Discriminating Between the Bivariate Generalized Exponential and Bivariate Weibull Distributions

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

Abstract

Recently Kundu and Gupta ("Bivariate generalized exponential distribution", *Journal of Multivariate Analysis*, vol. 100, 581 - 593, 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 Marshall-Olkin bivariate Weibull distribution. The two singular bivariate distributions functions have very similar joint probability density functions. In this paper we consider the discrimination between the 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.

Key Words and Phrases: Likelihood ratio test; asymptotic distribution; probability of correct selection; Monte Carlo simulations; EM algorithm; maximum likelihood estimator.

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

[‡] Corresponding author. e-mail: kundu@iitk.ac.in

1 Introduction

Recently, the two-parameter generalized exponential (GE) distribution proposed by Gupta and Kundu (1999) has received some attention. The two-parameter GE distribution, which has one shape parameter, and one scale parameter is a positively skewed distribution. It has several desirable properties, and many of its properties 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 distribution and Weibull distribution are very similar, and both the distributions have nice compact distribution functions. Both the distributions contain exponential distribution as a special case, therefore, they are extensions of the exponential distributions but in different manners. It is further observed that the GE distribution also can be used quite successfully in analyzing positively skewed data sets, in place of Weibull distribution, and often it is very difficult to distinguish between the two. For some recent developments on GE distribution, and for its different applications, the readers are referred to the review article by Gupta and Kundu (2007).

The problem of testing whether some given observations follow one of the two (or more) probability distribution functions, is quite an old statistical problem. Cox (1961), see also Cox (1962), first considered this problem in his classical paper, and he also discussed the effect of choosing the wrong model. Since then extensive work has been done in discriminating between two or more distribution functions, see for example Atkinson (1969, 1970), Bain and Englehardt (1980), Marshall, Meza and Olkin (2001), Dey and Kundu (2009, 2010) and the references cited therein.

In recent times it is observed, see Gupta and Kundu (2003, 2006), that due to the closeness between Weibull and GE distributions, it is extremely difficult to discriminate between these two distribution functions. Note that if the shape parameter is one, the two distribution functions are not distinguishable. For small sample sizes the probability of correct selection (PCS) can be quite low 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 distribution functions, but no work has been found in discriminating between two bivariate distribution functions.

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

In this paper we consider discriminating between BVGE and MOBW distributions. We have used the difference of the maximized log-likelihood values in discriminating between the two distribution functions. 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 has been used to compute the probability of correct selection (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 a one dimensional optimization problem only. Therefore, computationally it 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. We perform the analysis of a data set for illustrative purposes.

Rest of the paper is organized as follows. In Section 2, we briefly discuss about the BVGE and MOBW distributions. The discrimination procedure is presented in Section 3. The asymptotic distribution of the test statistics for both the cases are provided in Section 4. The calculation of the misspecified parameters are discussed in Section 5. In Section 6, we present the Monte Carlo simulation results and the analysis of a data set is presented in Section 7. Finally we conclude the paper in Section 8.

2 MOBW and BVGE Distributions

In this section we will briefly discuss about the MOBW and BVGE distributions. We will be using 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 the following probability density function (PDF), for $x > 0$;

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

the corresponding cumulative distribution function (CDF) and survival function (SF) are

$$F_{WE}(x; \alpha, \lambda) = 1 - e^{-\lambda x^{\alpha}}, \quad \text{and} \quad S_{WE}(x; \alpha, \lambda) = e^{-\lambda x^{\alpha}}, \tag{2}$$

respectively. From now on a Weibull distribution with the PDF as given in (1) will be denoted by WE(α, λ). The GE distribution with the shape parameter $\alpha > 0$ and the scale parameter $\lambda > 0$, has the PDF

$$f_{GE}(x; \alpha, \lambda) = \alpha\lambda e^{-\lambda x} \left(1 - e^{-\lambda x}\right)^{\alpha-1}. \tag{3}$$

The corresponding CDF and SF are

$$F_{GE}(x;\alpha,\lambda) = \left(1-e^{-\lambda x}\right)^{\alpha}, \quad \text{and} \quad S_{GE}(x;\alpha,\lambda) = 1-\left(1-e^{-\lambda x}\right)^{\alpha} \tag{4}$$

respectively. A GE distribution with the PDF given in (3) will be denoted by GE(α, λ).

2.1 MOBW Distribution

Suppose $U_0 \sim \text{WE}(\alpha, \lambda_0)$, $U_1 \sim \text{WE}(\alpha, \lambda_1)$ and $U_2 \sim \text{WE}(\alpha, \lambda_2)$ 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 MOBW distribution with parameters $\alpha, \lambda_0, \lambda_1, \lambda_2$, and it will be denoted as MOBW(Γ), where $\Gamma = (\alpha, \lambda_0, \lambda_1, \lambda_2)$.

If $(X_1, X_2) \sim \text{MOBW}(\Gamma)$, then their joint SF takes the following form, for $z = \max\{x_1, x_2\}$,

$$\begin{aligned} S_{MOBW}(x_1,x_2;\Gamma) &= P(X_1 > x_1, X_2 > x_2) = P(U_1 > x_1, U_2 > x_2, U_0 > z) \\ &= S_{WE}(x_1;\alpha,\lambda_1)S_{WE}(x_2;\alpha,\lambda_2)S_{WE}(z;\alpha,\lambda_0). \end{aligned} \tag{5}$$

The joint PDF of (X_1, X_2) can be written as

$$f_{MOBW}(x_1,x_2;\Gamma) = \begin{cases} f_{1W}(x_1,x_2;\Gamma) & \text{if} \quad 0 < x_1 < x_2 \\ f_{2W}(x_1,x_2;\Gamma) & \text{if} \quad 0 < x_2 < x_1 \\ f_{0W}(x;\Gamma) & \text{if} \quad 0 < x_1 = x_2 = x, \end{cases} \tag{6}$$

where

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

Note that the function $f_{MOBW}(\cdot)$ may be considered to be a density function for MOBW distribution, if it is understood that the first two terms are densities with respect to two

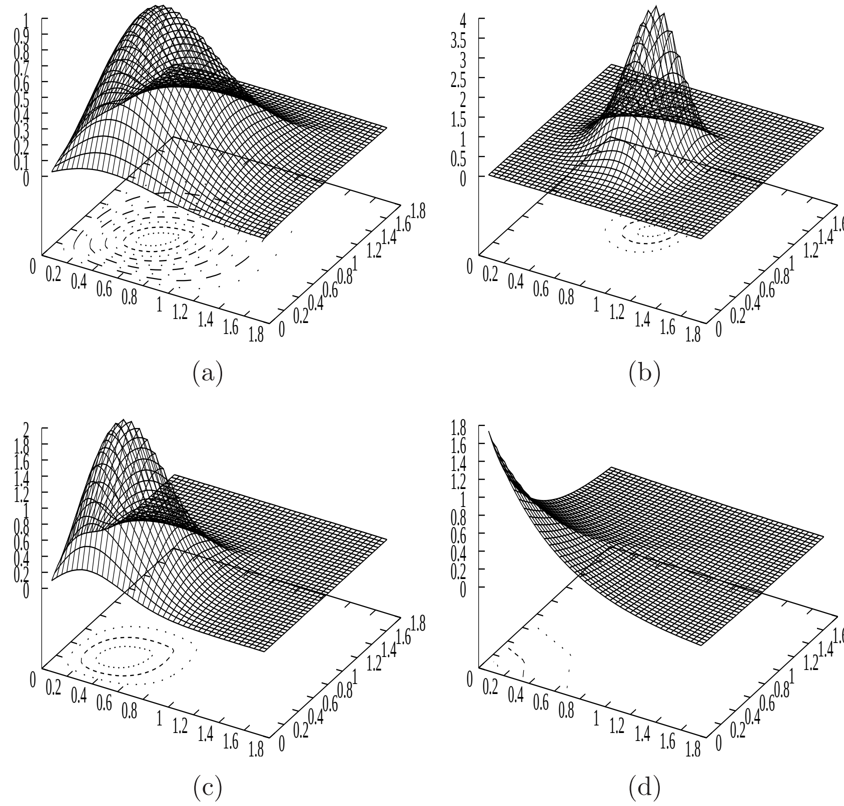

Figure 1: Surface plots of the absolute continuous part of the joint PDF of MOBW for $(\alpha, \lambda_1, \lambda_2, \lambda_3):$ (a) (2.0, 1.0, 1.0, 1.0) (b) (5.0, 1.0, 1.0, 1.0) (c) (2.0, 2.0, 2.0, 2.0) (d) (1.0, 1.0, 1.0, 1.0).

dimensional Lebesgue measure, and the third term is a density function with respect to a one dimensional Lebesgue measure, see for example Bemis *et al.* (1972). It is clear that MOBW 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 PDF of MOBW distribution can take variety of shapes, therefore, it can be used quite effectively in analyzing singular bivariate data.

The following probabilities will be used later in deriving the asymptotic probability of correct selection. If $(X_1, X_2) \sim \text{MOBW}(\Gamma)$, then

$$\begin{aligned} p_{1W} = P[X_1 < X_2] &= \int_0^\infty \int_0^y f_{WE}(x; \alpha, \lambda_1) f_{WE}(y; \alpha, \lambda_0 + \lambda_2) dx dy \\ &= \frac{\lambda_1}{\lambda_0 + \lambda_1 + \lambda_2}, \qquad (7) \\ p_{2W} = P[X_1 > X_2] &= \int_0^\infty \int_y^\infty f_{WE}(x; \alpha, \lambda_0 + \lambda_1) f_{WE}(y; \alpha, \lambda_2) dx dy \\ &= \frac{\lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}, \qquad (8) \\ p_{0W} = P[X_1 = X_2] &= \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \int_0^\infty f_{WE}(z; \alpha, \lambda_0 + \lambda_1 + \lambda_2) dz \\ &= \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}. \qquad (9) \end{aligned}$$

2.2 BVGE Distribution

Suppose $V_0 \sim \text{GE}(\alpha_0, \lambda)$, $V_1 \sim \text{GE}(\alpha_1, \lambda)$ and $V_2 \sim \text{GE}(\alpha_2, \lambda)$. 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 BVGE distribution with parameters $\alpha_0, \alpha_1, \alpha_2, \lambda$, and it will be denoted by BVGE(Σ), where $\Sigma = (\alpha_0, \alpha_1, \alpha_2, \lambda)$. It is immediate $Y_1 \sim$ GE $(\alpha_0 + \alpha_1, \lambda)$, and $Y_2 \sim \text{GE}(\alpha_0 + \alpha_2, \lambda)$. The joint CDF of (Y_1, Y_2) can be expressed as follows for $v = \min\{y_1, y_2\}$.

$$\begin{aligned} F_{BVGE}(y_1, y_2; \Sigma) &= P(Y_1 \le y_1, Y_2 \le y_2) = P(V_1 \le y_1, V_2 \le y_2, V_0 \le v) \\ &= (1 - e^{-\lambda y_1})^{\alpha_1} (1 - e^{-\lambda y_2})^{\alpha_2} (1 - e^{-\lambda v})^{\alpha_0}. \qquad (10) \end{aligned}$$

In this case also, the joint CDF of Y_1 and Y_2 can be written as

$$f_{BVGE}(y_1, y_2; \Sigma) = \begin{cases} f_{1G}(y_1, y_2) & \text{if} \quad 0 < y_1 < y_2 \\ f_{2G}(y_1, y_2) & \text{if} \quad 0 < y_2 < y_1 \\ f_{0G}(y) & \text{if} \quad 0 < y_1 = y_2 = y, \end{cases} \qquad (11)$$

where

$$f_{1G}(y_1, y_2; \Sigma) = f_{GE}(y_1; \alpha_0 + \alpha_1, \lambda) f_{GE}(y_2; \alpha_2, \lambda)$$

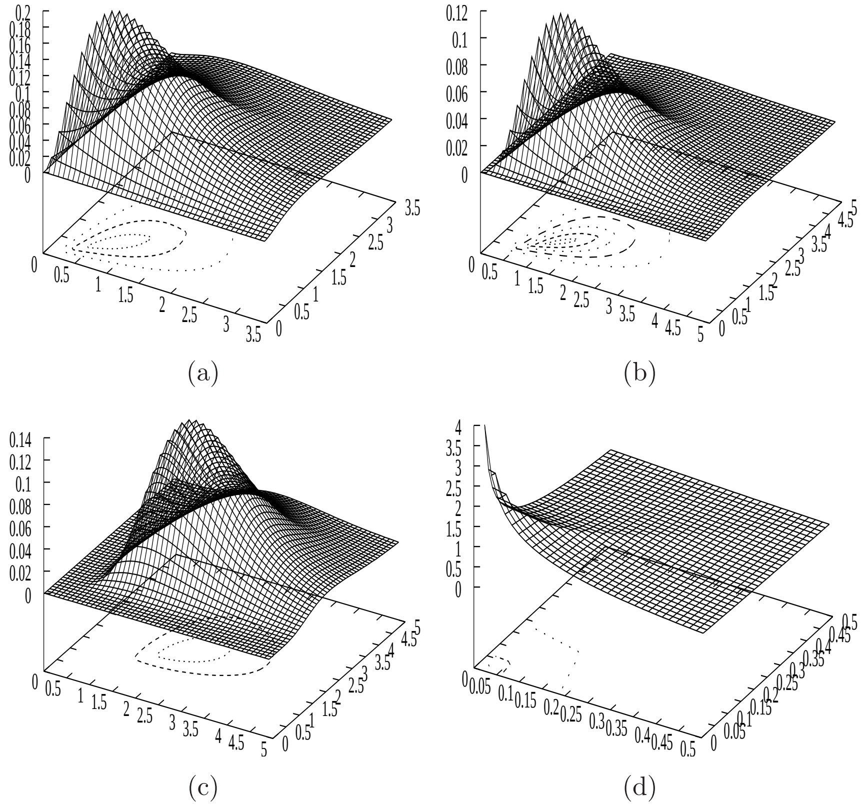

Figure 2: Surface plots of the absolute continuous part of the joint PDF of BVGE for $(\alpha_1, \alpha_2, \alpha_3, \lambda)$: (a) (1.0, 1.0, 2.0, 1.0) (b) (1.0, 1.0, 1.0, 4.0) (c) (5.0, 5.0, 5.0, 1.0) (d) (0.5, 0.5, 0.5, 1.0).

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

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

The following probabilities will be needed later. If $(Y_1, Y_2) \sim \text{BVGE}(\Sigma)$, then

$$\begin{aligned} p_{1G} = P[Y_1 < Y_2] &= \int_0^\infty \int_0^y f_{GE}(x; \alpha_0 + \alpha_1, \lambda) f_{GE}(y; \alpha_2, \lambda) dx dy \\ &= \frac{\alpha_2}{\alpha_0 + \alpha_1 + \alpha_2}, \end{aligned} \tag{12}$$

$$
\begin{aligned}
p_{2G} = P[Y_1 > Y_2] &= \int_0^\infty \int_y^\infty f_{GE}(x;\alpha_1,\lambda) f_{GE}(y;\alpha_0+\alpha_2,\lambda)\,dx\,dy \\
&= \frac{\alpha_1}{\alpha_0+\alpha_1+\alpha_2}, \qquad (13)\\
p_{0G} = P[Y_1 = Y_2] &= \frac{\alpha_0}{\alpha_0+\alpha_1+\alpha_2}\int_0^\infty f_{WE}(z;\alpha_0+\alpha_1+\alpha_2,\lambda)\,dz \\
&= \frac{\alpha_0}{\alpha_0+\alpha_1+\alpha_2}. \qquad (14)
\end{aligned}
$$

3 Discrimination Procedure

Suppose $\{(X_{11}, X_{21}), \cdots, (X_{1n}, X_{2n})\}$ is a random bivariate sample of size n generated either form a BVGE(Σ) or from a MOBW(Γ). Based on the above sample, we want to decide from which distribution function the data set has been obtained. We use the following notations and sets for the rest of the paper; $I_0 = \{(x_{1i}, x_{2i}), x_{1i} = x_{2i} = x_i, i = 1\cdots,n\}$, $I_1 = \{(x_{1i}, x_{2i}), x_{1i} < x_{2i}, i = 1\cdots,n\}$, $I_2 = \{(x_{1i}, x_{2i}), x_{1i} > x_{2i}, i = 1\cdots,n\}$, $I = I_0 \cup I_1 \cup I_2$, $n_0 = |I_0|$, $n_1 = |I_1|$ 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 $\widehat{\Gamma} = (\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\lambda})$ be the maximum likelihood estimators (MLEs) of Σ, based on the assumption that the data have been obtained from BVGE(Σ). Similarly, let $\widehat{\Gamma} = (\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1, \widehat{\lambda}_2)$, be the MLE of Γ based on the assumption that the data have been obtained from MOBW(Γ). 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 of the BVGE can be written as;

$$
\begin{aligned}
L_1(\Sigma) &= (n_0 + 2n_1 + 2n_2)\ln\lambda + n_1\ln(\alpha_0+\alpha_1) + n_1\ln\alpha_2 \qquad (15)\\
&+ (\alpha_0+\alpha_1-1)\sum_{i\in I_1}\ln(1-e^{-\lambda x_{1i}}) + (\alpha_2-1)\sum_{i\in I_1}\ln(1-e^{-\lambda x_{2i}})\\
&+ n_2\ln\alpha_1 + n_2\ln(\alpha_0+\alpha_2) + (\alpha_1-1)\sum_{i\in I_2}\ln(1-e^{-\lambda x_{1i}})\\
&+ (\alpha_0+\alpha_2-1)\sum_{i\in I_2}\ln(1-e^{-\lambda x_{2i}}) + n_0\ln\alpha_0\\
&+ (\alpha_0+\alpha_1+\alpha_2-1)\sum_{i\in I_0}\ln(1-e^{-\lambda x_i}) - \lambda\Big(\sum_{i\in I_0} x_i + \sum_{i\in I_1\cup I_2} x_{1i} + \sum_{i\in I_1\cup I_2} x_{2i}\Big) \qquad (16)
\end{aligned}
$$

and the log-likelihood of BVWE can be written as

$$\begin{aligned}
L_2(\Gamma) &= (n_0 + 2n_1 + 2n_2)\ln\alpha + n_1 \ln\lambda_1 + n_2 \ln\lambda_2 + n_0 \ln\lambda_0 + n_1 \ln(\lambda_0+\lambda_2) \\
&+ n_2 \ln(\lambda_0+\lambda_1) + (\alpha-1)\left[\sum_{i\in I_0} \ln x_{1i} + \sum_{i\in I_1\cup I_2} \ln x_{2i} + \sum_{i\in I_0} \ln x_i\right] \\
&- \lambda_1 \left[\sum_{i\in I_1\cup I_2} x_{1i}^{\alpha} + \sum_{i\in I_0} x_i^{\alpha}\right] - \lambda_2 \left[\sum_{i\in I_1\cup I_2} x_{2i}^{\alpha} + \sum_{i\in I_0} x_i^{\alpha}\right] - \lambda_0 \left[\sum_{i\in I_2} x_{1i}^{\alpha} + \sum_{i\in I_1} x_{2i}^{\alpha} + \sum_{i\in I_0} x_i^{\alpha}\right].
\end{aligned} \tag{17}$$

We use the following discrimination procedure. Consider the following statistics:

$$T = L_2(\widehat{\alpha}, \widehat{\lambda}_0, \widehat{\lambda}_1, \widehat{\lambda}_2) - L_1(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\lambda}). \tag{18}$$

If $T > 0$, we choose MOBW distribution, otherwise we prefer BVGE 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 (16) and (17) respectively. Computationally both are quite challenging problems, and to maximize directly one needs to solve a four dimensional optimization problem in each case. In both the cases EM algorithm can be used quite effectively to compute the MLEs of the unknown parameters. See for example Kundu and Gupta (2009) and Kundu and Dey (2009) for BVGE and MOBW 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 will help to compute the asymptotic PCS.

4 Asymptotic Distributions

In this section, we use the following notations. For any functions, $f_1(U)$ and $f_2(U)$, $E_{BVGE}(f_1(U))$, $V_{BVGE}(f_1(U))$ and $Cov_{BVGE}(f_2(U), f_1(U))$ will 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 $U \sim BVGE(\Sigma)$. Similarly, we define $E_{BVWE}(f_1(U))$, $V_{BVWE}(f_1(U))$ and $Cov_{BVWE}(f_2(U), f_1(U))$

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 $U \sim BVWE(\Gamma)$. We have the following two main results:

THEOREM 1: Under the assumption that data are from MOBW$(\alpha, \lambda_0, \lambda_1, \lambda_2)$ the distribution of T as defined in (18), is approximately normally distributed with mean $E_{MOBW}(T)$ and variance $V_{MOBW}(T)$. The expressions of $E_{MOBW}(T)$ and $V_{MOBW}(T)$ are provided below.

PROOF OF THEOREM 1: It is provided in the Appendix. ∎

Now we provide the expressions for $E_{MOBW}(T)$ and $V_{MOBW}(T)$. We denote

$$\lim_{n\to\infty} \frac{E_{MOBW}(T)}{n} = AM_{MOBW}, \quad \text{and} \quad \lim_{n\to\infty} \frac{E_{MOBW}(T)}{n} = AV_{MOBW}.$$

Therefore,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} E_{MOBW}(T) &= AM_{MOBW} = E_{MOBW}[\ln(f_{MOBW}(X_1, X_2; \Gamma)) - \ln(f_{BVGE}(X_1, X_2; \widetilde{\Sigma}))] \\
\lim_{n\to\infty} \frac{1}{n} V_{MOBW}(T) &= AV_{MOBW} = V_{MOBW}[\ln(f_{MOBW}(X_1, X_2; \Gamma)) - \ln(f_{BVGE}(X_1, X_2; \widetilde{\Sigma}))].
\end{aligned}$$

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

THEOREM 2: Under the assumption that data are from BVGE(Σ) the distribution of T as defined in (18), is approximately normally distributed with mean $E_{BVGE}(T)$ and variance $V_{BVGE}(T)$. The expressions of $E_{BVGE}(T)$ and $V_{BVGE}(T)$ are provided below.

PROOF OF THEOREM 2: It is provided in the Appendix. ∎

Now we provide the expressions for $E_{BVGE}(T)$ and $V_{BVGE}(T)$. In this case also we denote

$$\lim_{n\to\infty} \frac{E_{BVGE}(T)}{n} = AM_{BVGE} \quad \text{and} \quad \lim_{n\to\infty} \frac{V_{BVGE}(T)}{n} = AV_{BVGE}.$$

Therefore,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}E_{BVGE}(T) &= AM_{BVGE} = E_{BVGE}[\ln(f_{MOBW}(X_1,X_2;\widetilde{\Gamma})) - \ln(f_{BVGE}(X_1,X_2;\Sigma))]\\ \lim_{n\to\infty}\frac{1}{n}V_{BVGE}(T) &= AV_{BVGE} = V_{BVGE}[\ln(f_{MOBW}(X_1,X_2;\widetilde{\Gamma})) - \ln(f_{BVGE}(X_1,X_2;\Sigma))].\end{aligned}$$

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

Based on the asymptotic distributions, it is possible to compute the probability of correct selection (PCS) for both the cases.

5 Misspecified Parameter Estimates

Estimation of $\widetilde{\Sigma}$

In this case it is assumed that the data have been obtained from MOBW(Γ) and we would like to compute $\widetilde{\Sigma}$, the misspecified BVGE parameters, as defined in Lemma 1. Suppose $(X_1, X_2) \sim$ MOBW(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 $E_{MOBW}(\ln(f_{BVGE}(X_1,X_2;\Sigma))) = \Pi_1(\Sigma)$ (say), where

$$\begin{aligned}\Pi_1(\Sigma) &= \ln(\lambda) + p_{1W}\ln(\alpha_0+\alpha_1) + p_{1W}\ln\alpha_2 + (\alpha_0+\alpha_1-1)E_{MOBW}(\ln(1-e^{-\lambda X_1})\cdot 1_{A_1})\\ &\quad +(\alpha_2-1)E_{MOBW}(\ln(1-e^{-\lambda X_2})\cdot 1_{A_1}) - \lambda E_{MOBW}((X_1+X_2)\cdot 1_{A_1})\\ &\quad +p_{2W}\ln\alpha_1 + (\alpha_1-1)E_{MOBW}(\ln(1-e^{-\lambda X_1})\cdot 1_{A_2}) + p_{2W}\ln(\alpha_0+\alpha_2)\\ &\quad +(\alpha_0+\alpha_2-1)E_{MOBW}(\ln(1-e^{-\lambda X_2})\cdot 1_{A_2}) - \lambda E_{MOBW}((X_1+X_2)\cdot I_{A_2})\\ &\quad +(\alpha_0+\alpha_1+\alpha_2-1)E_{MOBW}(\ln(1-e^{-\lambda X})\cdot 1_{A_0}) - \lambda E_{MOBW}(X\cdot 1_{A_0}) + p_{0W}\ln\alpha_0.\end{aligned}$$

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 $\Pi_1(\Sigma)$, as follows:

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

Here

$$u_1 = \frac{\lambda_0}{\lambda_0 + \lambda_2}, \quad u_2 = \frac{\lambda_2}{\lambda_0 + \lambda_2}, \quad w_1 = \frac{\lambda_0}{\lambda_0 + \lambda_1}, \quad w_2 = \frac{\lambda_1}{\lambda_0 + \lambda_1}, \tag{19}$$

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

$$\Pi_1^*(\Sigma) = \lim_{n \to \infty} \frac{1}{n} E\left\{l_{pseudo}(\alpha_0, \alpha_1, \alpha_2, \lambda \mid (X_{1i}, X_{2i}; i = 1, \cdots, n)\right\}. \tag{20}$$

Here $l_{pseudo}(\cdot)$ 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 $(X_{1i}, X_{2i}) \sim \text{MOBW}(\alpha, \lambda_1, \lambda_2, \lambda_3)$, therefore the expressions of u_1, u_2, w_1, w_2 are as (19), 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

$$\widetilde{\alpha}_0(\lambda) \;=\; \frac{p_{0W} + u_2 p_{1W} + w_2 p_{2W}}{E\left(\ln(1 - e^{-\lambda X_1}) \cdot 1_{A_0}\right) + E\left(\ln(1 - e^{-\lambda X_1}) \cdot 1_{A_1}\right) + E\left(\ln(1 - e^{-\lambda X_2}) \cdot 1_{A_2}\right)},$$

$$\widetilde{\alpha}_1(\lambda) = \frac{u_1 p_{1W} + p_{2W}}{E\left(\ln(1-e^{-\lambda X_1})\cdot 1_{A_0}\right) + E\left(\ln(1-e^{-\lambda X_1})\cdot 1_{A_1}\right) + E\left(\ln(1-e^{-\lambda X_1})\cdot 1_{A_2}\right)}$$
$$\widetilde{\alpha}_2(\lambda) = \frac{p_{1W} + w_1 p_{2W}}{E\left(\ln(1-e^{-\lambda X_2})\cdot 1_{A_0}\right) + E\left(\ln(1-e^{-\lambda X_2})\cdot 1_{A_2}\right) + E\left(\ln(1-e^{-\lambda X_2})\cdot 1_{A_1}\right)},$$

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 a one dimensional optimization problem only.

Estimation of $\widetilde{\Gamma}$

In this case it is assumed that the data have been obtained from BVGE(Σ) and we compute $\widetilde{\Gamma}$, the misspecified MOBW parameters, as defined in Lemma 2. In this case, $\widetilde{\Gamma}$ can be obtained as the argument maximum of $E_{BVGE}(\ln(f_{MOBW}(X_1, X_2; \Gamma))) = \Pi_2(\Gamma)$ (say), where

$$\begin{aligned}
\Pi_2(\Gamma) = {} & (p_{0G} + 2p_{1G} + 2p_{2G}) \ln\alpha + p_{1G}\ln\lambda_1 + p_{2G}\ln\lambda_2 + p_{0G}\ln\lambda_0 + p_{1G}\ln(\lambda_0+\lambda_2) \\
& + p_{2G}\ln(\lambda_0+\lambda_1) + (\alpha-1)\left[E_{BVGE}(\ln X_1 \cdot 1_{A_1}) + E_{BVGE}(\ln X_1 \cdot 1_{A_2})\right] \\
& + (\alpha-1)\left[E_{BVGE}(\ln X_2 \cdot 1_{A_1}) + E_{BVWE}(\ln X_2 \cdot 1_{A_2}) + E_{BVWE}(\ln X_1 \cdot 1_{A_0})\right] \\
& - \lambda_1\left[E_{BVWE}(X_1^\alpha \cdot 1_{A_1}) + E_{BVWE}(X_1^\alpha \cdot 1_{A_2}) + E_{BVWE}(X_1^\alpha \cdot 1_{A_0})\right] \\
& - \lambda_2\left[E_{BVWE}(X_2^\alpha \cdot 1_{A_1}) + E_{BVWE}(X_2^\alpha \cdot 1_{A_2}) + E_{BVWE}(X_1^\alpha \cdot 1_{A_0})\right] \\
& - \lambda_0\left[E_{BVWE}(X_1^\alpha \cdot 1_{A_2}) + E_{BVWE}(X_2^\alpha \cdot 1_{A_1}) + E_{BVWE}(X_1^\alpha \cdot 1_{A_0})\right]
\end{aligned}$$

In this case also 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

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

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

It may be similarly observed as before that

$$\Pi_2^*(\Gamma) = \lim_{n\to\infty} \frac{1}{n} E\left(l_{pseudo}(\alpha, \lambda_0, \lambda_1, \lambda_2 \mid (X_{1i}, X_{2i}); i = 1, \cdots, n)\right), \tag{21}$$

where $(X_{1i}, X_{2i}) \sim \text{BVGE}(\alpha_0, \alpha_1, \alpha_2, \lambda)$. The explicit expression of $l_{pseudo}(\cdot)$ 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

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

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). They can be obtained from the authors on request. We compute PCS based on Monte Carlo simulation(MC), 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.

Case 1: Parent Distribution is MOBW

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 MOBW distribution. Then we compute the MLEs of the unknown parameters and the associated log-likelihood values, assuming that the data are coming from MOBW or BVGE distribution. In computing the MLEs 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 maximized log-likelihood values 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 the Tables 1 to 4.

Table 1: Probability of correct selection based on Monte Carlo (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

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 following

Table 2: Probability of correct selection based on Monte Carlo (MC) simulations and based on asymptotic distribution (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: Probability of correct selection based on Monte Carlo (MC) simulations and based on asymptotic distribution (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 5. In each case we need to compute AM_{MOBW} and AV_{MOBW}, as defined in Theorem 1. Since the exact expressions of AM_{MOBW} and AV_{MOBW} are quite complicated, we have used simulation consistent estimates of AM_{MOBW} and AV_{MOBW}, which can be obtained very easily. The simulation consistent estimators of AM_{MOBW} and AV_{MOBW} are obtained using 10,000 replications, and they are reported in Table 6

Now using Theorem 1, based on the asymptotic distribution of T, the discrimination statistic, we compute the probability of correct selection, *i.e.* $P(T > 0)$ 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 probability of correct selections are very high, and the asymptotic results match very well with the simulated results.

Case 2: Parent Distribution is BVGE

In this case we consider the following parameter sets;

Set 5: $\alpha_0 = 1.5$, $\alpha_1 = 2.0$, $\alpha_2 = 1.0$, $\lambda = 1.0$; Set 6: $\alpha_0 = 1.0$, $\alpha_1 = 1.0$, $\alpha_2 = 1.0$, $\lambda = 1.0$;

Table 4: Probability of correct selection based on Monte Carlo (MC) simulations and based on asymptotic distribution (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

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

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 BVGE, and using the same procedure as before we compute the proportion of correct selection. The results are reported in the first rows of the Tables 7 to 10.

Now to compute the asymptotic probability of correct selection, 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_{BVGE} and AV_{BVGE} in Table 12.

Now similarly as before, based on the asymptotic distribution of T, as provided in Theorem 2, we compute the probability of correct selection in this case, *i.e.* $P(T < 0)$ 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 also, it observed that the asymptotic results match extremely well with the simulated results.

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

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

Table 7: Probability of correct selection based on Monte Carlo (MC) simulations and based on asymptotic distribution (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

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.

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)

Table 8: Probability of correct selection based on Monte Carlo (MC) simulations and based on asymptotic distribution (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: Probability of correct selection based on Monte Carlo (MC) simulations and based on asymptotic distribution (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

$X_1 > X_2$ 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 following random variables:

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

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 Marshall-Olkin bivariate exponential model. Csorgo and Welsh (1989) analyzed the data using the Marshall-Olkin bivariate 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 hazard functions of both X_1 and X_2 are increasing functions.

Since both MOBW and BVGE can have increasing marginal hazard functions, we fit

Table 10: Probability of correct selection based on Monte Carlo (MC) simulations and based on asymptotic distribution (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

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

both the models to the data set. For MOBW distribution using EM algorithm as suggested in Kundu and Dey (2009), we compute the MLEs of the unknown parameters as $\widehat{\alpha} = 1.2889$, $\widehat{\lambda}_0 = 11.2073$, $\widehat{\lambda}_1 = 8.3572$, $\widehat{\lambda}_2 = 0.4720$, 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 corresponding log-likelihood value is 47.8041. In case of BVGE distribution using the EM algorithm as suggested in Kundu and Gupta (2009), we obtained the MLEs 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 corresponding log-likelihood value is 38.0042. Therefore, based on the log-likelihood values we prefer to use the MOBW model rather than BVGE model to analyze this data set.

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

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

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

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

Table 13: American Football League (NFL) data

probability of correct selection is 0.98.

8 Conclusion

In this paper we have considered discrimination between two singular bivariate distributions namely, MOBW and BVGE distributions. Both the distributions have singular part and absolute continuous part. The difference of the maximized log-likelihood values has been used as the discrimination statistic. We have obtained the asymptotic distribution of the

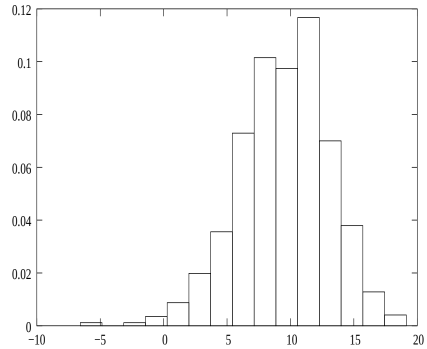

Figure 3: Histogram of the bootstrap sample of the discrimination statistic.

discrimination statistic, which can be used to compute the asymptotic probability of correct selection. Monte Carlo 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 MOBW and BVGE is relatively much easier. Even with small sample sizes the probability of correct selection is quite high. Moreover the asymptotic probability of correct selection matches very well with the simulated probability of correct selection even for moderate sample sizes. We have performed the analysis of a data set, and computed the probability of correct selection 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 probability of correct selection in this case. More work is needed in this direction.

Appendix :

To prove Theorem 1, we need the following Lemma 1. Here $\stackrel{a.s.}{\to}$ means converges almost surely.

Lemma 1: Under the assumption that data are from $BVWE(\alpha, \lambda_0, \lambda_1, \lambda_2)$, as $n \to \infty$, we have

(i) $\widehat{\alpha} \xrightarrow{a.s.} \alpha$, $\widehat{\lambda}_0 \xrightarrow{a.s} \lambda_0$, $\widehat{\lambda}_1 \xrightarrow{a.s} \lambda_1$ and $\widehat{\lambda}_2 \xrightarrow{a.s} \lambda_2$ where for $\Gamma = (\alpha, \lambda_0, \lambda_1, \lambda_2)$

$$E_{MOBW}(\ln(f_{MOBW}(X_1, X_2; \Gamma))) = \max_{\bar{\Gamma}} E_{MOBW}(\ln(f_{MOBW}(X_1, X_2; \bar{\Gamma}))) \tag{22}$$

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

$$E_{MOBW}(\ln(f_{BVGE}(X_1, X_2; \widetilde{\Sigma}))) = \max_{\Sigma} E_{MOBW}(\ln(f_{BVGE}(X_1, X_2; \Sigma))) \tag{23}$$

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 $n^{-\frac{1}{2}}\,[T - E_{MOBW}(T)]$ is asymptotically equivalent to $n^{-\frac{1}{2}}\,[T^* - E_{MOBW}(T^*)]$

Proof of Lemma 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 1: Using Central limit theorem and part (ii) of Lemma 1, it follows that $n^{-\frac{1}{2}}\,[T^* - E_{BVWE}(T^*)]$ is asymptotically normally distributed with mean zero and variance $V_{MOBW}(T^*)$. Therefore, using part (iii) of Lemma 1, the result immediately follows. ∎

To prove Theorem 2, and for defining the misspecified parameter $\widetilde{\Gamma}$ we need the following Lemma 2, whose proof is same as the proof of Lemma 1. ∎

Lemma 2: Suppose the data follow $BVGE(\alpha_0, \alpha_1, \alpha_2, \lambda)$, as $n \to \infty$, we have

(i) $\widehat{\alpha}_0 \xrightarrow{a.s.} \alpha_0$, $\widehat{\alpha}_1 \xrightarrow{a.s.} \alpha_1$, $\hat{\alpha}_2 \xrightarrow{a.s.} \alpha_2$ and $\hat{\lambda} \to \lambda$ where

$$E_{BVGE}(\ln(f_{BVGE}(X_1, X_2; \Sigma))) = \max_{\bar{\Sigma}} E_{BVGE}(\ln(f_{BVGE}(X_1, X_2; \bar{\Sigma}))) \tag{24}$$

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

$$E_{BVGE}(\ln(f_{MOBW}(X_1, X_2; \widetilde{\Gamma}))) = \max_{\Gamma} E_{BVGE}(\ln(f_{MOBW}(X_1, X_2; \Gamma))) \tag{25}$$

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 $n^{-\frac{1}{2}}[T - E_{BVGE}(T)]$ is asymptotically equivalent to $n^{-\frac{1}{2}}[T_* - E_{BVGE}(T^*)]$

Proof of Theorem 2: Along the same line as the Proof of Lemma 1, it also follows using Lemma 2. ∎

The following Lemmas will be 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 $W_0 \sim$ GE$(\alpha_0 + \alpha_1 + \alpha_2, \lambda)$, $W_1 \sim$ GE$(\alpha_0 + \alpha_1, \lambda)$, $W_2 \sim$ GE$(\alpha_0 + \alpha_2, \lambda)$ and $(X_1, X_2) \sim$ BVGE$(\alpha_0, \alpha_1, \alpha_2, \lambda)$. If $g(\cdot)$ is any Borel measurable function, then

$$\begin{aligned} E(g(X_1) \cdot 1_{A_1}) &= E(g(W_1)) + \frac{\alpha_0 + \alpha_1}{\alpha_0 + \alpha_1 + \alpha_2} E(g(W_0)). \\ E(g(X_1) \cdot 1_{A_2}) &= \frac{\alpha_1}{\alpha_0 + \alpha_1 + \alpha_2} E(g(W_0)). \\ E(g(X_1) \cdot 1_{A_0}) &= E(g(X_2) \cdot 1_{A_0}) = \frac{\alpha_0}{\alpha_0 + \alpha_1 + \alpha_2} E(g(W_0)). \\ E(g(X_2) \cdot 1_{A_1}) &= \frac{\alpha_2}{\alpha_0 + \alpha_1 + \alpha_2} E(g(W_0)). \\ E(g(X_2) \cdot 1_{A_2}) &= E(g(W_2)) + \frac{\alpha_0 + \alpha_2}{\alpha_0 + \alpha_1 + \alpha_2} E(g(W_0)). \end{aligned}$$

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

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

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

$$
\begin{aligned}
E(g(X_1)\cdot 1_{A_2}) &= E(g(Z_1)) - \frac{\lambda_0+\lambda_1}{\lambda_0+\lambda_1+\lambda_2}E(g(Z_0)). \\
E(g(X_1)\cdot 1_{A_0}) &= E(g(X_2)\cdot 1_{A_0}) = \frac{\lambda_0}{\lambda_0+\lambda_1+\lambda_2}E(g(Z_0)). \\
E(g(X_2)\cdot 1_{A_1}) &= E(g(Z_2)) - \frac{\lambda_0+\lambda_2}{\lambda_0+\lambda_1+\lambda_2}. \\
E(g(X_2)\cdot 1_{A_2}) &= \frac{\lambda_2}{\lambda_0+\lambda_1+\lambda_2}E(g(Z_2)).
\end{aligned}
$$

PROOF OF LEMMA A.1: They can be obtained along the same line as in Lemma A.1. ∎

References

[1] Atkinson, A. (1969), "A test for discriminating between models", *Biometrika*, vol. 56, 337 - 341.

[2] Atkinson, A. (1970), "A method for discriminating between models (with discussions)", *Journal of the Royal Statistical Society*, Ser. B. vol. 32, 323 - 353.

[3] Bain, L.J. and Englehardt, M. (1980), "Probability of correct selection of Weibull versus gamma based on likelihood ratio test", *Communications in Statistics - Theory and Methods*, vol. 9, 375 - 381.

[4] Bemis, B., Bain, L.J. and Higgins, J.J. (1972), "Estimation and hypothesis testing for the parameters of a bivariate exponential distribution", *Journal of the American Statistical Association*, vol. 67, 927-929.

[5] Cox, D. R. (1961), "Tests of separate families of hypotheses", *Proceedings of the Fourth Berkley Symposium on Mathematical Statistics and Probability*, Berkley, University of California Press, 105 - 123.

[6] Cox, D.R. (1962), "Further results on tests of separate families of hypotheses", *Journal of the Royal Statistical Society*, Ser B, vol. 24, 406 - 424.

[7] Csorgo, S. and Welsh, A.H. (1989), "Testing for exponential and Marshall-Olkin distribution", *Journal of Statistical Planning and Inference*, vol. 23, 287 - 300.

[8] Dey, A.K. and Kundu, D. (2009), "Discriminating among the log-normal, Weibull and generalized exponential distributions", *IEEE Transactions on Reliability*, vol. 58, 416-424.

[9] Dey, A. K. and Kundu, D. (2010), "Discriminating between the log-normal and log-logistic distributions", *Communications in Statistics - Theory and Methods*, vol. 39, 280 - 292.

[10] Gupta, R.D. and Kundu, D. (2003), "Discriminating between Weibull and generalized exponential distributions", *Computational Statistics and Data Analysis*, vol. 43, 179 - 196.

[11] Gupta, R.D. and Kundu, D. (2006), "On the comparison of Fisher information of the Weibull and GE distributions", *Journal of Statistical Planning and Inference*, vol. 136, 3130 - 3144.

[12] Gupta, R.D. and Kundu, D. (2007), "Generalized exponential distribution: existing methods and recent developments", *Journal of the Statistical Planning and Inference*, vol. 137, 3537 - 3547.

[13] Kotz, S., Balakrishnan, N. and Johnson, N. (2000), *Continuous Multivariate Distributions: Models And Applications*, Wiley and Sons, New York.

[14] Kundu D., Dey, A. K. (2009), "Estimating the parameters of the Marshall-Olkin bivariate Weibull distribution by EM algorithm", *Computational Statistics and Data Analysis*, vol. 53, 956-965.

[15] Kundu, D. and Gupta, R.D. (2009), "Bivariate generalized exponential distribution", *Journal of Multivariate Analysis*, vol 100, 581 - 593.

[16] Marshall, A.W., Meza, J.C. and Olkin, I. (2001), "Can data recognize its parent distribution?", *Journal of Computational and Graphical Statistics*, vol. 10, 555 - 580.

[17] White, H. (1982), "Nonlinear Regression on Cross-section Data", *Econometrica*, vol. 48, 721-746.