mathrm{BMOW}}[\log(f_{\mathrm{BGE}}(X_1, X_2; \Sigma))].$$

It may be noted that $\widetilde{\Sigma}$ may depend on Γ, but we do not make it explicit for brevity;
(iii) If we denote

$$T^* = L_2(\alpha, \lambda_0, \lambda_1, \lambda_2) - L_1(\widetilde{\alpha}_0, \widetilde{\alpha}_1, \widetilde{\alpha}_2, \widetilde{\lambda}),$$

then $n^{-\frac{1}{2}}(T - \mathrm{E}_{\mathrm{BMOW}}[T])$ is asymptotically equivalent to $n^{-\frac{1}{2}}(T^* - \mathrm{E}_{\mathrm{BMOW}}[T^*])$.

Proof of Lemma 8.1 It is quite standard and it follows along the same line as the proof of Lemma 2.2 of White (1982), and it is avoided. ■

Proof of Theorem 4.1 Using Central limit theorem and part (ii) of Lemma 8.1, it follows that $n^{-\frac{1}{2}}(T^* - \mathrm{E}_{BWE}[T^*])$ is asymptotically normally distributed with mean zero and variance $\mathrm{V}_{\mathrm{BMOW}}[T^*]$. Therefore, using part (iii) of Lemma 8.1, the result immediately follows. ■

To prove Theorem 4.2 and for defining the misspecified parameter $\widetilde{\Gamma}$, we need of Lemma 8.2, whose proof is same as the proof of Lemma 8.1.

Lemma 8.2 Suppose the data follow the BGE$(\alpha_0, \alpha_1, \alpha_2, \lambda)$ distribution, as $n \to \infty$, we have
(i) $\widehat{\alpha}_0 \xrightarrow{\text{a.s.}} \alpha_0$, $\widehat{\alpha}_1 \xrightarrow{\text{a.s.}} \alpha_1$, $\hat{\alpha}_2 \xrightarrow{\text{a.s.}} \alpha_2$ and $\hat{\lambda} \to \lambda$ where

$$\mathrm{E}_{\mathrm{BGE}}[\log(f_{\mathrm{BGE}}(X_1, X_2; \Sigma))] = \max_{\bar{\Sigma}} \mathrm{E}_{\mathrm{BGE}}[\log(f_{\mathrm{BGE}}(X_1, X_2; \bar{\Sigma}))];$$

(ii) $\widehat{\alpha} \xrightarrow{\text{a.s.}} \widetilde{\alpha}$, $\widehat{\lambda}_0 \xrightarrow{\text{a.s.}} \widetilde{\lambda}_0$, $\widehat{\lambda}_1 \xrightarrow{\text{a.s.}} \widetilde{\lambda}_1$, $\widehat{\lambda}_2 \xrightarrow{\text{a.s.}} \widetilde{\lambda}_2$, where $\widetilde{\Gamma} = (\widetilde{\alpha}, \widetilde{\lambda}_0, \widetilde{\lambda}_1, \widetilde{\lambda}_2)$,

$$\mathrm{E}_{\mathrm{BGE}}[\log(f_{\mathrm{BMOW}}(X_1, X_2; \widetilde{\Gamma}))] = \max_{\Gamma} \mathrm{E}_{\mathrm{BGE}}[\log(f_{\mathrm{BMOW}}(X_1, X_2; \Gamma))];$$

here also $\widetilde{\Gamma}$ depend on Σ, but we do not make it explicit for brevity;
(iii) If we denote

$$T_* = L_2(\widetilde{\alpha}, \widetilde{\lambda}_0, \widetilde{\lambda}_1, \widetilde{\lambda}_2) - L_1(\alpha_0, \alpha_1, \alpha_2, \lambda),$$

then $n^{-\frac{1}{2}}(T - \mathrm{E}_{\mathrm{BGE}}[T])$ is asymptotically equivalent to $n^{-\frac{1}{2}}(T_* - \mathrm{E}_{\mathrm{BGE}}[T^*])$.

Proof of Theorem 4.2 Along the same line as the Proof of Lemma 8.1, it also follows using Lemma 8.2. ■

The following lemmas are useful in computing the different expected values needed in $\Pi_1^*(\Sigma)$ and in $\Pi_2^*(\Gamma)$. Here 1_{A_0}, 1_{A_1} and 1_{A_2} are same as defined before.

Lemma A.1 Let $W_0 \sim \mathrm{GE}(\alpha_0 + \alpha_1 + \alpha_2, \lambda)$, $W_1 \sim \mathrm{GE}(\alpha_0 + \alpha_1, \lambda)$, $W_2 \sim \mathrm{GE}(\alpha_0 + \alpha_2, \lambda)$ and $(X_1, X_2) \sim \mathrm{BGE}(\alpha_0, \alpha_1, \alpha_2, \lambda)$. If $g(\cdot)$ is any Borel measurable function, then

$$\mathrm{E}[g(X_1) \cdot 1_{A_1}] = \mathrm{E}[g(W_1)] + \frac{\alpha_0 + \alpha_1}{\alpha_0 + \alpha_1 + \alpha_2}\mathrm{E}[g(W_0)].$$

$$\mathrm{E}[g(X_1) \cdot 1_{A_2}] = \frac{\alpha_1}{\alpha_0 + \alpha_1 + \alpha_2}\mathrm{E}[g(W_0)].$$

$$\mathrm{E}[g(X_1) \cdot 1_{A_0}] = \mathrm{E}[g(X_2) \cdot 1_{A_0}] = \frac{\alpha_0}{\alpha_0 + \alpha_1 + \alpha_2}\mathrm{E}[g(W_0)].$$

$$\mathrm{E}[g(X_2) \cdot 1_{A_1}] = \frac{\alpha_2}{\alpha_0 + \alpha_1 + \alpha_2}\mathrm{E}[g(W_0)].$$

$$\mathrm{E}[g(X_2) \cdot 1_{A_2}] = \mathrm{E}[g(W_2)] + \frac{\alpha_0 + \alpha_2}{\alpha_0 + \alpha_1 + \alpha_2}\mathrm{E}[g(W_0)].$$

Proof of Lemma A.1 See Kundu and Gupta (2009). ■

Lemma A.2 Let $Z_0 \sim \mathrm{WE}(\alpha,\, \lambda_0 + \lambda_1 + \lambda_2)$, $Z_1 \sim \mathrm{WE}(\alpha,\, \lambda_0 + \lambda_1)$, $Z_2 \sim \mathrm{WE}(\alpha,\, \lambda_0 + \lambda_2)$ and $(X_1, X_2) \sim \mathrm{BMOW}(\alpha, \lambda_0, \lambda_1, \lambda_2)$. If $g(\cdot)$ is any Borel measurable function, then

$$\mathrm{E}[g(X_1) \cdot 1_{A_1}] = \frac{\lambda_1}{\lambda_0 + \lambda_1 + \lambda_2}\mathrm{E}[g(Z_1)].$$

$$\mathrm{E}[g(X_1) \cdot 1_{A_2}] = \mathrm{E}[g(Z_1)] - \frac{\lambda_0 + \lambda_1}{\lambda_0 + \lambda_1 + \lambda_2}\mathrm{E}[g(Z_0)].$$

$$\mathrm{E}[g(X_1) \cdot 1_{A_0}] = \mathrm{E}[g(X_2) \cdot 1_{A_0}] = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}\mathrm{E}[g(Z_0)].$$

$$\mathrm{E}[g(X_2) \cdot 1_{A_1}] = \mathrm{E}[g(Z_2)] - \frac{\lambda_0 + \lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}.$$

$$\mathrm{E}[g(X_2) \cdot 1_{A_2}] = \frac{\lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}\mathrm{E}[g(Z_2)].$$

Proof of Lemma A.1 They can be obtained along the same line as in Lemma A.1. ■

References

Atkinson, A., 1969. A test for discriminating between models. Biometrika, 56, 337–341.
Atkinson, A., 1970. A method for discriminating between models (with discussions). Journal of The Royal Statistical Society Series B - Statistical Methodology, 32, 323–353.
Bain, L.J., Englehardt, M., 1980. Probability of correct selection of Weibull versus gamma based on likelihood ratio test. Communications in Statistics - Theory and Methods, 9, 375–381.
Bemis, B., Bain, L.J., Higgins, J.J., 1972. Estimation and hypothesis testing for the parameters of a bivariate exponential distribution. Journal of the American Statistical Association, 67, 927–929.
Cox, D.R., 1961. Tests of separate families of hypotheses. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, pp. 105–123.
Cox, D.R., 1962. Further results on tests of separate families of hypotheses. Journal of The Royal Statistical Society Series B - Statistical Methodology, 24, 406–424.

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

Dey, A.K., Kundu, D., 2009. Discriminating among the log-normal, Weibull and generalized exponential distributions. IEEE Transactions on Reliability, 58, 416–424.

Dey, A.K., Kundu, D., 2010. Discriminating between the log-normal and log-logistic distributions. Communications in Statistics - Theory and Methods, 39, 280–292.

Gupta, R.D., Kundu, D., 1999. Generalized exponential distributions. Australian and New Zealand Journal of Statistics, 41, 173–188.

Gupta, R.D., Kundu, D., 2003. Discriminating between Weibull and generalized exponential distributions. Computational Statistics and Data Analysis, 43, 179–196.

Gupta, R.D., Kundu, D., 2006. On the comparison of Fisher information of the Weibull and GE distributions. Journal of Statistical Planning and Inference, 136, 3130–3144.

Gupta, R.D., Kundu, D., 2007. Generalized exponential distribution: existing methods and recent developments. Journal of Statistical Planning and Inference, 137, 3537–3547.

Kotz, S., Balakrishnan, N., Johnson, N., 2000. Continuous Multivariate Distributions: Models and Applications. Wiley and Sons, New York.

Kundu, D., Dey, A.K., 2009. Estimating the parameters of the Marshall-Olkin bivariate Weibull distribution by EM algorithm. Computational Statistics and Data Analysis, 53, 956–965.

Kundu, D., Gupta, R.D., 2009. Bivariate generalized exponential distribution. Journal of Multivariate Analysis, 100, 581–593.

Marshall, A.W., Meza, J.C., Olkin, I., 2001. Can data recognize its parent distribution? Journal of Computational and Graphical Statistics, 10, 555–580.

White, H., 1982. Nonlinear regression on cross-section data. Econometrica, 48, 721–746.