Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110105

Generalization of the Ratio of Minimized Kullback-Leibler Divergence Discrimination Technique to Bivariate Marshall-Olkin Family

Hanan. M. Aly and Ola. A. Abuelamayem∗

Department of Statistics, Faculty of Economics and Political Science, Cairo University, Egypt

Received: 20 Mar. 2020, Revised: 20 Nov. 2020, Accepted: 13 Feb. 2021
Published online: 1 Jan. 2022

Abstract: Different bivariate lifetime distributions are used to analyze lifetime data and study the reliability of the products. Sometimes, we find more than one distribution that fit the data well. In this case, we should select the best one. That is why the discriminant analysis is used. In literature, there is only one method used for bivariate lifetime distributions, which is the likelihood ratio test. In this paper, we try to generalize the ratio of minimized Kullback-Leibler divergence to be used as a discrimination method in the bivariate case and it could be applied on the bivariate Marsall-Olkin family. we will select two lifetime distributions which belong to the bivariate Marshall-Olkin family. The distributions are, bivariate generalized exponential distribution and a recently proposed distribution which is the bivariate inverted Kumaraswamy distribution. We compared the proposed method with likelihood ratio test after deriving its asymptotic distribution. The minimum sample size required for discrimination is obtained using the derived asymptotic distribution. A simulation study is performed to illustrate the results and it is found that ratio of minimized Kullback-Leibler divergence method performs better than likelihood ratio test method. Finally, A real dataset is analyzed.

Keywords: Bivariate inverted Kumaraswamy distribution,Bivariate Marshall-Olkin family, Discriminant analysis,Likelihood ratio test, The ratio of minimized Kullback-Leibler divergence.

1 Introduction

Higher manufacturing technology in companion with global competition increase the need to study the lifetime of products. The better we analyze the performance of products, the more we can guarantee the quality and consumer satisfaction. In literature, There are several bivariate families that can be used to analyze lifetime data. one of the most heavily used families in reliability engineering is the bivariate Marshall-Olkin family. This family has the advantage of analyzing different scenarios (the lifetime of the first component is less than, greater than or equal to the second component). Also it can be applied in different models, for example;

Maintenance Model: consider a two-component system and suppose that there is separate maintenance for each component, as well as, an overall maintenance. The lifetime of the first component is increased by T_1 due to the separate maintenance; similarly the lifetime of the second component is increased by T_2. Also, the lifetime of each component is increased by T_3 due to the overall maintenance. Accordingly, the increases in lifetime are X_1=max(T_1,T_3) and X_2=max(T_2,T_3). The increase in the system's lifetime can be modeled by the bivariate distribution of X_1 and X_2 (for more details, see [1]).

Stress Model: consider a two-component system with three different sources of stresses, a stress U_1 corresponds to the first component, a stress U_2 corresponds to the second component, and a stress U_3 affects both components equally. The observed stress is $X_1 = \max(U_1, U_3)$ and $X_2 = \max(U_2, U_3)$ for the first and second components, respectively. The observed stress can be modeled by the bivariate distribution of X_1 and X_2 (for more details, see [1]).

∗ Corresponding author e-mail: ola.abuelamayem@feps.edu.eg

There are several bivariate lifetime distributions which belong to the bivariate Marshall-Olkin family. Sometimes, we find more than one distribution that fit the data well. In this case, we should select the best one. That is why discriminant analysis is used.

Discrimination between two univariate distributions was first discussed by [2,3]. [4] used a combined distribution and applied it to compare between exponential and log-normal distributions. The most heavily used method in discrimination between two univariate distributions is the likelihood ratio test (LRT). Several papers applied it to make discrimination, for example, [[5]-[6]]. Other discrimination procedures were also applied, for example, [7] compared between generalized extreme value, Pearson Type III and generalized logistic using normality transformation followed by applying the Shapiro-Wilk and the probability plot correlation coefficient statistics.

[8] presented the ratio of minimized Kullback-Leibler divergence (RMKLD) as a new measure of discrimination. They applied this measure to discriminate between Gamma and log normal distribution. [9] discriminated between Gamma and log logistic distributions using both LRT and RMKLD. This method gives a better discrimination in terms of higher probability of correct selection.

Although, extensive work was performed in the univariate case. To the best of our knowledge, only two papers handled the discrimination problem in the bivariate case and using LRT method. [10] selected between bivariate generalized exponential and bivariate Weibull distributions. [11] decided between bivariate exponential and exponentiated exponential distributions.

In this paper, we generalize the RMKLD to the bivariate case and it could be applied on the bivariate Marshall-Olkin family. Also, the presented method is used to discriminate between the recently derived bivariate inverted Kumaraswamy (BIK) and bivariate generalized exponential (BVGE) distributions. Moreover, we compared the proposed method with LRT after deriving its asymptotic distribution.

The paper is organized as follows; In Section 2, The Kullback-Leibler divergence test statistic is presented. Also, the generalization of the RMKLD test statistic to the bivariate Marshall-Olkin family is performed. In Section 3, The LRT is explained, Besides the asymptotic distribution of the likelihood ratio test statistic is derived. Also, the misspecified parameters estimation for both BIK and BVGE distributions is performed. In Section 4, the maximum likelihood estimation technique is used to estimate the unknown parameters for both BIK and BVGE distributions. In Section 5, asymptotic distribution is used to determine the minimum sample size needed to discriminate between two distributions. In Section 6, simulation studies are performed using 10000 replications. Also, to select the best method, a comparison between RMKLD, LRT (using simulation and asymptotic distribution) is made. Moreover a real dataset is analyzed. Finally, the paper is concluded in Section 7.

2 Kullback-Leibler Divergence Test Statistic

The Kullback-Leibler divergence (KLD) between two models say, f and g (KLD($f \parallel g$)) is a method to measure the information lost when g is used to approximate f, or in other words it measures the distance from g to f. For continuous distributions, KLD is defined to be as follows (for more details, see [12]);

$$KLD(f\|g) = \int f(x)log(\frac{f(x)}{g(x)})dx.$$

KLD is a non-symmetric measure of the difference between the two density functions f and g. It is a directed distance as the distance from f to g is not the same as that from g to f. The following are some properties of KLD ([12]):

1- The distributions in their entirety are the subject of comparison and not only the mean and variance.

2- KLD is always positive (i.e. KLD ($f\|$ g) > 0, f and g are not identical).

3- KLD($f\|$ g) = 0 iff f=g.

4- The smaller KLD($f\|$ g) means that f is preferable, while large values is in favor of g.

[13] used KLD to choose between Weibull and log-normal distribution. However, the applicability of his approach needs to generate critical values and then judge to reject or accept the null hypothesis (i.e. for example H_0: Weibull , H_1: log-normal). Hence, [8] presented the RMKLD as a new test statistic, where there is no need to generate critical values

and hence much easier to be applied. Suppose $h_1(x), h_2(x)$ are continuous probability density functions, the test statistic is as follows;

$$RMKLD(h_1(x), h_2(x)) = ln(\frac{KLD(h_1(x)\|h_2(x))}{KLD(h_2(x)\|h_1(x))}).$$

$h_1(x)$ is selected if RMKLD < 0. This measure was presented by [8] for univariate distributions. Here, we generalize this method to be applied in the bivariate Marshall-Olkin family. First, we will explain the structure of the bivariate Marshall-Olkin family and then present the general form of RMKLD.

2.1 Bivariate Marshall-Olkin Family

[14] presented a bivariate exponential distribution with exponential marginals and loss of memory property. This type of bivariate distributions have the following form.

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} f_1(x_1,x_2) \; if \quad x_1 < x_2 \\ f_2(x_1,x_2) \; if \quad x_2 < x_1 \\ \quad f_3(x) \quad if \; x_1 = x_2 = x \end{cases}, \qquad (1)$$

which has both an absolutely continuous part (i.e $f_1(x_1,x_2)$ and $f_2(x_1,x_2)$) and singular part ($f_3(x)$).

where;

$$\begin{aligned} f_1(x_1,x_2) &= f_E(x_1;\lambda_1)f_E(x_2;\lambda_2+\lambda_3), \\ f_2(x_1,x_2) &= f_E(x_1;\lambda_1+\lambda_3)f_g(x_2;\lambda_2), \\ f_3(x) &= \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3} f_E(x;\lambda_1+\lambda_2+\lambda_3), \\ f_E(x;\lambda) &= \lambda e^{-\lambda x}, \end{aligned}$$

and $\lambda_1, \lambda_2, \lambda_3, \lambda > 0$.

Due to wide applicability and importance of this family in reliability engineering (illustrated in Section 1), several distributions that belong to the bivariate Marshall-Olkin family were constructed. For example; [1] introduced bivariate generalized exponential distribution, [15] presented bivariate models with proportional reversed hazard marginals, [16] illustrated bivariate Kumaraswamy distribution, [17] introduced bivariate inverse Weibull distribution. Recently, the authors presented BIK distribution.

BIK distribution and BVGE distribution have similar shapes under different combinations of the parameters (as illustrated in Figure 1 and Figure 3). Hence, we will focus here on discriminating between them. The following is the probability density function of BIK;

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} f_1(x_1,x_2) = f_{IK}(x_1,\beta_1+\beta_3,\alpha)f_{IK}(x_2,\beta_2,\alpha) \\ f_2(x_1,x_2) = f_{IK}(x_1,\beta_1,\alpha)f_{IK}(x_2,\beta_2+\beta_3,\alpha) \\ \quad f_3(x) \quad = \frac{\beta_3}{\beta_1+\beta_2+\beta_3} f_{IK}(x,\beta_1+\beta_2+\beta_3,\alpha) \end{cases}, \qquad (2)$$

where;
$f_{IK}(x,\beta,\alpha) = \alpha\beta(1+x)^{-(\alpha+1)}(1-(1+x)^{-\alpha})^{\beta-1}, x > 0, \alpha, \beta > 0.$

The joint probability density function can take different shapes according to the parameters' values. Here, we explain only some of them (see Figure1),

a) ($\alpha = 1$, $\beta_1 = 0.5$, $\beta_2 = 0.5$, $\beta_3 = 0.5$) the surface plot of the absolutely continuous part of the joint probability density function is concave up and decreasing.

b)($\alpha = 0.9$, $\beta_1 = 3$, $\beta_2 = 1.5$, $\beta_3 = 0.5$) the surface plot of the absolutely continuous part of the joint probability density function is concave down and increasing till it reaches the mode, then it is decreasing.

Because different combinations of parameters results in different shapes for the joint probability density function, the BIK distribution can be effectively used in analyzing bivariate data.

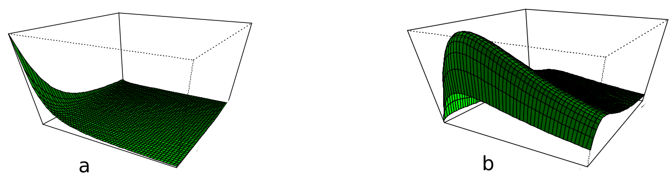

Fig. 1: The absolutely continuous part of BIK pdf with parameters (β_1,β_2,β_3,α): (a) (0.5,0.5,0.5,1), (b) (3,1.5,0.5,0.9).

BIK distribution may has several applications in different fields. For example, as a lifetime distribution it can be used in reliability and life testing problems. Also as a member of bivariate Marshall-Olkin family it has applications in Maintenance models, shock models (described in Section 1) and failure rate models in medical research and biological studies such as frailty model (see for Example [18]). Moreover, it can be seen from Figure 2 that the hazard rate of BIK distribution is increasing with time (i.e. wear out). Hence, it can be used in engineering sciences: for example, degradation of mechanical components such as pistons, crankshafts of diesel engines.

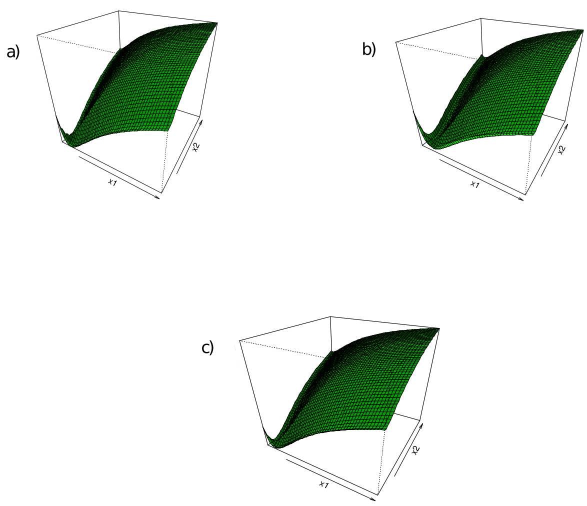

Fig. 2: The absolutely continuous part of BIK hazard rate with parameters (β_1,β_2,β_3,α): (a) (0.6,0.6,0.6,0.6), (b) (0.9,0.9,0.4,0.2), (c) (0.7, 0.7, 0.4, 0.4).

The BVGE density function is as follows;

$$g_{X_1,X_2}(x_1,x_2) = \begin{cases} g_1(x_1,x_2) &= g_{GE}(x_1,\alpha_1+\alpha_3,\lambda)g_{GE}(x_2,\alpha_2,\lambda) \\ g_2(x_1,x_2) &= g_{GE}(x_1,\alpha_1,\lambda)g_{GE}(x_2,\alpha_2+\alpha_3,\lambda) \\ g_3(x) &= \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3} g_{GE}(x,\alpha_1+\alpha_2+\alpha_3) \end{cases}, \quad (3)$$

where;

$g_{GE}(x) = \alpha\lambda e^{-\lambda x}(1-e^{-\lambda x})^{(\alpha-1)}, x > 0, \alpha, \lambda > 0.$

Analyzing some of the different shapes presented [10], we can see from Fig.3 that

a) ($\lambda = 1$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 2$) the surface plot of the absolutely continuous part of the joint probability density function is concave up and decreasing.

b)($\lambda = 1$, $\alpha_1 = 0.5$, $\alpha_2 = 0.5$, $\alpha_3 = 0.5$) the surface plot of the absolutely continuous part of the joint probability density function is concave down and increasing till it reaches the mode, then it is decreasing.

It can be seen that, under some combinations of the parameters, the joint probability density function is similar to that of the BIK distribution.

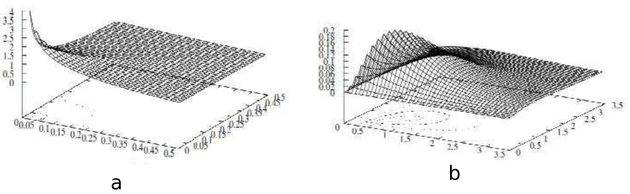

Fig. 3: The absolutely continuous part of BVGE pdf with parameters (α_1,α_2,α_3,λ):(a) (1,1,2,1), (b) (0.5,0.5,0.5,1). Source: [10].

2.2 Ratio of Minimized Kullback-Leibler Divergence for Bivariate Marshall-Olkin Family

Suppose f and g are two continuous probability density functions, the two functions are members of the bivariate Marshall-Olkin family. Hence, both functions can be written as illustrated in Equation (1) in the previous subsection. RMKLD for bivariate Marshall-Olkin family can be written as follows;

$$RMKLD(f(x_1,x_2),g(x_1,x_2)) = ln(\frac{KLD(f(x_1,x_2)\|g(x_1,x_2))}{KLD(g(x_1,x_2)\|f(x_1,x_2))}),$$

where

$$\begin{aligned} KLD(f(x_1,x_2)\|g(x_1,x_2)) &= \int_0^\infty \int_0^{x_2} f_1(x_1,x_2)\log(\frac{f_1(x_1,x_2)}{g_1(x_1,x_2)})dx_1dx_2 \\ &+ \int_0^\infty \int_0^{x_1} f_2(x_1,x_2)\log(\frac{f_2(x_1,x_2)}{g_2(x_1,x_2)})dx_2dx_1 \\ &+ \int_0^\infty f_3(x)\log(\frac{f_3(x)}{g_3(x)})dx. \end{aligned}$$

which can be rewritten as follows;

$$\begin{aligned} KLD(f(x_1,x_2)\|g(x_1,x_2)) &= \int_0^\infty \int_0^{x_2} f_1(x_1,x_2)\log(f_1(x_1,x_2))dx_1dx_2 \\ &- \int_0^\infty \int_0^{x_2} f_1(x_1,x_2)\log(g_1(x_1,x_2))dx_1dx_2 \\ &+ \int_0^\infty \int_0^{x_1} f_2(x_1,x_2)\log(f_2(x_1,x_2))dx_2dx_1 \\ &- \int_0^\infty \int_0^{x_1} f_2(x_1,x_2)\log(g_2(x_1,x_2))dx_2dx_1 \\ &+ \int_0^\infty f_3(x)\log(f_3(x))dx \\ &- \int_0^\infty f_3(x)\log(g_3(x))dx. \end{aligned}$$

To illustrate the procedure, the following steps can be used to apply RMKLD to choose between two distributions;

1- Calculate $KLD(f(x_1,x_2)\|g(x_1,x_2))$ and $KLD(g(x_1,x_2)\|f(x_1,x_2))$.

2- Calculate $RMKLD(f(x_1,x_2),g(x_1,x_2)) = ln(\frac{KLD(f(x_1,x_2)\|g(x_1,x_2))}{KLD(g(x_1,x_2)\|f(x_1,x_2))})$.

3- Select $f(x_1,x_2)$ if RMKLD < 0.

In the next section, we will explain the likelihood ratio test. Also, we will derive its asymptotic distribution.

3 Likelihood Ratio Test

The likelihood ratio test was first applied for the bivariate Marshall-Olkin family by [10], the test has the following form;

$T = L_{BIK}(x_1,x_2,\hat{\xi}) - L_{BVGE}(x_1,x_2,\hat{\Psi})$,

where

$\hat{\xi} = (\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\hat{\alpha})$: is the maximum likelihood estimators obtained by maximizing the log likelihood function of BIK distribution $L_{BIK}(x_1,x_2,\xi)$.

$\hat{\psi} = (\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3,\hat{\lambda})$: is the maximum likelihood estimators obtained by maximizing the log likelihood function of BVGE distribution $L_{BVGE}(x_1,x_2,\psi)$.

$L_{BIK}(x_1,x_2,\hat{\xi})$: the maximized log likelihood function of the BIK distribution.

$L_{BVGE}(x_1,x_2,\hat{\psi})$: the maximized log likelihood function of the BVGE distribution.

The discrimination criteria is as follows;

- If $T > 0$, select BIK distribution.

- If $T < 0$, select BVGE distribution.

In the next subsection, the asymptotic distribution for the test statistic (T) will be derived.

3.1 Asymptotic distribution

Assume that U follows $BIK(\xi)$, $f_1(u)$ and $f_2(u)$ are two probability density functions of U, $E_{BIK}(f_1(u))$, $V_{BIK}(f_1(u))$ and $Cov_{BIK}(f_1(u),f_2(u))$ are the mean of $f_1(u)$, the variance of $f_1(u)$, and the covariance of $f_1(u)$ and $f_2(u)$, respectively. Similarly for $BVGE(\Psi)$ distribution. To derive the Asymptotic distribution we will introduce the following lemma;

Lemma 1. Assume that the data come from $BIK(\beta_1,\beta_2,\beta_3,\alpha)$ and under the assumptions illustrated in [19], let $n \to \infty$, then

a) $\hat{\beta}_1 \xrightarrow{a.s} \beta_1$, $\hat{\beta}_2 \xrightarrow{a.s} \beta_2$, $\hat{\beta}_3 \xrightarrow{a.s} \beta_3$, and $\hat{\alpha} \xrightarrow{a.s} \alpha$. For $\xi = (\beta_1,\beta_2,\beta_3,\alpha)$, we have

$E_{BIK}[\log(f_{BIK}(x_1,x_2;\xi))] = \max_{\bar{\xi}} E_{BIK}[\log(f_{BIK}(x_1,x_2;\bar{\xi}))]$

b) $\hat{\alpha}_1 \xrightarrow{a.s} \tilde{\alpha}_1$, $\hat{\alpha}_2 \xrightarrow{a.s} \tilde{\alpha}_2$, $\hat{\alpha}_3 \xrightarrow{a.s} \tilde{\alpha}_3$ and $\hat{\lambda}_1 \xrightarrow{a.s} \tilde{\lambda}_1$ where for $\psi = (\alpha_1,\alpha_2,\alpha_3,\lambda)$, $E_{BIK}[\log(f_{BVGE}(X_1,X_2,\tilde{\psi}))] = max_{\psi} E_{BIK}[\log(f_{BVGE}(X_1,X_2,\psi))]$.

c) Define $T^* = L_{BIK}(x_1,x_2,\xi) - L_{BVGE}(x_1,x_2,\tilde{\psi})$, then $n^{-\frac{1}{2}}(T - E_{BIK}(T))$ is asymptotically equivalent to $n^{-\frac{1}{2}}(T^* - E_{BIK}[T^*])$.

Proof: The proof follows using similar arguments as ([19], Theorem 1).

Theorem 1. Suppose the data come from BIK(ξ) distribution, then the test statistic (T), defined in Section 3, is approximately normally distributed with mean $E_{BIK}(T)$ and variance $V_{BIK}(T)$.

Proof: Using the central limit theorem and part (b) of Lemma 1, then $n^{-\frac{1}{2}}(T^* - E_{BIK}[T^*])$ is asymptotically normally distributed with mean zero and variance $V_{BIK}(T^*)$. Now, using part (c) in Lemma 1 the result follows.

Lemma 2. Under the assumption that the data come from BVGE$(\alpha_1, \alpha_2, \alpha_3, \lambda)$ distribution and under the assumptions illustrated in [19], let n $\to \infty$, then

a) $\hat{\alpha}_1 \xrightarrow{a.s} \alpha_1$, $\hat{\alpha}_2 \xrightarrow{a.s} \alpha_2$, $\hat{\alpha}_3 \xrightarrow{a.s} \alpha_3$ and $\hat{\lambda} \xrightarrow{a.s} \lambda$, where for $\psi = (\alpha_1, \alpha_2, \alpha_3, \lambda)$,
$E_{BVGE}[log(f_{BVGE}\ (X_1, X_2; \psi))] = max_{\bar{\psi}} E_{BVGE}[log(f_{BVGE}(X_1, X_2; \bar{\psi}))]$.

b) $\hat{\beta}_1 \xrightarrow{a.s} \tilde{\beta}_1$, $\hat{\beta}_2 \xrightarrow{a.s} \tilde{\beta}_2$, $\hat{\beta}_3 \xrightarrow{a.s} \tilde{\beta}_3$ and $\hat{\alpha} \xrightarrow{a.s} \tilde{\alpha}$, where for $\xi = (\beta_1, \beta_2, \beta_3, \alpha)$,
$E_{BVGE}[log(f_{BIK}(X_1\ , X_2; \tilde{\xi}))] = max_{\xi} E_{BVGE}[log(f_{BIK}(X_1, X_2; \xi))]$.

c) Define $T^* = L_{BIK}(x_1, x_2, \tilde{\xi}) - L_{BVGE}(x_1, x_2, \psi)$, then $n^{-\frac{1}{2}}(T - E_{BVGE}[T])$ is asymptotically equivalent to $n^{-\frac{1}{2}}(T^* - E_{BVGE}[T^*])$.

Proof: The proof for Lemma 2 is the same as Lemma 1.

Theorem 2. Suppose the data come from BVGE(ψ) distribution, then the test statistic (T), defined in Section 3, is approximately normally distributed with mean $E_{BVGE}(T)$ and variance $V_{BVGE}(T)$.

Proof: The proof for Theorem 2 is the same as Theorem 1, respectively.

Now, we denote:

$\lim_{n\to\infty} \frac{E_{BIK}[T]}{n} = AM_{BIK}(T)$, and $\lim_{n\to\infty} \frac{V_{BIK}(T)}{n} = AV_{BIK}(T)$, hence

$AM_{BIK}(T) = E_{BIK}[log(f_{BIK}(x_1, x_2, \xi)) - log(f_{BVGE}(x_1, x_2, \tilde{\psi}))]$,

$AV_{BIK}(T) = V_{BIK}[log(f_{BIK}(x_1, x_2, \xi)) - log(f_{BVGE}(x_1, x_2, \tilde{\psi}))]$.

$\lim_{n\to\infty} \frac{E_{BVGE}[T]}{n} = AM_{BVGE}(T)$, and $\lim_{n\to\infty} \frac{V_{BVGE}(T)}{n} = AV_{BVGE}(T)$, hence

$AM_{BVGE}(T) = E_{BVGE}[log(f_{BIK}(x_1, x_2, \tilde{\xi})) - log(f_{BVGE}(x_1, x_2, \psi))]$,

$AV_{BVGE}(T) = V_{BVGE}[log(f_{BIK}(x_1, x_2, \tilde{\xi})) - log(f_{BVGE}(x_1, x_2, \psi))]$.

Here, $AM_{BIK}[T]$, $AV_{BIK}[T]$, $AM_{BVGE}[T]$ and $AV_{BVGE}[T]$ could not be obtained in closed form and therefore need to be computed numerically.

The computation of the misspecified parameters $\tilde{\psi}$ and $\tilde{\xi}$ will be illustrated more in the next subsection.

3.2 Misspecified Parameters

Let f(x;θ) and g(x;α) be two probability density functions that fit our data well. Now, suppose that f(x;θ) is the true model and g(x;α) is incorrect. The parameters corresponding to the incorrect model (i.e. α) is defined as misspecified parameters. For more details, see [20] and [21].

In this subsection we will illustrate the misspecified parameter estimators.

Estimation of $\tilde{\psi}$

Suppose $(x_{11}, x_{21}), ..., (x_{1n}, x_{2n})$ follow BIK distribution, and the objective is to estimate the misspecified parameters of BVGE distribution ($\tilde{\psi}$). This can be done by maximizing E_{BIK}[log(f_{BVGE}(X_1 ,X_2, ψ))] with respect to ψ and fixed ξ.

$$\begin{aligned}
E_{BIK}[\log(f_{BVGE}(x_1,x_2,\psi))] &= (2p_{1IK}+2p_{2IK}+p_{3IK})\log\lambda + p_{1IK}\log(\alpha_1+\alpha_3)\\
&+ p_{1IK}\log(\alpha_2)+(\alpha_1+\alpha_3-1)E_{BIK}[\log(1-e^{\lambda x_1}).I_{A_1}]\\
&+ (\alpha_2-1)E_{BIK}[\log(1-e^{\lambda x_2}).I_{A_1}]-\lambda E_{BIK}[(x_1+x_2).I_{A_1}]\\
&+ p_{2IK}\log(\alpha_2+\alpha_3)+p_{2IK}\log(\alpha_1)\\
&+ (\alpha_1-1)E_{BIK}[\log(1-e^{\lambda x_1}).I_{A_2}]\\
&+ (\alpha_2+\alpha_3-1)E_{BIK}[\log(1-e^{\lambda x_2}).I_{A_2}]\\
&- \lambda E_{BIK}[(x_1+x_2).I_{A_2}]\\
&+ p_{3IK}\log(\alpha_3)+(\alpha_1+\alpha_2+\alpha_3-1)E_{BIK}[\log(1-e^{-\lambda x}).I_{A_3}]\\
&- \lambda E_{BIK}[x.I_{A_3}],
\end{aligned}$$

where

$$\begin{aligned}
I_{A_1} &= 1, X_1 < X_2,\\
I_{A_2} &= 1, X_1 > X_2,\\
I_{A_3} &= 1, X_1 = X_2 = X,\\
p_{1IK} &= p(X_1<X_2) = \int_0^\infty\int_0^{x_2} f_{IK}(x_1;\beta_1+\beta_3,\alpha)f_{IK}(x_2;\beta_2,\alpha)dx_1dx_2\\
&= \frac{\beta_2}{\beta_1+\beta_2+\beta_3},\\
p_{2IK} &= p(X_1>X_2) = \int_0^\infty\int_{x_2}^{\infty} f_{IK}(x_1;\beta_1,\alpha)f_{IK}(x_2;\beta_2+\beta_3,\alpha)dx_1dx_2\\
&= \frac{\beta_1}{\beta_1+\beta_2+\beta_3},\\
p_{3IK} &= p(X_1=X_2=X) = \frac{\beta_3}{\beta_1+\beta_2+\beta_3}\int_0^\infty f_{IK}(x;\beta_1+\beta_2+\beta_3,\alpha)dx\\
&= \frac{\beta_3}{\beta_1+\beta_2+\beta_3}.
\end{aligned}$$

In order to calculate the previous expectations, we will introduce following lemma;

Lemma 3.

Let $V_1 \sim \text{IK}(\beta_1+\beta_3,\alpha)$, $V_2 \sim \text{IK}(\beta_2+\beta_3,\alpha)$, $V_0 \sim \text{IK}(\beta_1+\beta_2+\beta_3,\alpha)$ and $(X_1, X_2) \sim \text{BIK}(\beta_1, \beta_2, \beta_3, \alpha)$. If $g(.)$ is any Borel measurable function (for more details see, [22]), then

a) $E(g(X_1).I_{A_1}) = E(g(V_1)) - \frac{\beta_1+\beta_3}{\beta_1+\beta_2+\beta_3}E(g(V_0))$.

b) $E(g(X_1).I_{A_2}) = \frac{\beta_1}{\beta_1+\beta_2+\beta_3}E(g(V_0))$.

c) $E(g(X_1).I_{A_3}) = E(g(X_2).I_{A_3}) = \frac{\beta_3}{\beta_1+\beta_2+\beta_3}E(g(V_0))$.

d) $E(g(X_2).I_{A_1}) = \frac{\beta_2}{\beta_1+\beta_2+\beta_3}E(g(V_0))$.

e) $E(g(X_2).I_{A_2}) = E(g(V_2)) - \frac{\beta_2+\beta_3}{\beta_1+\beta_2+\beta_3}E(g(V_0))$.

Proof

$$
\begin{aligned}
a)E(g(X_1).I_{A_1}) &= \int_0^\infty \int_{x_1}^\infty g(x_1) f_{IK}(x_1;\beta_1+\beta_3,\alpha) f_{IK}(x_2;\beta_2,\alpha) dx_2 dx_1 \\
&= \int_0^\infty \int_{x_1}^\infty g(x_1)\alpha(\beta_1+\beta_3)(1+x_1)^{-(\alpha+1)}(1-(1+x_1)^{-\alpha})^{\beta_1+\beta_3-1} \\
&* \ \alpha\beta_2(1+x_2)^{-(\alpha+1)}(1-(1+x_2)^{-\alpha})^{\beta_2-1} dx_2 dx_1 \\
&= \int_0^\infty g(x_1)\alpha(\beta_1+\beta_3)(1+x_1)^{-(\alpha+1)}(1-(1+x_1)^{-\alpha})^{\beta_1+\beta_3-1} \\
&* \ [1-(1-(1+x_1)^{-\alpha})^{\beta_2}] dx_1 \\
&= \int_0^\infty g(x_1)\alpha(\beta_1+\beta_3)(1+x_1)^{-(\alpha+1)}(1-(1+x_1)^{-\alpha})^{\beta_1+\beta_3-1} dx_1 \\
&- \int_0^\infty g(x_1)\alpha(\beta_1+\beta_3)(1+x_1)^{-(\alpha+1)}(1-(1+x_1)^{-\alpha})^{\beta_1+\beta_2+\beta_3-1} dx_1 \\
&= E(g(V_1)) - \frac{\beta_1+\beta_3}{\beta_1+\beta_2+\beta_3} E(g(V_0)).
\end{aligned}
$$

Estimation of $\tilde{\xi}$

Assume the data is obtained from BVGE(ψ) distribution, to estimate the misspecified parameters of the BIK distribution ($\tilde{\xi}$). This can be done by maximizing
$E_{BVGE}[\log(f_{BIK}(X_1, X_2, \xi))]$ with respect to ξ and fixed ψ.

$$
\begin{aligned}
E_{BVGE}[\log(f_{BIK}(x_1,x_2,\xi))] &= (2p_{1GE}+2p_{2GE}+p_{3GE})\log\alpha + p_{1GE}\log(\beta_1+\beta_3) \\
&+ p_{1GE}\log(\beta_2) + p_{2GE}\log(\beta_1) + p_{2GE}log(\beta_2+\beta_3) \\
&+ p_{3GE}\log(\beta_3) - (\alpha+1)E_{BVGE}[\log(1+x_1).I_{A_1}] \\
&- (\alpha+1)E_{BVGE}[\log(1+x_2).I_{A_1}] \\
&+ (\beta_1+\beta_3-1)E_{BVGE}[\log(1-(1+x_1)^{-\alpha}).I_{A_1}] \\
&+ (\beta_2-1)E_{BVGE}[\log(1-(1+x_2)^{-\alpha}).I_{A_1}] \\
&- (\alpha+1)E_{BVGE}[\log(1+x_1).I_{A_2}] \\
&- (\alpha+1)E_{BVGE}[\log(1+x_2).I_{A_2}] \\
&+ (\beta_1-1)E_{BVGE}[\log(1-(1+x_1)^{-\alpha}).I_{A_2}] \\
&+ (\beta_2+\beta_3-1)E_{BVGE}[\log(1-(1+x_2)^{-\alpha}).I_{A_2}] \\
&- (\alpha+1)E_{BVGE}[\log(1+x).I_{A_3}] \\
&+ (\beta_1+\beta_2+\beta_3-1)E_{BVGE}[\log(1-(1+x)^{-\alpha}).I_{A_3}].
\end{aligned}
$$

where

$$
\begin{aligned}
I_{A_1} &= 1, X_1 < X_2, \\
I_{A_2} &= 1, X_1 > X_2, \\
I_{A_3} &= 1, X_1 = X_2 = X, \\
p_{1GE} &= p(X_1 < X_2) = \frac{\alpha_2}{\alpha_1+\alpha_2+\alpha_3}, \\
p_{2GE} &= p(X_1 > X_2) = \frac{\alpha_1}{\alpha_1+\alpha_2+\alpha_3}, \\
p_{3GE} &= p(X_1 = X_2 = X) = \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}.
\end{aligned}
$$

In order to calculate the previous expectations, we will use the following lemma;

Lemma 4

Let $W_1 \sim \text{GE}(\alpha_1+\alpha_3,\lambda)$, $W_2 \sim \text{GE}(\alpha_2+\alpha_3,\lambda)$, $W_0 \sim \text{GE}(\alpha_1+\alpha_2+\alpha_3,\lambda)$ and $(X_1, X_2) \sim \text{BVGE}(\alpha_1, \alpha_2, \alpha_3, \lambda)$. If $g(.)$ is any Borel measurable function, then

a) $E(g(X_1).I_{A_1}) = E(g(W_1)) - \frac{\alpha_1+\alpha_3}{\alpha_1+\alpha_2+\alpha_3}E(g(W_0))$.

b) $E(g(X_1).I_{A_2}) = \frac{\alpha_1}{\alpha_1+\alpha_2+\alpha_3}E(g(W_0))$.

c) $E(g(X_1).I_{A_3}) = E(g(X_2).I_{A_3}) = \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}E(g(W_0))$.

d) $E(g(X_2).I_{A_1}) = \frac{\alpha_2}{\alpha_1+\alpha_2+\alpha_3}E(g(W_0))$.

e) $E(g(X_2).I_{A_2}) = E(g(W_2)) - \frac{\alpha_2+\alpha_3}{\alpha_1+\alpha_2+\alpha_3}E(g(W_0))$.

Proof

Similar to Lemma 3.

To apply the previous methodologies, we need to derive the maximum likelihood estimators, which will be illustrated in the next section.

4 Maximum Likelihood Estimation

Maximum likelihood estimation technique has been used to estimate the unknown parameters. To do so we need to obtain the parameters that maximize the log likelihood function for both BIK and BVGE distributions.

4.1 Bivariate inverted Kumaraswamy distribution

The likelihood function for the BIK distribution has the following form

$$L_{BIK} = \prod_{i=1}^{n_1} f_1(x_1,x_2) \prod_{i=1}^{n_2} f_2(x_1.x_2) \prod_{i=1}^{n_3} f_3(x),$$

where

$f_1(x_1,x_2), f_2(x_1,x_2)$ and $f_3(x)$ are as defined in (2).

n_i are the number of observations falling in the range corresponding to f_i, i= 1, 2, 3.

The first derivatives with respect to the parameters are as follows

$$\begin{aligned}\frac{\partial \ln L_{BIK}}{\partial \beta_1} &= \frac{n_1}{\beta_1+\beta_3} + \frac{n_2}{\beta_1} + \sum_{i=1}^{n_1} \ln(1-(1+x_1)^{-\alpha}) + \sum_{i=1}^{n_2} \ln(1-(1+x_1)^{-\alpha}) \\ &+ \sum_{i=1}^{n_3} \ln(1-(1+x)^{-\alpha}), \\ \frac{\partial \ln L_{BIK}}{\partial \beta_2} &= \frac{n_1}{\beta_2} + \frac{n_2}{\beta_2+\beta_3} + \sum_{i=1}^{n_1} \ln(1-(1+x_2)^{-\alpha}) + \sum_{i=1}^{n_2} \ln(1-(1+x_2)^{-\alpha}) \\ &+ \sum_{i=1}^{n_3} \ln(1-(1+x)^{-\alpha}), \\ \frac{\partial \ln L_{BIK}}{\partial \beta_3} &= \frac{n_1}{\beta_1+\beta_3} + \frac{n_2}{\beta_2+\beta_3} + \frac{n_3}{\beta_3} + \sum_{i=1}^{n_1} \ln(1-(1+x_1)^{-\alpha}) + \sum_{i=1}^{n_2} \ln(1-(1+x_1)^{-\alpha}) \\ &+ \sum_{i=1}^{n_3} \ln(1-(1+x)^{-\alpha}),\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln L_{BIK}}{\partial \alpha} &= \frac{2n_1+2n_2+n_3}{\alpha} - \sum_{i=1}^{n_1} \ln(1+x_{1i}) + (\beta_1+\beta_3-1)\sum_{i=1}^{n_1} \frac{\ln(1+x_{1i})}{(1+x_{1i})^{\alpha}-1} \\ &- \sum_{i=1}^{n_1} \ln(1+x_{2i}) + (\beta_2-1)\sum_{i=1}^{n_1} \frac{\ln(1+x_{2i})}{(1+x_{2i})^{\alpha}-1} - \sum_{i=1}^{n_2} \ln(1+x_{1i}) \\ &+ (\beta_1-1)\sum_{i=1}^{n_2} \frac{\ln(1+x_{1i})}{(1+x_{1i})^{\alpha}-1} - \sum_{i=1}^{n_2} \ln(1+x_{2i}) \\ &+ (\beta_2+\beta_3-1)\sum_{i=1}^{n_2} \frac{\ln(1+x_{2i})}{(1+x_{2i})^{\alpha}-1} \\ &- \sum_{i=1}^{n_3} \ln(1+x_i) + (\beta_1+\beta_2+\beta_3-1)\sum_{i=1}^{n_3} \frac{\ln(1+x)}{(1+x)^{\alpha}-1}.\end{aligned}$$

Maximum likelihood estimators could not be obtained in closed forms, hence numerical analysis will be used. This will be illustrated in the Section 6.

4.2 Bivariate generalized exponential distribution

The likelihood function has following form

$$L_{BVGE} = \prod_{i=1}^{n_1} g_1(x_1,x_2)\prod_{i=1}^{n_2} g_2(x_1.x_2)\prod_{i=1}^{n_3} g_3(x),$$

where

$g_1(x_1,x_2), g_2(x_1,x_2)$ and $g_3(x)$ are as defined in (3).

n_i are the number of observations falling in the range corresponding to f_i, i= 1, 2, 3.

The first derivatives with respect to the parameters are as follows;

$$\begin{aligned}\frac{\partial \ln L_{BVGE}}{\partial \alpha_1} &= \frac{n_1}{\alpha_1+\alpha_3} + \frac{n_2}{\alpha_1} + \sum_{i=1}^{n_1} \ln(1-e^{-\lambda x_1}) + \sum_{i=1}^{n_2} \ln(1-e^{-\lambda x_1}) + \sum_{i=1}^{n_3} \ln(1-e^{-\lambda x}), \\ \frac{\partial \ln L_{BVGE}}{\partial \alpha_2} &= \frac{n_1}{\alpha_2} + \frac{n_2}{\alpha_2+\alpha_3} + \sum_{i=1}^{n_1} \ln(1-e^{-\lambda x_2}) + \sum_{i=1}^{n_2} \ln(1-e^{-\lambda x_2}) + \sum_{i=1}^{n_3} \ln(1-e^{-\lambda x}), \\ \frac{\partial \ln L_{BVGE}}{\partial \alpha_3} &= \frac{n_1}{\alpha_1+\alpha_3} + \frac{n_2}{\alpha_2+\alpha_3} + \frac{n_3}{\alpha_3} + \sum_{i=1}^{n_1} \ln(1-e^{-\lambda x_1}) + \sum_{i=1}^{n_2} \ln(1-e^{-\lambda x_2}) \\ &+ \sum_{i=1}^{n_3} \ln(1-e^{-\lambda x}), \\ \frac{\partial \ln L_{BVGE}}{\partial \lambda} &= \frac{2n_1+2n_2+n_3}{\lambda} - \sum_{i=1}^{n_1} x_{1i} - \sum_{i=1}^{n_1} x_{2i} - \sum_{i=1}^{n_2} x_{1i} - \sum_{i=1}^{n_2} x_{2i} - \sum_{i=1}^{n_3} x_i \\ &+ \sum_{i=1}^{n_1} \frac{(\alpha_1+\alpha_3-1)}{e^{\lambda x_{1i}}-1} + \sum_{i=1}^{n_1} \frac{(\alpha_2-1)}{e^{\lambda x_{2i}}-1} + \sum_{i=1}^{n_2} \frac{(\alpha_1-1)}{e^{\lambda x_{1i}}-1} + \sum_{i=1}^{n_2} \frac{(\alpha_2+\alpha_3-1)}{e^{\lambda x_{2i}}-1} \\ &+ \sum_{i=1}^{n_3} \frac{(\alpha_1+\alpha_2+\alpha_3-1)}{e^{\lambda x_i}-1}.\end{aligned}$$

Similar to the case of BIK, maximum likelihood estimates will be obtained numerically.

Also, to compare the two illustrated methodologies of discrimination, we need to compute the probability of correct selection (PCS). The method with higher PCS is better as it is corresponding to minimum Type-I error (1- PCS). All the results will be illustrated in details in Section 6. Probability of correct selection is calculated as follows;

In case data are from BIK distribution, probability of correct selection using asymptotic distribution of likelihood ratio test is calculated as follows;

$$\begin{aligned} PCS &= p(T>0) = p(z > \frac{-E_{BIK}(T)}{\sqrt{Var_{BIK}(T)}}) \\ &= p(z > \frac{-\sqrt{n}AM_{BIK}(T)}{\sqrt{AV_{BIK}(T)}}) \\ &= \Phi(\frac{\sqrt{n}AM_{BIK}(T)}{AV_{BIK}(T)}), \end{aligned} \quad (4)$$

where, Φ is the distribution function of standard normal random variable and $z = \frac{T-E_{BIK}(T)}{Var_{BIK}(T)}$.

Suppose data are from BVGE distribution, PCS using asymptotic distribution of LRT is obtained as follows;

$$\begin{aligned} PCS &= p(T<0) = p(z < \frac{-E_{BVGE}(T)}{\sqrt{Var_{BVGE}(T)}}) \\ &= p(z < \frac{-\sqrt{n}AM_{BVGE}(T)}{\sqrt{AV_{BVGE}(T)}}) \\ &= \Phi(\frac{-\sqrt{n}AM_{BVGE}(T)}{AV_{BVGE}(T)}). \end{aligned} \quad (5)$$

5 Determination of Sample Size

In this section, we illustrate a method to determine the minimum sample size needed to discriminate between BIK and BVGE distributions, for a specified probability of correct selection (p^*) and tolerance level (D^*).

Tolerance level is determined according to the closeness between BIK and BVGE distributions, which is measured by the distance between them. Here, we generalize the distance method to be applied in case of bivariate distributions.

One of the most commonly used method to measure the distance between two density functions is Hellinger distance (H), which can be calculated as follows (see [23]);

$$H = 0.5 \int \dots \int (\sqrt{f(\underline{x})} - \sqrt{g(\underline{x})}\,)^2 d\underline{x},$$

it can be rewritten in case of bivariate Marshall-Olkin family as follows;

$$\begin{aligned} H = 0.5\{ &\int_0^\infty \int_0^{x_2} (\sqrt{f_1(x_1,x_2)} - \sqrt{g_1(x_1,x_2)}\,)^2 dx_1 dx_2 \\ &+ \int_0^\infty \int_0^{x_1} (\sqrt{f_2(x_1,x_2)} - \sqrt{g_2(x_1,x_2)}\,)^2 dx_2 dx_1 \\ &+ \int_0^\infty (\sqrt{f_3(x)} - \sqrt{g_3(x)})^2 dx\}, \end{aligned} \quad (6)$$

where;

$f_1(x_1,x_2)$, $f_2(x_1,x_2)$ and $f_3(x)$ are as defined in (2).

$g_1(x_1,x_2)$, $g_2(x_1,x_2)$ and $g_3(x)$ are as defined in (3).

Tolerance level is defined in terms of some distance between BIK and BVGE distributions. Once tolerance level is specified, the two functions are considered significantly different if the difference between them is greater than or equal the tolerance level. Here the tolerance level is defined based on Hellinger distance defined in (6).

Now, the asymptotic distribution derived in Section 3.1 will be used with Hellinger distance to determine the sample size.

Suppose Data come from BIK distribution, using Equation (4) it can be seen that for a specified probability of correct selection (p^*) that;

$$\Phi(\frac{\sqrt{n}AM_{BIK}(T)}{\sqrt{AV_{BIK}(T)}}) = p^*,$$

where, Φ is the distribution function of standard normal random variable. Hence, the sample size (n) can be obtained as

$$n = \frac{z_{p^*}^2 AV_{BIK}(T)}{[AM_{BIK}(T)]^2},$$

where z_{p^*} is the p^*-th percentile point of a standard normal distribution. In Table 1, n is reported for different values of the parameters and $p^* = 0.9$.

Similarly, using Equation (5) we can obtain the sample size in case of BVGE distribution (reported in Table 2, for different values of the parameters and $p^* = 0.9$).

Now we illustrate how we can determine the minimum sample size using both tolerance level (D^*) and specified PCS (p^*). If data come from BIK distribution, suppose the specified probability of correct selection p^* = 0.9 and the tolerance level is given in terms of Hellinger distance as $D^* = 0.10$. Tolerance level $D^* = 0.10$ means that, one wants to discriminate between BVGE and BIK when the Hellinger distance is at least 0.10. From Table 1, n = max (30, 32) = 32. Now, for data from BVGE distribution with p^* = 0.9 and $D^* = 0.34$, from Table 2, n = max (428, 415) = 428. It can be seen that, minimum sample size (n) in case of BIK is about $\frac{1}{14}$ times that in case of BVGE. This may give a preference of using BIK in representing bivariate lifetime data.

Table 1: Sample size(n) and Hellinger distance (H) between BIK(β_1, $\beta_2 = 1.2$, $\beta_3 = 1.2$, $\alpha = 2.5$) and BVGE($\tilde{\alpha}_1$, $\tilde{\alpha}_2$, $\tilde{\alpha}_3$, $\tilde{\lambda}$).

β_1	2.2	2.4	2.6	2.8	3.8
n	30	30	30	30	32
H	0.13	0.12	0.11	0.11	0.10

Table 2: Sample size(n) and Hellinger distance (H) between BVGE($\alpha_1 = 0.5$, $\alpha_2 = 0.5$, $\alpha_3 = 0.5$, λ) and BIK($\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\beta}_3$, $\tilde{\alpha}$).

λ	1.1	0.9	0.7	0.5	0.4
n	337	402	431	428	415
H	0.32	0.32	0.32	0.34	0.51

6 Numerical Results

In this section, we will explain a numerical analysis to study the performance of the new generalized method (RMKLD), and illustrate the performance of the asymptotic distribution of the LRT.

The comparison is done using the PCS. For LRT method, we compute it based on both simulation study (SIM) and asymptotic distribution (AD). Moreover, we compute PCS for RMKLD method.

In case of BIK and using LRT, PCS $= p(T > 0)$. While for BVGE, PCS $= p(T < 0)$. Using asymptotic distribution, PCS can be calculated using Equation (4) and (5) for BIK and BVGE, respectively.

6.1 A simulation study

We will present two cases, the first one when data is generated from BIK distribution and the second when data is obtained from BVGE distribution. The following steps are used to perform the simulation for LRT method from BIK distribution.

Step 1: n independent samples (X_1, X_2) from BIK distribution have been generated as follows

a) Generate U_1, U_2 and U_3 from uniform(0,1).

b) Compute $T_1 = (1-U_1^{\beta_1})^{\frac{1}{\alpha}}, T_2 = (1-U_2^{\beta_2})^{\frac{1}{\alpha}}$ and $T_3 = (1-U_3^{\beta_3})^{\frac{1}{\alpha}}$.

c) Define $Z_1 = \frac{1}{T_1} - 1, Z_2 = \frac{1}{T_2} - 1$ and $Z_3 = \frac{1}{T_3} - 1$.

d) Obtain $X_1 = \max(Z_1, Z_3)$ and $X_2 = \max(Z_2, Z_3)$.

Step 2: The maximum likelihood estimates for both L_{BIK} and L_{BVGE} have been obtained, and then used to get $T = L_{BIK}(x_1, x_2, \hat{\xi}) - L_{BVGE}(x_1, x_2, \hat{\psi})$ which has been stored.

Step 3: The three previous steps have been repeated 10,000 times.

Step 4: The number of times T $>$ 0 has been counted and then the approximate probability of correct selection has been obtained

$$\text{PCS} = \frac{\text{number of times T} > 0}{10000}.$$

To obtain the PCS for asymptotic distribution, one needs to compute the asymptotic mean and variance for the test statistic T. Since as proved in Section 3, T statistic is asymptotically normally distributed, we used consistent unbiased estimators (i.e. $\bar{X}$ and S^2) for the mean and variance using 10,000 replications. Then from the normal distribution, PCS could be obtained (i.e. using Equation 4). Now, we will illustrate the steps used to get the PCS using RMKLD method.

Repeat Step 1 and 2.

Step 3: $KLD(f(x_1,x_2)\|g(x_1,x_2))$ and $KLD(g(x_1,x_2)\|f(x_1,x_2))$ has been calculated.

Step 4: $RMKLD(f(x_1,x_2), g(x_1,x_2)) = ln(\frac{KLD(f(x_1,x_2)\|g(x_1,x_2))}{KLD(g(x_1,x_2)\|f(x_1,x_2))})$ has been obtained.

Step 5: The four previous steps have been repeated 10,000 times.

Step 6: The number of times RMKLD $<$ 0 has been counted, then the approximate probability of correct selection has been obtained

$$\text{PCS} = \frac{\text{number of times RMKLD} < 0}{10000}.$$

To apply LRT and RMKLD methods when data is obtained from BVGE distribution, similar steps can be used except that:

- The data is generated from BVGE distribution using the following steps. a) Generate U_1, U_2 and U_3 from generalized exponential distribution.

b) Obtain $X_1 = \max(U_1, U_3)$ and $X_2 = \max(U_2, U_3)$.

- For LRT method we count number of times T $<$ 0.

- For AD, PCS = P(T $<$ 0).

- For RMKLD method we count the number times RMKLD $>$ 0.

We used different sample sizes varying from 40 to 800 in Tables 3 and 4. The results from BIK and BVGE distributions are illustrated in Tables 3 and 4, respectively. It can be seen that:

As the sample size increases, PCS increases as expected for all used methods. For example in Table 3, for $\beta_1 = 2.8$, the PCS increased from 0.9940 at n = 40 to 1 at n = 100 using RMKLD method. Also, it increased from 0.8384 at n = 40 to 0.9996 at n = 800 using AD. Moreover, PCS increased from 0.9031 at n = 40 to 1 at n =300 using SIM.

when BIK is the parent distribution, we can see that as β_1 increases the PCS increases for different sample sizes and different used methods(RMKLD and LRT method using both SIM and AD). Besides, for BVGE distribution, it can be noticed that as λ decreases the PCS increases for different sample sizes and different used methods (RMKLD and LRT method using both SIM and AD).

Moreover, the AD provides a good approximation for PCS when compared to the values obtained from simulation study. For example, In Table 4, at n=40 and $\lambda = 1.1$, PCS from AD is 0.9012 and from SIM is 0.8997.

In addition, RMKLD provides the best results, that is, it has the highest PCS or equivalently the minimum Type-I error. For example, from Table 4, for $\lambda = 1.1$ and n=40, Type-I error for AD equals 0.0988 and for SIM equals 0.1003 while for RMKLD method equals 0.0094.

Table 3: PCS for BIK distribution based on simulation (SIM), asymptotic distribution (AD) and RMKLD.

				β_2=1.2	β_3 = 1.2	α =2.5			
β_1	n	40	60	80	100	200	300	500	800
2.2	AD	0.8344	0.8799	0.9022	0.9228	0.9671	0.9864	0.9964	0.9996
	SIM	0.9011	0.9528	0.9734	0.9840	0.9989	1	1	1
	RMKLD	0.9534	0.9777	0.9887	0.9940	0.9999	1	1	1
2.4	AD	0.8349	0.8768	0.8963	0.9145	0.9676	0.9866	0.9965	0.9996
	SIM	0.9012	0.9533	0.9725	0.9865	0.9987	1	1	1
	RMKLD	0.9756	0.9890	0.9948	0.9990	1	1	1	1
2.6	AD	0.8378	0.8777	0.8971	0.9156	0.9679	0.9868	0.9966	0.9996
	SIM	0.9020	0.9536	0.9735	0.9865	0.9989	1	1	1
	RMKLD	0.9876	0.9959	0.9990	0.9999	1	1	1	1
2.8	AD	0.8384	0.8786	0.8980	0.9187	0.9683	0.9870	0.9967	0.9996
	SIM	0.9031	0.9542	0.9736	0.9873	0.9988	1	1	1
	RMKLD	0.9940	0.9983	0.9997	1	1	1	1	1
3.8	AD	0.8392	0.8784	0.8983	0.9211	0.9700	0.9879	0.9971	0.9996
	SIM	0.9069	0.9568	0.9759	0.9884	0.9989	1	1	1
	RMKLD	0.9997	1	1	1	1	1	1	1

Table 4: PCS for BVGE distribution based on simulation (SIM), asymptotic distribution (AD) and RMKLD.

			α_1=0.5	α_2 = 0.5	α_3 =0.5				
λ	n	40	60	80	100	200	300	500	800
1.1	AD	0.9012	0.9482	0.9698	0.9808	0.9984	0.9998	0.9999	1
	SIM	0.8997	0.9426	0.9691	0.9767	0.9976	0.9996	1	1
	RMKLD	0.9906	0.9979	0.9999	0.9999	1	1	1	1
0.9	AD	0.9207	0.9625	0.9802	0.9884	0.9994	0.9999	1	1
	SIM	0.9157	0.9581	0.9785	0.9856	0.9993	0.9998	1	1
	RMKLD	0.9923	0.9986	0.9999	0.9999	1	1	1	1
0.7	AD	0.9409	0.9757	0.9887	0.9942	0.9998	0.9999	1	1
	SIM	0.9359	0.9729	0.9868	0.9928	0.9998	0.9999	1	1
	RMKLD	0.9939	0.9990	0.9999	0.9999	1	1	1	1
0.5	AD	0.9608	0.9869	0.9949	0.9978	1	1	1	1
	SIM	0.9544	0.9850	0.9929	0.9970	1	1	1	1
	RMKLD	0.9953	0.9989	0.9999	0.9999	1	1	1	1
0.4	AD	0.9704	0.9914	0.9971	0.9989	1	1	1	1
	SIM	0.9653	0.9912	0.9957	0.9981	1	1	1	1
	RMKLD	0.9956	0.9993	1	1	1	1	1	1

6.2 A Real Dataset

We analyze a real dataset obtained from [24]. The variables in this dataset are;

X_1 : the time in minutes of the first kick goal scored by any team.

X_2: the time of the first goal of any type scored by the home team.

In the data structure, $X_1 = X_2$ has positive probability, hence a bivariate Marshall- Olkin family is a good choice for analyzing this dataset as it has a singular part. Here, we have analyzed this dataset using two members of the family, namely, BIK and BVGE distributions. The data has been divided by 100 for the parameters to be of the same order, this is not going to make any changes in the statistical inference.

Before analyzing the data, as a part of model checking, we applied goodness of fit tests in order to see if the fit based on univariate inverted Kumaraswamy (IK) distributions were reasonable in this case. We computed the Cramer-Von Mises (W^*) and Anderson-Darling statistic (A^*). The values of these statistics and the corresponding p-values (in brackets) for X_1, X_2 and $\max(X_1, X_2)$ are:

Table 5: Cramer-Von Mises (W^*) and Anderson-Darling statistic (A^*) for IK.

	X_1	X_2	$\max(X_1,X_2)$
A^*	0.966 (0.375)	0.549 (0.696)	1.395 (0.204)
W^*	0.139 (0.426)	0.083 (0.679)	0.233 (0.212)

Similarly, W^* and A^* are computed for generalized exponential (GE) distributions. The values of these statistics and the corresponding p-values (in brackets) for X_1, X_2 and $\max(X_1, X_2)$ are:

Table 6: Cramer-Von Mises (W^*) and Anderson-Darling statistic (A^*) for GE.

	X_1	X_2	$\max(X_1,X_2)$
A^*	0.656 (0.595)	0.394 (0.854)	0.985 (0.365)
W^*	0.092 (0.627)	0.061 (0.811)	0.165 (0.347)

Based on the values of the statistics A^* and W^* and the corresponding p-values, it is decided that the inverted Kumaraswamy distribution and generalized exponential distribution cannot be rejected for modeling the marginals and the maximum.

Since both BIK and BVGE can be used to analyze this dataset, we need to select the distribution that best fit the data. Maximum likelihood estimators for the unknown parameters of BIK and BVGE have been obtained. From the simulation study, RMKLD provided the best fit, so we apply it here as our test statistic for discrimination. In this dataset, RMKLD = -1.147659 which suggests that BIK provides a better fit.

7 Conclusion

The main aim of this paper is to generalize the RMKLD test to the bivariate case. Here, we generalized it with application to the bivariate Marshall-Olkin family.

To study RMKLD, we applied it to discriminate between the recently derived BIK and BVGE distributions. Also, the LRT was used and its asymptotic distribution was derived and used to obtain the minimum sample size required for discrimination. Moreover, a comparison to select the best method is performed using PCS.

It is observed that RMKLD test has the same trend as LRT method, in the sense that when sample size increases the probability of correct selection increases for both methods. Also, it is found that RMKLD works better than LRT method as it has a smaller Type-I error. Moreover, it can be observed that, the asymptotic distribution results of LRT method work quite well.

Finally, a real dataset is analyzed. Using the RMKLD method, we found that bivariate inverted Kumaraswamy distribution is chosen for this dataset.

The generalization of RMKLD method to the bivariate case has several applications as it can be effectively used to discriminate between other different bivariate lifetime distributions.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] D. Kundu and R.D. Gupta, Bivariate generalized exponential distribution, *J. Multivar. Anal.*, **100 (4)**, 581-593 (2009).
[2] D.R. Cox, *Tests of separate families of hypotheses,* in Proc. Fourth Berkeley Symp. on Math. Statist. and Prob. Berkely, University of California Press, 105-123,(1961).
[3] D. R. Cox, Further results on tests of separate families of hypotheses, *J. R. Stat. Soc. Series B Stat.l Methodol.*, **24 (2)**, 406-424 (1962).
[4] A.C. Atkinson, A method for discriminating between models, *J. R. Stat. Soc. Series B Methodol.*, **32 (3)**, 323-353 (1970).
[5] R. Dumonceaux and C.E. Antle, Discrimination between the log-normal and the Weibull distributions, *Technometrics.*, **15 (4)**, 923-926 (1973).
[6] A.Z. Algamal and H.T.M. Ali, Discrimination between the Weibull and the inverse Gaussian distributions, *IOSR J. Mathematics.*, **13 (2)**, 74-83 (2017).
[7] I. Ba and F. Ashkar, Discrimination between a group of three-parameter distributions for hydro-meteorological frequency modeling, *Can. J. Civ. Eng.*, **45 (5)**, 351-365 (2018).
[8] A.A. Bromideh and R. Valizadeh, Discrimination between Gamma and log-normal distributions by ratio of minimized Kullback-Leibler divergence, *Pak. j. stat. oper. res.*, **9 (4)**, 441-451 (2013).
[9] E.A. Elsherpieny, H.Z. Muhammed and N.U. Radwan, On discriminating between Gamma and log-logistic distributions in case of progressive Type II censoring, *Pak. j. stat. oper. res.*, **13 (1)**, 157-183 (2017).
[10] A.K. Dey, D. Kundu, Discriminating between the bivariate generalized exponential and bivariate Weibull distributions, *Chil. J. Stat.*, **3 (1)**, 93-110 (2012).
[11] M.M. Ristić, B.V. Popović, K. Zografos and N. Balakrishnan, Discrimination among bivariate beta-generated distributions, *J. Theoretical. Appl. Stat.*, **52 (2)**, 303-320 (2018).
[12] K.P. Burnham and D.R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach,* 2nd eds., Springer, New York, (2002).
[13] A.A. Bromideh, Discrimination between Weibull and log-normal distributions based on Kullback-Leibler divergence, *Istanbul University Econometrics and Statistics e-Journal.*, **16 (1)**, 44-54 (2012).
[14] A.W. Marshall and I. Olkin, A multivariate exponential distribution, *J. Am. Stat. Assoc.*, **62 (317)**, 30-44 (1967).
[15] D. Kundu and R.D. Gupta, A class of bivariate models with proportional reversed hazard marginals, *Sankhya B.*, **72 (2)**, 236-253 (2010).
[16] W. Barreto-Souza and A.J. Lemonte, Bivariate Kumaraswamy distribution: properties and a new method to generate bivariate classes, *J. Theoretical. Appl. Stat.*, **47 (6)**, 1321-1342 (2013).
[17] D. Kundu and A.K. Gupta, On bivariate inverse Weibull distribution, *Braz. J. Probab. Stat.* **31 (2)**, 275-302 (2017).
[18] D.D. Hanagal and R. Sharma, Bayesian inference in Marshall Olkin bivariate Exponential shared Gamma frailty regression model under random censoring, *Commun. Stat. Theory Methods.*, **44 (1)**, 24-47 (2015).
[19] H. White, Regularity conditions for cox's test of non-nested hypotheses, *J. Econom.*, **19 (2)**, 301-318 (1982).
[20] F.G. Pascual, Maximum likelihood estimation under misspecified lognormal and Weibull distributions, *Commun. Stat. Simul. Comput.*, **34 (3)**, 503-524 (2005).
[21] H. White, Maximum likelihood estimation of misspecified models, *Econometrica.*, **19 (2)**, 301-318 (1982).
[22] D.L. Cohen, *Measure Theory*, Birkhauser, Boston. (1980).
[23] L. Pardo, *Statistical Inference Based on Divergence Measures*, Taylor & Francis Group, New York, (2006).
[24] S.G. Meintanis, Test of fit for Marshall-Olkin distributions with applications, *J. Stat. Plan. Inference.*, **137 (12)**, 3954-3963 (2007).