A new Bivariate Distribution with Weighted Exponential Marginals and its Multivariate Generalization

D. K. Al-Mutairi† & M. E. Ghitany† & D. Kundu‡

Abstract

Gupta and Kundu [6] recently introduced a new class of weighted exponential distribution. It is observed that the proposed weighted exponential distribution is very flexible and can be used quite effectively to analyze skewed data. In this paper we propose a new bivariate distribution with the weighted exponential marginals. Different properties of this new bivariate distribution have been investigated. This new family has three unknown parameters, and it is observed that the maximum likelihood estimators of the unknown parameters can be obtained by solving a one-dimensional optimization procedure. We obtain the asymptotic distribution of the maximum likelihood estimators. Small simulation experiments have been performed to see the behavior of the maximum likelihood estimators, and one data analysis has been presented for illustrative purposes. Finally we discuss the multivariate generalization of the proposed model.

Keywords and Phrases: Weighted exponential distribution; maximum likelihood estimators; skew-normal distribution; hidden truncation models; confidence intervals.

Postal address: † Department of Statistics and Operations Research, Faculty of Science, Kuwait University, P.O. Box 5969 Safat, Kuwait 13060.

Corresponding author: ‡Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, INDIA; Phone: 91-512-2597141, Fax: 91-512-2597500; e-mail: kundu@iitk.ac.in.

1 Introduction

Azzalini [2]'s skew-normal distribution has received considerable attention in recent times and it has been used quite successfully to analyze skew data sets. Azzalini [2] in his seminal paper first introduced a shape parameter to a normal distribution. Since then extensive work has been done to introduce shape parameter to different symmetric distributions. See for example the recent edited monograph by Genton [5] in this respect. Interestingly, although extensive work has been done to introduce shape parameter to a symmetric distribution, but not much attention has been paid to introduce an extra shape parameter to a skewed distribution. Recently, Gupta and Kundu [6] adopt the similar approach as of Azzalini [2] and introduce a shape parameter to an exponential distribution and they name it as a weighted exponential (WE) distribution. They study different properties of this two-parameter WE distribution and study their inferential procedures. It is observed that the proposed WE distribution has several interesting properties and it can be used quite effectively to analyze skewed data. In many situation, it may perform better than the well known gamma, log-normal, Weibull or generalized exponential distributions, see Gupta and Kundu [6].

The main aim of this paper is to consider a new bivariate absolute continuous distribution whose marginals are WE distributions. We call this new bivariate distribution as the bivariate distribution with the weighted exponential marginals and we will denote it by BWE. The BWE has three parameters, similarly as the Block and Basu's [3] bivariate exponential distribution. Here also there are two scale parameters and one shape parameter. It has also several interesting properties. It has some other nice interpretations also. The generation of random samples from the BWE is quite straight forward, which makes it very convenient to perform the simulation experiments or parametric bootstrapping.

The cumulative distribution function and also the joint moment generating function can be obtained in explicit forms. The single and product moments of the BWE can be obtained explicitly. The conditional distribution and different moments of the conditional distributions have been derived. It is observed that the correlation between the two components is always positive and it can vary between 0 and 1. The distribution of the convolutions of n identically distributed BWE random variables has been investigated, and it is observed that the joint probability density function of the convolution can be expressed explicitly.

The maximum likelihood estimators (MLEs) of the unknown parameters can be obtained by solving a one-dimensional optimization problem. The asymptotic distribution of the MLEs have been derived. Monte Carlo simulations have been performed to see the behavior of the MLEs, and one data analysis has been performed for illustrative purposes. The multivariate generalization is also quite straight forward, and can be developed along the same lines as done here for the bivariate case.

The rest of the paper is organized as follows. In section 2, we briefly describe the WE distribution. In section 3 we introduce the BWE. Different properties are discussed in section 4. The MLEs and their properties are discussed in section 5. In section 6 we present the simulation and one data analysis results. Multivariate generalization has been suggested in section 7, and finally we conclude the paper in section 8.

2 Weighted Exponential Distribution

The random variable X is said to have a weighted exponential (WE) distribution with the shape and scale parameters as $\alpha > 0$ and $\lambda > 0$ respectively, if it has the following probability density (PDF)

$$f_X(x;\alpha,\lambda) = \frac{\alpha+1}{\alpha}\lambda e^{-\lambda x}\left(1 - e^{-\alpha\lambda x}\right), \qquad x \geq 0. \tag{1}$$

The corresponding cumulative distribution function (CDF) for $x \geq 0$, becomes;

$$F_X(x;\alpha,\lambda) = \frac{\alpha+1}{\alpha}\left[1 - e^{-\lambda x} - \frac{1}{\alpha+1}\left(1 - e^{-(\alpha+1)\lambda x}\right)\right]. \tag{2}$$

From now on a WE distribution with the shape and scale parameters as α and λ respectively will be denoted WE(α, λ).

Note that the WE(α, λ) can be obtained exactly the same way Azzalini [2] obtained the skew-normal distribution from two *i.i.d.* normal distributions. Suppose X_1 and X_2 are *i.i.d.* $\exp(\lambda)$, *i.e.* an exponential random variable with mean $\frac{1}{\lambda}$, then for $\alpha > 0$ consider a new random variable $X = X_1$, if $\alpha X_1 \geq X_2$. Then X follows WE(α, λ).

It is observed that the PDF of WE(α, λ) is unimodal, and it has increasing hazard rate for all values of α. The shapes of the PDFs of the WE distribution are very similar to the corresponding shapes of the PDFs of the well studied Weibull, gamma, log-normal and generalized exponential distributions. Therefore, it can be used quite effectively to analyze positively skewed data, and it can be used quite conveniently for censored data also, which often arises in practice.

It is interesting to observe that the WE model can be observed as a hidden truncation model in the sense of Arnold and Beaver [1]. Moreover, if U and V are two independent exponential random variables with mean $\frac{1}{\lambda}$ and $\frac{1}{\lambda(1+\alpha)}$ respectively, then it can be seen that

$$X \stackrel{d}{=} U + V, \tag{3}$$

here $\stackrel{d}{=}$ means equal in distribution. The above (3) representation of X is very useful to generate data from the WE distribution. For several other interesting properties and interpretations, the readers are referred to the original paper by Gupta and Kundu [6]

3 Bivariate Distribution with Weighted Exponential Marginals

In this section we introduce the BWE and provide its joint cumulative distribution function (CDF) and the joint probability density function (PDF). Suppose X_1 follows ($\sim$) $\exp(\lambda_1)$, $X_2 \sim \exp(\lambda_2)$ and $X_3 \sim \exp(1)$. For any $\lambda_3 > 0$, consider the following bivariate random variables $X = X_1$ and $Y = X_2$, if $\lambda_3 X_1 \geq X_3$ and $\lambda_3 X_2 \geq X_3$, *i.e.* $X_3 \leq \lambda_3 \min\{X_1, X_2\}$. The new bivariate random variables (X, Y) is called BWE with parameters $\lambda_1, \lambda_2, \lambda_3$, and it will be denoted by BWE$(\lambda_1, \lambda_2, \lambda_3)$. The joint CDF and the joint PDF can be obtained as follows.

Theorem 3.1: If $(X, Y) \sim$ BWE$(\lambda_1, \lambda_2, \lambda_3)$, then the joint CDF of (X, Y) is

$$\begin{aligned} F_{X,Y}(x,y) &= \left(1 - e^{-\lambda z}\right) - \frac{\lambda}{\lambda_1 + \lambda_3} e^{-\lambda_2 y}\left(1 - e^{-(\lambda_1+\lambda_3)z}\right) - \frac{\lambda}{\lambda_2 + \lambda_3} e^{-\lambda_1 x}\left(1 - e^{-(\lambda_2+\lambda_3)z}\right) \\ &\quad + \frac{\lambda}{\lambda_3} e^{-\lambda_1 x} e^{-\lambda_2 y}\left(1 - e^{-\lambda_3 z}\right), \end{aligned} \tag{4}$$

where $z = \min\{x, y\}$ and $\lambda = \lambda_1 + \lambda_2 + \lambda_3$.

Proof: Note that,

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \leq x, Y \leq y) = P(X_1 \leq x, X_2 \leq y | \lambda_3 X_1 > X_3, \lambda_3 X_2 > X_3) \\ &= \frac{P(X_3 < \lambda_3 X_1 \leq \lambda_3 x, X_3 < \lambda_3 X_2 \leq \lambda_3 y)}{P(\lambda_3 X_1 > X_3, \lambda_3 X_2 > X_3)}. \end{aligned} \tag{5}$$

Since the numerator of the right hand side of (5) is

$$\int_0^{\lambda_3 z} e^{-u}\left(e^{-\frac{\lambda_1 u}{\lambda_3}} - e^{-\lambda_1 x}\right)\left(e^{-\frac{\lambda_2 u}{\lambda_3}} - e^{-\lambda_2 y}\right) du \tag{6}$$

and the denominator is

$$\int_0^{\infty}\left(e^{-u} e^{-\frac{\lambda_1 u}{\lambda_3}} e^{-\frac{\lambda_2 u}{\lambda_3}}\right) du = \frac{\lambda_3}{\lambda}, \tag{7}$$

the result follows after simplifications.

Theorem 3.2: Suppose $(X, Y) \sim \text{BWE}(\lambda_1, \lambda_2, \lambda_3)$ and z is same as before.

(a) The joint PDF of X and Y for $x > 0$ and $y > 0$, is

$$f_{X,Y}(x, y) = \frac{\lambda\lambda_1\lambda_2}{\lambda_3} e^{-\lambda_1 x} e^{-\lambda_2 y} \left(1 - e^{-\lambda_3 z}\right). \tag{8}$$

(b) $X \sim \text{WE}\left(\frac{\lambda_2 + \lambda_3}{\lambda_1}, \lambda_1\right), \quad Y \sim \text{WE}\left(\frac{\lambda_1 + \lambda_3}{\lambda_2}, \lambda_2\right).$

(c) The conditional PDF of X given $Y = y$ is

$$f_{X|Y=y}(x) = \frac{\lambda_1(\lambda_1 + \lambda_3)}{\lambda_3 \left(1 - e^{-(\lambda_1+\lambda_3)y}\right)} \, e^{-\lambda_1 x} \left(1 - e^{-\lambda_3 z}\right). \tag{9}$$

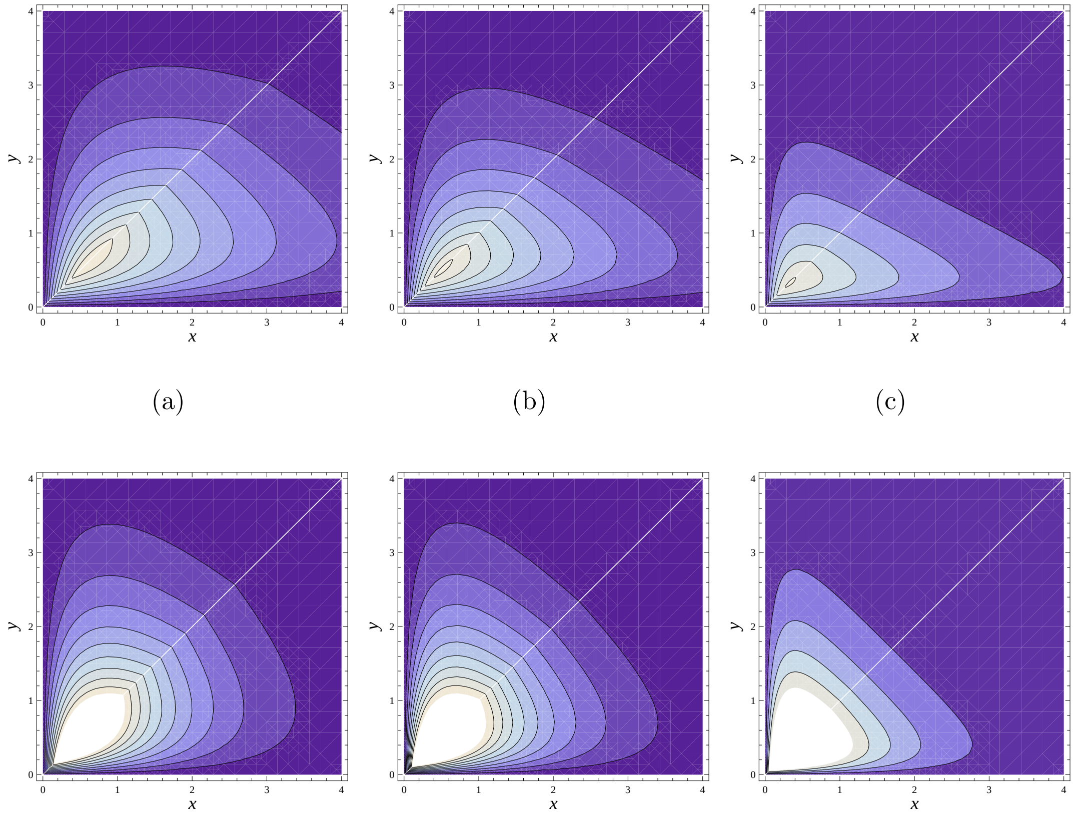
(d) The conditional moment generating function of X given $Y = y$ for $|t| < \lambda_1$, is

$$\begin{aligned} E(e^{tX}|Y = y) \quad &= \quad \frac{\lambda_1(\lambda_1 + \lambda_3)}{\lambda_3 \left(1 - e^{-(\lambda_1+\lambda_3)y}\right)} \left[\frac{1}{\lambda_1 - t}\left(1 - e^{-(\lambda_1 - t)}\right) - \frac{1}{\lambda_1 + \lambda_3 - t}\left(1 - e^{-(\lambda_1+\lambda_3-t)y}\right)\right. \\ &\quad \left. + \left(1 - e^{-\lambda_3 y}\right) \frac{e^{-(\lambda_1 - t)}}{1 - e^{-\lambda_3 y}}\right]. \end{aligned} \tag{10}$$

Proof: (a) Can be obtained by observing the fact $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$. The proofs of (b), (c) and (d) are quite routine and therefore they are omitted.

Before progressing further, we would like to mention how the BWE may occur in practice. Suppose a system has two components. Each component is subject to independent stress say X_1 and X_2 respectively. Moreover, X_1 and X_2 are observed only if $X_1 \geq cX_3$ and $X_2 \geq cX_3$ for some $c > 0$. Here X_3 is a random threshold and it is independent of X_1 and X_2. The observed (X_1, X_2) has BWE distribution if X_1, X_2 and X_3 are independent exponential random variables.

The contour plots of $f_{X,Y}(x, y)$ for different values of λ_1, λ_2 and λ_3 are provided in Figure 1. From Figure 1, it is clear that the joint PDF can take different shapes and therefore BWE can be used quite effectively to analyze skewed bivariate data sets. The following result indicates that the joint PDF of (X, Y) is unimodal.



(a) (b) (c)

(d) (e) (f)

Figure 1: Contour plots of the joint PDF for $(\lambda_1, \lambda_2, \lambda_3)$: (a) (0.5, 1, 0.25) (b) (0.5, 1, 1) (c) (0.5, 1, 4) (d) (1, 1, 0.25) (e) (1, 1, 1) (f) (1, 1, 4).

Theorem 3.3: For all $\lambda_1, \lambda_2, \lambda_3 > 0$, $f_{X,Y}(x, y)$ is unimodal in $x, y > 0$. The mode occurs at the point (x_0, y_0) where

$$x_0 = y_0 = \frac{1}{\lambda_3} \ln\Big(1 + \frac{\lambda_3}{\lambda_1 + \lambda_2}\Big).$$

Proof. Consider the following three cases:

Case 1. Consider the region $R_1 = \{(x, y) : 0 < x < y\}$ where

$$f_{X,Y}(x, y) = \frac{\lambda\lambda_1\lambda_2}{\lambda_3}\, e^{-\lambda_1 x}\, e^{-\lambda_2 y}\, (1 - e^{-\lambda_3 x}), \qquad x < y.$$

The first partial derivative of $f_{X,Y}(x,y)$ are given by

$$\begin{aligned}\frac{\partial f_{X,Y}(x,y)}{\partial x} &= -\frac{\lambda\lambda_1\lambda_2}{\lambda_3}\; e^{-\lambda_1 x}\; e^{-\lambda_2 y}\; [\; \lambda_1 - (\lambda_1+\lambda_3)\; e^{-\lambda_3 x}\;],\\ \frac{\partial f_{X,Y}(x,y)}{\partial y} &= -\frac{\lambda\lambda_1\lambda_2^2}{\lambda_3}\; e^{-\lambda_1 x}\; e^{-\lambda_2 y}\; (1-e^{-\lambda_3 x}).\end{aligned}$$

Since $\frac{\partial f_{X,Y}(x_1,y_1)}{\partial x} = 0$ and $\frac{\partial f_{X,Y}(x_1,y_1)}{\partial y} = 0$, respectively, lead to $x_1 = \frac{1}{\lambda_3}\; \ln(\frac{\lambda_1+\lambda_3}{\lambda_1})$ and y_1 does not exist. That is, there does not exist critical point (x_1, y_1) in the region R_1, and hence no local minima or maxima can occur in this region.

Case 2. Consider the region $R_3 = \{(x,y) : x > y > 0\}$ where

$$f_{X,Y}(x,y) = \frac{\lambda\lambda_1\lambda_2}{\lambda_3}\; e^{-\lambda_1 x}\; e^{-\lambda_2 y}\; (1-e^{-\lambda_3 y}), \qquad x > y.$$

By symmetry, as in case 1, there does not exist critical point (x_2, y_2) in the region R_2, and hence no local minima or maxima can occur in this region.

Case 3. Consider the region $R_3 = \{(x,y) : x = y > 0\}$ where

$$f_{X,Y}(x,x) = \frac{\lambda\lambda_1\lambda_2}{\lambda_3}\; e^{-(\lambda_1+\lambda_2)x}\; (1-e^{-\lambda_3 x}) = \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2} g(x), \qquad x > 0.$$

with $g(x)$ representing the PDF of $WE(\frac{\lambda_3}{\lambda_1+\lambda_2}, \lambda_1+\lambda_2)$. Gupta and Kundu [6] showed that $g(x)$ is log-concave, and hence unimodal, with mode at

$$x_0 = \frac{1}{\lambda_3} \ln\Big(1 + \frac{\lambda_3}{\lambda_1+\lambda_2}\Big).$$

From the above analysis, we conclude that, for all $\lambda_1, \lambda_2, \lambda_3 > 0$, $f_{X,Y}(x,y)$ has a unique global maximum at the point $(x,y) = (x_0, x_0)$. This completes the proof.

Remark: The following table gives the value of x_0 for the 6 parts of Figure 1.

$(\lambda_1, \lambda_2, \lambda_3)$	(0.5,1,0.25) Fig. 1 (a)	(0.5,1,1) Fig. 1 (b)	(0.5,1,4) Fig. 1 (c)	(1,1,0.25) Fig. 1 (d)	(1,1,1) Fig. 1 (e)	(1,1,4) Fig. 1 (f)
x_0	0.617	0.511	0.325	0.471	0.405	0.275

4 Properties

In this section we discuss different properties of BWE. First we provide joint moment generating function (MGF) and it will be helpful to characterize the BWE.

Theorem 4.1: If $(X,Y) \sim \text{BWE}(\lambda_1, \lambda_2, \lambda_3)$, then the joint MGF of (X,Y) for $|t_1| < \lambda_1$ and $|t_2| < \lambda_2$ is

$$M_{X,Y}(t_1,t_2) = \left(1 - \frac{t_1}{\lambda_1}\right)^{-1} \left(1 - \frac{t_2}{\lambda_2}\right)^{-1} \left(1 - \frac{t_1+t_2}{\lambda}\right)^{-1}, \tag{11}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_3$.

Proof: Trivial.

From Theorem 4.1, we can easily obtain the joint cumulant generating function as

$$K_{X,Y}(t_1,t_2) = \ln M_{X,Y}(t_1,t_2) = -\left[\ln\left(1 - \frac{t_1}{\lambda_1}\right) + \ln\left(1 - \frac{t_2}{\lambda_2}\right) + \ln\left(1 - \frac{t_1+t_2}{\lambda}\right)\right]. \tag{12}$$

From the joint cumulant generating function we can easily obtain the means, variances and covariance of X and Y as follows;

$$\begin{aligned} E(X) &= \frac{\partial}{\partial t_1} K_{X,Y}(t_1,0)\bigg|_{t_1=0} = \frac{1}{\lambda_1} + \frac{1}{\lambda}, \qquad E(Y) = \frac{\partial}{\partial t_2} K_{X,Y}(0,t_2)\bigg|_{t_2=0} = \frac{1}{\lambda_2} + \frac{1}{\lambda}, \\ V(X) &= \frac{\partial^2}{\partial t_1^2} K_{X,Y}(t_1,0)\bigg|_{t_1=0} = \frac{1}{\lambda_1^2} + \frac{1}{\lambda^2} \qquad V(Y) = \frac{\partial^2}{\partial t_2^2} K_{X,Y}(0,t_2)\bigg|_{t_2=0} = \frac{1}{\lambda_2^2} + \frac{1}{\lambda^2}, \end{aligned}$$

and

$$Cov_{X,Y} = \frac{\partial^2}{\partial t_1 \partial t_2} K_{X,Y}(t_1,t_2)\bigg|_{t_1=0,t_2=0} = \frac{1}{\lambda^2}. \tag{13}$$

Therefore, the correlation coefficient of X and Y becomes

$$\rho_{X,Y} = \frac{\lambda_1 \lambda_2}{\sqrt{(\lambda^2 + \lambda_1^2)(\lambda^2 + \lambda_2^2)}}. \tag{14}$$

It is clear that $\rho_{X,Y}$ is always positive, and it is a decreasing function of λ_3. Moreover, if we take $\lambda_2 = c\lambda_1$, then it can be easily seen that $\rho_{X,Y} \to 1$ as $c \to \infty$ and $\lambda_3 \to 0$.

Alternative representation or characterization of BWE can be easily obtained from the joint MGF of X and Y. The following corollary provides the alternative representation BWE, which can be easily proved using the joint MGF. The alternative representation becomes very useful in deriving several properties of the BWE.

Corollary 4.1: Suppose $U_1 \sim \exp(\lambda_1)$, $U_2 \sim \exp(\lambda_2)$ and $V \sim \exp(\lambda)$ and they are independent. Let us define

$$X = U_1 + V, \quad \text{and} \quad Y = U_2 + V, \tag{15}$$

then $(X, Y) \sim \text{BWE}(\lambda_1, \lambda_2, \lambda_3)$.

From the Corollary 4.1, it is clear that V plays the role of dependence between X and Y. The representation (15) is very useful to generate BWE. Moreover, it can be used to establish different properties of BWE. For example, it can be used very easily to compute the product moments of X and Y. Using the fact that if $U \sim \exp(\theta)$, then $E(U^m) = \dfrac{\Gamma(m+1)}{\theta^m}$, it easily follows from the Binomial expansion that

$$\begin{aligned} E\left(X^m Y^n\right) &= E\left(U_1+V\right)^m \left(U_2+V\right)^n = \sum_{i=0}^{m}\sum_{j=0}^{n}\binom{m}{i}\binom{n}{j} E\left(U_1^i\right) E\left(U_2^j\right) E\left(V^{m+n-i-j}\right) \\ &= \sum_{i=0}^{m}\sum_{j=0}^{n}\binom{m}{i}\binom{n}{j} \frac{\Gamma(i+1)}{\lambda_1^i}\,\frac{\Gamma(j+1)}{\lambda_2^j}\,\frac{\Gamma(n+m-i-j+1)}{\lambda^{n+m-i-j}}. \end{aligned}$$

Here m and n are non-negative integers.

The stress-strength parameter $R = P(Y > X)$ also can be obtained using (15) as follows;

$$R = P(Y > X) = P(U_2 > U_1) = \frac{\lambda_2}{\lambda_1 + \lambda_2}. \tag{16}$$

We also have the following convolution properties of the n *i.i.d.* random variable from BWE. Suppose $(X_i, Y_i) \sim \text{BWE}(\lambda_1, \lambda_2, \lambda_3)$, then

$$(S, \;\; T) = (\sum_{i=1}^{n} X_i, \;\; \sum_{i=1}^{n} Y_i) \stackrel{d}{=} (P + R, \;\; Q + R), \tag{17}$$

here $\overset{d}{=}$ means equal in distribution, and $P \sim \text{gamma}(n, \lambda_1)$, $Q \sim \text{gamma}(n, \lambda_2)$, $R \sim \text{gamma}(n, \lambda_3)$ and they are independently distributed. For completeness purposes we provide the joint PDF of $(S,\ T)$ in the integral form as follows:

$$f_{S,T}(x,y) = \frac{(\lambda_1\lambda_2\lambda_3)^n}{(\Gamma(n))^3} e^{-\lambda_1 x} e^{-\lambda_2 y} \int_0^{\min\{x,y\}} (x-z)^{n-1}(y-z)^{n-1} z^{n-1} e^{-\lambda_3 z} dz. \tag{18}$$

Note that (18) can be written as a finite sum, but it has not attempted here. Now we provide the following total positivity result of BWE.

Theorem 4.2: If $(X,Y) \sim \text{BWE}(\lambda_1, \lambda_2, \lambda_3)$, then (X,Y) has total positivity of order two (TP_2) property.

Proof: Note that (X,Y) has the TP_2 property if and only if for any $t_{11}, t_{12}, t_{21}, t_{22}$, whenever $0 < t_{11} < t_{12}$ and $0 < t_{21} < t_{22}$, we have

$$f_{X,Y}(t_{11}, t_{21}) f_{X,Y}(t_{12}, t_{22}) - f_{X,Y}(t_{12}, t_{21}) f_{X,Y}(t_{11}, t_{22}) \geq 0. \tag{19}$$

Now let us consider different cases separately. For example if $t_{11} < t_{21} < t_{12} < t_{22}$, then the left hand side of (19) becomes

$$e^{-\lambda_3 t_{21}} - e^{-\lambda_3 t_{12}}. \tag{20}$$

Clearly, (20) is greater than or equal to zero as $t_{21} < t_{12}$. The other cases also can be proved along the same manner.

The following properties are very similar to the properties of the Block and Basu's bivariate exponential model, see proposition 5.1 of Block and Basu [3]. The proofs can be obtained using Corollary 4.1.

Theorem 4.3: Suppose $(X,Y) \sim \text{BWE}(\lambda_1, \lambda_2, \lambda_3)$, then
(a) $\min\{X,Y\} \sim \text{WE}\left(\frac{\lambda_3}{\lambda_1+\lambda_2}, \lambda_1 + \lambda_2\right)$.
(b) $X - Y$ has the cumulative distribution function

$$F_{X-Y}(u) = \begin{cases} \frac{\lambda_1}{\lambda_1+\lambda_2} e^{\lambda_2 u} & \text{if} \quad u \leq 0 \\ 1 - \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right) e^{-\lambda_1 u} & \text{if} \quad u > 0. \end{cases}$$

(c) $|X-Y|$ has the cumulative distribution function

$$G(u) = \frac{\lambda_1}{\lambda_1+\lambda_2}\left(1-e^{-\lambda_2 u}\right) + \frac{\lambda_2}{\lambda_1+\lambda_2}\left(1-e^{-\lambda_1 u}\right), \quad \text{if } u \geq 0.$$

(d) $\min\{X,Y\}$ is independent of $X-Y$ and also $|X-Y|$.

PROOF: (a) Using Corollary 4.1, we will be able to write

$$\min\{X,Y\} = \min\{U_1,U_2\} + V.$$

Since $\min\{U_1,U_2\}$ is also exponential, the result immediately follows.

(b) Note that $X-Y = U_1-U_2$. Now the result can be obtained easily directly, or it can be obtained from proposition 5.1 of Block and Basu [3] by substituting $\lambda_{12}=0$.

(c) Also can be obtained along the same line as (b).

(d) From proposition 5.1 of Block and Basu [3], it is known that $\min\{U_1,U_2\}$ is independent of U_1-U_2. Since V is also independent of U_1-U_2, both the results immediately follow.

5 Maximum Likelihood Estimation

In this section we discuss the maximum likelihood estimators (MLEs) of the unknown parameters. Based on the observations $(x_1,y_1),\ldots,(x_n,y_n)$, from BWE, the MLEs of the unknown parameters can be obtained by maximizing the log-likelihood function. The log-likelihood function of the observed data can be written as

$$l(\lambda_1,\lambda_2,\lambda_3) = n[\ln\lambda + \ln\lambda_1 + \ln\lambda_2 - \ln\lambda_3 - \lambda_1\bar{x} - \lambda_2\bar{y}] + \sum_{i=1}^{n}\ln\left(1-e^{-\lambda_3 z_i}\right), \tag{21}$$

here $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, and $z_i = \min\{x_i,y_i\}$ for $i=1,\ldots n$. Taking derivatives of (21) with respect to λ_1, λ_2 and λ_3, we obtain the following normal equations:

$$\frac{1}{\lambda} + \frac{1}{\lambda_1} - \bar{x} = 0 \tag{22}$$

$$\frac{1}{\lambda} + \frac{1}{\lambda_2} - \bar{y} = 0 \tag{23}$$

$$\frac{1}{\lambda} - \frac{1}{\lambda_3} + \frac{1}{n}\sum_{i=1}^{n} \frac{z_i e^{-\lambda_3 z_i}}{1 - e^{-\lambda_3 z_i}} = 0. \tag{24}$$

Instead of solving λ_1, λ_2 and λ_3 from (22), (23) and (24), first we solve λ_1 and λ_2 as a function of λ. For given λ, the solution for λ_1 and λ_2 become

$$\widehat{\lambda}_1(\lambda) = \frac{\lambda}{\lambda\bar{x} - 1} \quad \text{and} \quad \widehat{\lambda}_2(\lambda) = \frac{\lambda}{\lambda\bar{y} - 1}. \tag{25}$$

Substituting $\widehat{\lambda}_1(\lambda)$ and $\widehat{\lambda}_1(\lambda)$ in (24) we obtain the non-linear equation

$$\frac{1}{\lambda} - \frac{1}{\lambda\left(1 - \frac{1}{\lambda\bar{x}-1} - \frac{1}{\lambda\bar{y}-1}\right)} + \frac{1}{n}\sum_{i=1}^{n} \frac{z_i e^{-\lambda\left(1-\frac{1}{\lambda\bar{x}-1}-\frac{1}{\lambda\bar{y}-1}\right)z_i}}{1 - e^{-\lambda\left(1-\frac{1}{\lambda\bar{x}-1}-\frac{1}{\lambda\bar{y}-1}\right)z_i}} = 0. \tag{26}$$

Therefore, $\widehat{\lambda}$ can be obtained by solving (26), provided $\widehat{\lambda}_1(\widehat{\lambda}) > 0$, $\widehat{\lambda}_2(\widehat{\lambda}) > 0$ and $\widehat{\lambda}_3(\widehat{\lambda}) = \widehat{\lambda} - \widehat{\lambda}_1(\widehat{\lambda}) - \widehat{\lambda}_2(\widehat{\lambda}) > 0$. Alternatively, it can also be obtained by maximizing the profile log-likelihood function of λ, namely $l(\widehat{\lambda}_1(\lambda), \widehat{\lambda}_2(\lambda), \widehat{\lambda}_3(\lambda))$, directly as a function of λ, where the range of λ is such that $\widehat{\lambda}_1(\lambda) > 0$, $\widehat{\lambda}_2(\lambda) > 0$ and $\widehat{\lambda}_3(\lambda) = \lambda - \widehat{\lambda}_1(\lambda) - \widehat{\lambda}_2(\lambda) > 0$.

Now we discuss the asymptotic properties of the MLEs. It can be easily seen that the BWE satisfies all the required assumptions for the MLEs to be consistent and asymptotically normal. We have the following result.

Theorem 5.1: If $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\lambda}_3$ are the MLEs of λ_1, λ_2 and λ_3 respectively, then

$$\sqrt{n}\left(\widehat{\lambda}_1 - \lambda_1,\ \widehat{\lambda}_2 - \lambda_2,\ \widehat{\lambda}_3 - \lambda_3\right) \xrightarrow{d} N_3\left(0, I^{-1}\right).$$

Here $\xrightarrow{d}$ means convergence in distribution, $N_3\left(0, I^{-1}\right)$ is a 3-variate normal distribution with mean 0 and the dispersion matrix I^{-1}, where the elements of $I = (I_{ij})$ are as follows:

$$I_{11} = \frac{1}{\lambda^2} + \frac{1}{\lambda_1^2}, \quad I_{22} = \frac{1}{\lambda^2} + \frac{1}{\lambda_2^2}, \quad I_{33} = \frac{1}{\lambda^2} - \frac{1}{\lambda_3^2} + \frac{\lambda(\lambda_1+\lambda_2)}{\lambda_3}\int_0^\infty \frac{x^2 e^{-\lambda x}}{1 - e^{-\lambda_3 x}}dx,$$

$$I_{ij} = \frac{1}{\lambda^2} \quad \text{for} \quad i \neq j.$$

Proof: The proof follows using the standard arguments of a regular family. In calculating the expected Fisher information matrix, part (a) of Theorem 4.3 becomes useful.

6 Simulation and Data Analysis

In this section first we present some Monte Carlo simulation results to study the behavior of the MLEs and then present one data analysis results mainly for illustrative purposes.

6.1 Simulation Results

We have performed some small simulation to see the behavior of the MLEs. All the computations are performed with the FORTRAN code and using the random number generator RAN2 of Press *et al.* [7]. We have considered three different sets of model parameters; (1) $\lambda_1 = \lambda_2 = 1.0$, and $\lambda_3 = 0.5$, (2) $\lambda_1 = \lambda_2 = \lambda_3 = 1.0$ and (3) $\lambda_1 = \lambda_2 = 1.0$, and $\lambda_3 = 2.0$. In all the three cases we have used $n = 25$, 50, 75 and 100. We report the average estimates and square root of the mean squared errors based on 1000 replications. The results are reported in the following Table 1.

Some of the points are quite clear from the simulation results. As the sample size increases, the biases and the mean squared errors decrease, which verifies the consistency properties of MLEs. The behavior of $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ are very similar in nature both in terms of biases and mean squared errors, where as the behavior of $\widehat{\lambda}_3$ is quite different. Moreover, the behavior of $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ do not depend that much on the true value of λ_3, whereas the mean squared errors of $\widehat{\lambda}_3$ gradually increase as λ_3 increases.

6.2 Data Analysis

The data set represents the scores from twenty five first year graduate students in Probability-I and Inference-I of an premier Institute in India. For both the courses Analysis-I is a prerequisite. It is assumed that the knowledge of Analysis-I affects the scores in both the courses. For simplicity it is further assumed that it affects linearly. Therefore, according

Table 1: Average estimates (first row) and the square root of the mean squared errors (second row) are reported. Corresponds to each model, the first, second and third columns represent the results for λ_1, λ_2 and λ_3 respectively.

Model →	$\lambda_1 = \lambda_2 = 1$ & $\lambda_3 = 0.5$			$\lambda_1 = \lambda_2 = \lambda_3 = 1$			$\lambda_1 = \lambda_2 = 1$ & $\lambda_3 = 2.0$		
$n = 25$	1.0287	1.0321	0.7217	1.0581	1.0619	1.0237	1.0912	1.0956	1.6295
	0.2138	0.2197	1.0285	0.2337	0.2410	1.1339	0.2623	0.2712	1.4277
$n = 50$	1.0143	1.0172	0.6200	1.0329	1.0359	1.0085	1.0489	1.0523	1.7993
	0.1493	0.1552	0.8250	0.1632	0.1688	0.9788	0.1762	0.1820	1.2342
$n = 75$	1.0153	1.0118	0.5677	1.0291	1.0256	0.9962	1.0389	1.0356	1.8499
	0.1232	0.1241	0.6724	0.1341	0.1348	0.8690	0.1393	0.1410	1.0878
$n = 100$	1.0137	1.0084	0.5421	1.0229	1.0174	1.0064	1.0284	1.0230	1.9284
	0.1085	0.1116	0.5667	0.1166	0.1185	0.7787	0.1181	0.1212	0.9991

to Corollary 4.1, here V represents the score due to the knowledge of Analysis, and X (Y) represents the score purely due to the knowledge of Probability (Inference). The data set is presented below;

X : 53, 55, 85,87, 22, 23, 25,93, 51, 62, 53,32, 43, 47, 30, 88,59, 49, 42, 71,41, 82, 75, 93, 37

Y : 89, 90, 59,50, 25, 29, 54,62, 39, 25, 89,32, 33, 63, 38, 77,55, 41, 31, 66,57, 32, 43, 88, 34.

For the data, the sample means and sample covariance are, respectively, 55.92, 52.04 and 210.79. Based on the covariance and using (13), we obtain an initial estimate of $\widehat{\lambda}$ as $\dfrac{1}{\sqrt{210.79}} = 0.0689$. Using the above initial estimate, we obtain $\widehat{\lambda} = 0.0560$. Once we obtain $\widehat{\lambda}$, using (25), we obtain $\widehat{\lambda}_1 = 0.0263$, $\widehat{\lambda}_2 = 0.0293$. Finally by subtraction we also obtain $\widehat{\lambda}_3 = 0.0005$. We have also computed the 95% confidence intervals of λ_1, λ_2 and λ_3 using asymptotic distribution. The 95% confidence intervals of λ_1, λ_2 and λ_3 based on the asymptotic distribution are (0.0, 0.0733), (0.0, 0.0802), (0.0, 0.0129) respectively. Note that in all the three cases the lower limits are zero. Actually based on the asymptotic distributions, the lower limits were negative, but we replaced them by zero. Since the confidence lengths are quite large based on the asymptotic distributions, we compute 95% confidence intervals based on simple parametric bootstrapping with 1000 replications, and the confidence intervals for λ_1, λ_2 and λ_3 are (0.0014, 0.0671), (0.0018, 0.0718), (0.00001, 0.0015) respectively. They are significantly smaller than the corresponding confidence lengths based on asymptotic distributions.

Now we would like to see whether BWE model fits the data or not. It may be mentioned that although extensive work has been done on the goodness of fit tests for univariate continuous distribution, see for example D'Agostino and Stephens [4], not much work has been done on the multivariate continuous distributions except for the multivariate normal model, see for example Srivastava and Mudholkar [8]. Since we know that for the bivariate weighted exponential distribution, the marginals will be weighted exponential distributions, we are testing the marginals only. Although, it is not sufficient but at least it is necessary.

The Kolmogorov-Smirnov distances and the corresponding p-values (reported within brackets), between the empirical distribution function of the marginals and the estimated cumulative distribution functions based on the MLEs for X and Y are 0.208 (0.329) and 0.225 (0.286) respectively. Since the p-values are quite high for both the fitted marginals, we cannot reject the hypothesis that the marginals are weighted exponential distributions. It seems it is reasonable to use BWE model to analyze the above standardized data set.

7 Multivariate Generalization

In this section we discuss the multivariate generalization of the above model using the representation given in the Corollary 4.1. We define the p-variate distribution with weighted exponential marginals as follows. Some of the notations used in this section may be different than those used in the previous sections, but it should not create any ambiguity.

Suppose $\lambda_1, \ldots, \lambda_{p+1}$ are $p+1$ non-negative real numbers. Let $U_1, \ldots, U_p$ and V be independent exponential random variables with parameters $\lambda_1, \ldots, \lambda_p$ and λ_{p+1}, respectively.

Define

$$X_1 = U_1 + V, \ldots, X_p = U_p + V. \tag{27}$$

The p-variate random vector $(X_1, \ldots, X_p)$ is called the multivariate distribution with weighted exponential marginals and it will be denoted by MWE$(\lambda_1, \ldots, \lambda_{p+1})$. The joint PDF of $(X_1, \ldots, X_p)$ can be expressed as follows.

Theorem 7.1: The joint PDF of $(X_1, \ldots, X_p)$ is

$$f_{X_1,\ldots,X_p}(x_1, \cdots, x_p) = \frac{\lambda\lambda_1 \cdots \lambda_p}{\lambda_{p+1}} e^{-\lambda x_1} \cdots e^{-\lambda_p x_p} \left(1 - e^{-\lambda_{p+1} z}\right), \tag{28}$$

where $z = \min\{x_1, \cdots x_p\}$ and $\lambda = \lambda_1 + \cdots + \lambda_p$.

Proof: Using the transformation

$$\begin{bmatrix} X_1 \\ \vdots \\ X_p \\ V \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 0 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U_1 \\ \vdots \\ U_p \\ V \end{bmatrix},$$

the joint PDF of $X_1, \cdots, X_p, V$ can be obtained as

$$f_{X_1,\cdots,X_p,V}(x_1, \cdots, x_p, v) = \lambda_1 e^{-\lambda_1(x_1 - v)} \cdots \lambda_p e^{-\lambda_p(x_p - v)} \lambda e^{-\lambda v}, \tag{29}$$

for $0 < x_1, \cdots, 0 < x_p, 0 < v < z$. Now the result can be easily obtained by integrating out v from 0 to z.

It can be easily verified that the joint MGF of $X_1, \cdots X_p$ for $|t_1| < \lambda_1, \cdots, |t_p| < \lambda_p$ is

$$M_{X_1,\cdots X_p}(t_1, \cdots t_p) = \left(1 - \frac{t_1}{\lambda_1}\right)^{-1} \cdots \left(1 - \frac{t_p}{\lambda_p}\right)^{-1} \left(1 - \frac{\sum_{i=1}^{p} t_i}{\lambda}\right)^{-1}. \tag{30}$$

Therefore, it is immediately followed that if $q < p$, then $(X_1, \cdots X_q) \sim$ MWE$(\lambda_1, \cdots, \lambda_q, \lambda)$, where λ is same as defined before. Along the same line as the bivariate case the product moments of $X_1 \cdots X_p$ can be written as a finite summation as follows:

$$E\left(X_1^{n_1} \cdots X_p^{n_p}\right) = \sum_{i_1=0}^{n_1} \cdots \sum_{i_p=0}^{n_p} \binom{n_1}{i_1} \cdots \binom{n_p}{i_p} \frac{\Gamma(i_1+1)}{\lambda_1^{i_1}} \cdots \frac{\Gamma(i_p+1)}{\lambda_p^{i_p}} \frac{\Gamma(n - i_1 - \cdots i_p + 1)}{\lambda^{n - i_1 - \cdots - i_p}}.$$

It can be easily seen that

$$P(X_1 < \min\{X_2, \cdots, X_p\}) = \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_p}, \tag{31}$$

and

$$\min\{X_1, \cdots, X_p\} \sim \text{WE}\left(\frac{\lambda_{p+1}}{\lambda_1 + \cdots + \lambda_p}, \lambda_1 + \cdots + \lambda_p\right). \tag{32}$$

Now we would like to discuss the maximum likelihood estimators of the unknown parameters based on the sample $\{x_{i1}, \cdots, x_{ip}\}$ for $i = 1, \cdots, n$. We use the following notations: $\bar{x}_1 = \sum_{i=1}^{n} x_{i1}, \cdots, \bar{x}_p = \sum_{i=1}^{n} x_{ip}$, and $z_i = \min\{x_{i1}, \cdots, x_{ip}\}$, for $i = 1, \ldots, n$. Then, the log-likelihood function can be written as

$$l(\lambda_1, \cdots, \lambda_{p+1}) = n\left[\ln \lambda + \ln \lambda_1 + \cdots + \ln \lambda_p - \ln \lambda_{p+1} - \lambda_1 \bar{x}_1 - \cdots - \lambda_p \bar{x}_p\right] + \sum_{i=1}^{n} \ln\left(1 - e^{-\lambda_{p+1} z_i}\right). \tag{33}$$

It can be seen that the maximum likelihood estimate of λ can be obtained by solving the following non-linear equation;

$$\frac{1}{\lambda} - \frac{1}{g(\lambda)} + \frac{1}{n} \sum_{i=1}^{n} \frac{z_i e^{-g(\lambda)}}{1 - e^{-\lambda g(\lambda)}} = 0, \tag{34}$$

where

$$g(\lambda) = \lambda\left(1 - \widehat{\lambda}_1(\lambda) - \cdots - \widehat{\lambda}_p(\lambda)\right),$$

and

$$\widehat{\lambda}_1(\lambda) = \frac{\lambda}{\lambda \bar{x}_1 - 1}, \cdots, \widehat{\lambda}_p(\lambda) = \frac{\lambda}{\lambda \bar{x}_p - 1}. \tag{35}$$

Therefore, it is clear that the MLEs of the unknown parameters can be obtained by solving only one non-linear equation. Once $\widehat{\lambda}$ is obtained from (34) all the other estimators can be easily obtained from (35).

Now we provide the asymptotic distribution of the MLEs. We have the following result:

Theorem 7.2: If $\widehat{\lambda}_1, \cdots, \widehat{\lambda}_{p+1}$ are the MLEs of $\lambda_1, \cdots \lambda_{p+1}$ respectively, then

$$\sqrt{n}\left(\widehat{\lambda}_1 - \lambda_1, \cdots \quad \widehat{\lambda}_{p+1} - \lambda_{p+1}\right) \xrightarrow{d} N_{p+1}\left(0, I^{-1}\right).$$

Here $N_{p+1}\left(0, I^{-1}\right)$ is a (p+1)-variate normal distribution with mean 0 and the dispersion matrix I^{-1}, where the elements of $I = (I_{ij})$ are as follows:

$$I_{11} = \frac{1}{\lambda^2} + \frac{1}{\lambda_1^2}, \quad \cdots \quad I_{pp} = \frac{1}{\lambda^2} + \frac{1}{\lambda_p^2},$$

$$I_{(p+1)(p+1)} = \frac{1}{\lambda^2} - \frac{1}{\lambda_{p+1}^2} + \frac{\lambda(\lambda_1 + \cdots + \lambda_p)}{\lambda_{p+1}} \int_0^\infty \frac{x^2 e^{-\lambda x}}{1 - e^{-\lambda_{p+1} x}} dx,$$

$$I_{ij} = \frac{1}{\lambda^2} \quad \text{for} \quad i \neq j.$$

Proof: The proof follows using the standard arguments of a regular family. In calculating the expected Fisher information matrix, (32) becomes useful.

8 Conclusions

In this paper we have considered a new absolute continuous bivariate distribution which has three parameters and its marginals have weighted exponential distributions, recently proposed by Gupta and Kundu [6]. This new bivariate distribution has several interesting properties and it can be used as an alternative to the several three-parameter absolute continuous bivariate distribution, like Block and Basu bivariate distribution [3]. The generation of random samples from the proposed bivariate distribution is very simple, and therefore Monte Carlo simulations or the parametric bootstrapping can be performed very easily for different statistical inference purposes. It is observed that the maximum likelihood estimates of the unknown parameters can be obtained by solving one non-linear equation and Monte Carlo simulations indicate that the performance of the MLEs are quite satisfactory. Preliminary data analysis indicates that the performance of the confidence intervals based on asymptotic distribution is not very satisfactory, although the parametric bootstrapping seems to work quite well.

We have generalized the proposed bivariate model to the multivarite case and it is observed even the multivarite distribution also enjoys several interesting properties, similarly as the bivariate model. The most striking feature of the p variate distribution is, although it has $p+1$ parameters, but the MLEs of the unknown parameters can be obtained by solving only one non-linear equation. Therefore the proposed multivarite model can be easily used for analyzing higher dimensional data.

Acknowledgements

The authors would like to thank the referees for their helpful comments which improved significantly the earlier draft of the paper.

References

[1] Arnold, B. C. and Beaver, R. J. (2000), "Hidden truncation model", *Sankhya*, Ser. A, vol. 62, 23 - 35.

[2] Azzalini, A. (1985), "A class of distributions which includes the normal ones", *Scandinavian Journal of Statistics*, vol. 12, 171 - 185.

[3] Block, H. and Basu, A. P. (1974), "A continuous bivariate exponential extension", *Journal of the American Statistical Association*, vol. 69, 1031 - 1037.

[4] D'Agostino, R. B. and Stephens, M.A. (1986), *Goodness-of-fit techniques*, Marcel and Dekker, New York.

[5] Genton, M. G. (ed.) (2004), *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality*, Chapman and Hall/CRS, New York.

[6] Gupta, R. D. and Kundu, D. (2009), "A new class of weighted exponential distributions", *Statistics*, vol. 43, 621 - 643.

[7] Press, W.H., Teukolsky, S.A., Vellerling, W.T. and Flannery, B.P. (1992), *Numerical Recipes in FORTRAN, The Art of Scientific Computing*, 2nd. ed., Cambridge University, Cambridge, U.K.

[8] Srivastava, D. K. and Mudholkar, G.S. (2003), "Goodness-of-fit tests for univariate and multivariate normal models", *Handbook of Statistics*, vol. 22, North-Holland, Amsterdam, 869-956.