Bivariate Generalized Exponential Distribution

Debasis Kundu† and Rameshwar D. Gupta‡

Abstract

Recently it is observed that the generalized exponential distribution can be used quite effectively to analyze lifetime data in one dimension. The main aim of this paper is to define a bivariate generalized exponential distribution so that the marginals have generalized exponential distributions. It is observed that the joint probability density function, the joint cumulative distribution function and the joint survival distribution function can be expressed in compact forms. Several properties of this distribution have been discussed. We suggest to use the EM algorithm to compute the maximum likelihood estimators of the unknown parameters and also obtain the observed and expected Fisher information matrices. One data set has been re-analyzed and it is observed that the bivariate generalized exponential distribution provides a better fit than the bivariate exponential distribution.

Keywords: Joint probability density function; Conditional probability density function; Maximum likelihood estimators; Fisher information matrix; EM algorithm.

† Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, Pin 208016, INDIA. Phone no. 91-512-2597141, Fax No. 91-512-2597500, e-mail: kundu@iitk.ac.in. Corresponding author

† Department of Computer Science and Statistics, The University of New Brunswick at Saint John, New Brunswick, Canada E2L 4L5. Part of the work has been supported by a discovery grant from NSERC, CANADA.

1 Introduction

Gupta and Kundu (1999) introduced the generalized exponential (GE) distribution as a possible alternative to the well known gamma or Weibull distribution. The generalized exponential distribution has lots of interesting properties and it can be used quite effectively to analyze several skewed life time data. In many cases it is observed that it provides better fit than Weibull or gamma distributions. Since the distribution function of the GE is in closed form, it can be used quite easily for analyzing censored data also. The frequentest and Bayesian inferences have been developed for the unknown parameters of the GE distribution. The readers are referred to the review article of Gupta and Kundu (2007) for a current account on GE distribution.

Although quite a bit of work has been done in the recent years on GE distribution, but not much attempt has been made to extend this to the multivariate set up. Recently Sarhan and Balakrishnan (2007) has defined a new bivariate distribution using the GE distribution and exponential distribution and derived several interesting properties of this new distribution. Although they obtained the new bivariate distribution from the GE and exponential distributions, but the marginal distributions are not in known forms. In fact it is not known to the authors the existence of any bivariate distribution whose marginals are generalized exponential distributions.

The main aim of this paper is to provide a bivariate generalized exponential (BVGE) distribution so that the marginal distributions are GE distributions. In this connection, it may be mentioned here that Arnold (1967) provided some general techniques to construct multivarite distribution with specified marginals. We have adopted one of those techniques. The proposed BVGE distribution has three parameters but the scale and location parameters can be easily introduced. The joint cumulative distribution function (CDF), the joint probability density function (PDF) and the joint survival distribution function (SDF) are in closed forms, which make it convenient to use in practice.

The maximum likelihood estimators (MLEs) can be used to estimate the four unknown parameters when the scale parameter is also present. Although, the MLEs as expected can not be obtained in explicit forms, but the EM algorithm can be used quite effectively to obtain the MLEs. We also provide the observed and expected Fisher information matrices for practical users. Recently Meintanis (2007) analyzed one data and concluded that bivariate Marshal and Olkin (1967) exponential distribution provided a very good fit. We have re-analyzed the same data set and it is observed that the proposed BVGE distribution provides a much better fit than the Marshal and Olkin bivariate exponential model and provided some justification also.

The rest of the paper is organized as follows. In section 2, we define the BVGE distribution and discuss its different properties. The EM algorithm to compute the MLEs of the unknown parameters is provided in section 3. The analysis of a data set is provided in section 4. Finally we conclude the paper in section 5.

2 Bivariate Generalized Exponential Distribution

The univariate GE distribution has the following CDF and PDF respectively for $x > 0$;

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

Here $\alpha > 0$ and $\lambda > 0$ are the shape and scale parameters respectively. It is clear that for $\alpha = 1$, it coincides with the exponential distribution. From now on a GE distribution with the shape parameter α and the scale parameter λ will be denoted by GE(α,λ). For brevity when $\lambda = 1$, we will denote it by GE(α) and for $\alpha = 1$, it will be denoted by Exp(λ).

From now on unless otherwise mentioned, it is assumed that $\alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0, \lambda > 0$. Suppose $U_1 \sim \text{GE}(\alpha_1, \lambda)$, $U_1 \sim \text{GE}(\alpha_2, \lambda)$ and $U_3 \sim \text{GE}(\alpha_3, \lambda)$ and they are mutually independent. Here '$\sim$' means follows or has the distribution. Now define $X_1 = \max\{U_1, U_3\}$ and $X_2 = \max\{U_2, U_3\}$. Then we say that the bivariate vector (X_1, X_2) has a bivariate generalized exponential distribution with the shape parameters α_1, α_2 and α_3 and the scale parameter λ. We will denote it by $\text{BVGE}(\alpha_1, \alpha_2, \alpha_3, \lambda)$. Now for the rest of the discussions for brevity, we assume that $\lambda = 1$, although the results are true for general λ also. The BVGE distribution with $\lambda = 1$ will be denoted by $\text{BVGE}(\alpha_1, \alpha_2, \alpha_3)$. Before providing the joint CDF or PDF, we first mention how it may occur in practice.

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

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

The following results will provide the joint CDF, joint PDF and conditional PDF.

Theorem 2.1: If $(X_1, X_2) \sim \text{BVGE}(\alpha_1, \alpha_2, \alpha_3)$, then the joint CDF of (X_1, X_2) for $x_1 > 0$, $x_2 > 0$, is

$$F_{X_1,X_2}(x_1, x_2) = (1 - e^{-x_1})^{\alpha_1}(1 - e^{-x_2})^{\alpha_2}(1 - e^{-z})^{\alpha_3}, \tag{2}$$

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

PROOF: Trivial and therefore it is omitted. ∎

COROLLARY 2.1: The joint CDF of the BVGE$(\alpha_1, \alpha_2, \alpha_3)$ can also be written as

$$\begin{aligned} F_{X_1,X_2}(x_1,x_2) &= F_{GE}(x_1;\alpha_1)F_{GE}(x_2,\alpha_2)F_{GE}(z;\alpha_3) \\ &= F_{GE}(x_1;\alpha_1+\alpha_3)F_{GE}(x_2,\alpha_2) \quad \text{if} \quad x_1 < x_2 \\ &= F_{GE}(x_1;\alpha_1)F_{GE}(x_2,\alpha_2+\alpha_3) \quad \text{if} \quad x_2 < x_1 \\ &= F_{GE}(x;\alpha_1+\alpha_2+\alpha_3) \quad \text{if} \quad x_1 = x_2 = x. \end{aligned}$$

THEOREM 2.2: If $(X_1, X_2) \sim$ BVGE$(\alpha_1, \alpha_2, \alpha_3)$, then the joint PDF of (X_1, X_2) for $x_1 > 0$, $x_2 > 0$, is

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} f_1(x_1,x_2) & \text{if} \quad 0 < x_1 < x_2 < \infty \\ f_2(x_1,x_2) & \text{if} \quad 0 < x_2 < x_1 < \infty \\ f_0(x) & \text{if} \quad 0 < x_1 = x_2 = x < \infty, \end{cases}$$

where

$$\begin{aligned} f_1(x_1,x_2) &= f_{GE}(x_1;\alpha_1+\alpha_3)f_{GE}(x_2;\alpha_2) \\ &= (\alpha_1+\alpha_3)\alpha_2\left(1-e^{-x_1}\right)^{\alpha_1+\alpha_3-1}\left(1-e^{-x_2}\right)^{\alpha_2-1}e^{-x_1-x_2} \\ f_2(x_1,x_2) &= f_{GE}(x_1;\alpha_1)f_{GE}(x_2;\alpha_2+\alpha_3) \\ &= (\alpha_2+\alpha_3)\alpha_1\left(1-e^{-x_1}\right)^{\alpha_1-1}\left(1-e^{-x_2}\right)^{\alpha_2+\alpha_3-1}e^{-x_1-x_2} \\ f_0(x) &= \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}f_{GE}(x;\alpha_1+\alpha_2+\alpha_3) \\ &= \alpha_3\left(1-e^{-x}\right)^{\alpha_1+\alpha_2+\alpha_3-1}e^{-x}. \end{aligned}$$

PROOF: The expressions for $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$ can be obtained simply by taking $\dfrac{\partial^2 F_{X_1,X_2}(x_1,x_2)}{\partial x_1 \partial x_2}$ for $x_1 < x_2$ and $x_2 < x_1$ respectively. But $f_0(\cdot)$ can not be obtained in the same way. Using the facts that

$$\int_0^\infty \int_0^{x_2} f_1(x_1,x_2)dx_1dx_2 + \int_0^\infty \int_0^{x_1} f_2(x_1,x_2)dx_2dx_1 + \int_0^\infty f_0(x)dx = 1, \tag{3}$$

$$\int_0^\infty \int_0^{x_2} f_1(x_1,x_2)dx_1dx_2 = \alpha_2 \int_0^\infty \left(1-e^{-x}\right)^{\alpha_1+\alpha_2+\alpha_3-1} e^{-x}dx \tag{4}$$

and

$$\int_0^\infty \int_0^{x_1} f_2(x_1,x_2)dx_2dx_1 = \alpha_1 \int_0^\infty \left(1-e^{-x}\right)^{\alpha_1+\alpha_2+\alpha_3-1} e^{-x}dx \tag{5}$$

note that

$$\int_0^\infty f_0(x)dx = \alpha_3 \int_0^\infty \left(1-e^{-x}\right)^{\alpha_1+\alpha_2+\alpha_3-1} e^{-x}dx = \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}. \tag{6}$$

Therefore, the result follows. ∎

Comment 2.1: From Theorem 2.2 and Theorem 2.3, it easily follows that if we take $0 < \alpha_i < 1$, $i = 1,2,3$, and $\alpha_1 + \alpha_3 = \alpha_2 + \alpha_3 = 1$, then both X_1 and X_2 are exponentially distributed. Let, $\alpha_3 = \alpha$ and $\alpha_1 = 1 - \alpha$ and $\alpha_2 = 1 - \alpha$, then the joint PDF of (X_1, X_2) takes the form;

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} f_{GE}(x_1;1)f_{GE}(x_2;1-\alpha) & \text{if} \quad x_1 < x_2, \\ f_{GE}(x_1;1-\alpha)f_{GE}(x_2;1) & \text{if} \quad x_1 > x_2 \\ \frac{\alpha}{2-\alpha} f_{GE}(x;2-\alpha) & \text{if} \quad x_1 = x_2 = x. \end{cases} \tag{7}$$

Therefore the joint PDF as given in (7) has exponential marginals. ∎

The BVGE distribution has both an absolute continuous part and an singular part, similar to Marshall and Olkin's bivariate exponential model. The function $f_{X_1,X_2}(\cdot,\cdot)$ may be considered to be a density function for the BVGE distribution if it is understood that the first two terms are densities with respect to two-dimensional Lebesgue measure and the third term is a density function with respect to one dimensional Lebesgue measure, see for example Bemis, Bain and Higgins (1972). It is well known that although in one dimension the practical use of a distribution with this property is usually pathological, but they do arise quite naturally in higher dimension. In case of BVGE distribution, the presence of a singular part means that if X_1 and X_2 are BVGE distribution, then $X_1 = X_2$ has a positive probability. In many practical situations it may happen that X_1 and X_2 both are continuous random variables, but $X_1 = X_2$ has a positive probability, see Marshall and Olkin (1967) in this connection. The following result will provide explicitly the absolute continuous part and the singular part of the BVGE distribution function.

THEOREM 2.3: If $(X_1, X_2) \sim \text{BVGE}(\alpha_1, \alpha_2, \alpha_3)$, then

$$F_{X_1,X_2}(x_1, x_2) = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} F_a(x_1, x_2) + \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} F_s(x_1, x_2),$$

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

$$F_s(x_1, x_2) = \left(1 - e^{-z}\right)^{\alpha_1+\alpha_2+\alpha_3}$$

and

$$F_a(x_1, x_2) = \frac{\alpha_1 + \alpha_2 + \alpha_3}{\alpha_1 + \alpha_2} (1 - e^{-x_1})^{\alpha_1} (1 - e^{-x_2})^{\alpha_2} (1 - e^{-z})^{\alpha_3} - \frac{\alpha_3}{\alpha_1 + \alpha_2} (1 - e^{-z})^{\alpha_1+\alpha_2+\alpha_3},$$

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

PROOF: To find $F_a(x_1, x_2)$ from $F_{X_1,X_2}(x_1, x_2) = pF_a(x_1, x_2) + (1 - p)F_s(x_1, x_2)$, $0 \le p \le 1$, we compute

$$\frac{\partial^2 F_{X_1,X_2}(x_1, x_2)}{\partial x_1 \partial x_2} = p f_a(x_1, x_2) = \begin{cases} f_1(x_1, x_2) & \text{if} \quad x_1 < x_2, \\ f_2(x_1, x_2) & \text{if} \quad x_1 > x_2, \end{cases}$$

from which p may be obtained as

$$p = \int_0^\infty \int_0^{x_2} f_1(x_1, x_2) dx_1 dx_2 + \int_0^\infty \int_0^{x_1} f_2(x_1, x_2) dx_2 dx_1 = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3}$$

and

$$F_a(x_1, x_2) = \int_0^{x_1} \int_0^{x_2} f_a(u, v) du dv.$$

Once p and $F_a(\cdot,\cdot)$ are determined, $F_s(\cdot,\cdot)$ can be obtained by subtraction.

Alternatively, probabilistic arguments are also can be provided as follows. Suppose A is the following event: $A = \{U_1 < U_3\} \cap \{U_2 < U_3\}$, then $P(A) = \dfrac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}$.

Therefore,

$$F_{X_1,X_2}(x_1,x_2) = P(X_1 \le x_1, X_2 \le x_2|A)P(A) + P(X_1 \le x_1, X_2 \le x_2|A')P(A'). \tag{8}$$

Moreover for z as defined before,

$$P(X_1 \le x_1, X_2 \le x_2|A) = (1-e^{-z})^{\alpha_1+\alpha_2+\alpha_3}, \tag{9}$$

and $P(X_1 \le x_1, X_2 \le x_2|A')$ can be obtained by subtraction.

Clearly, $(1-e^{-z})^{\alpha_1+\alpha_2+\alpha_3}$ is the singular part as its mixed second partial derivative is zero when $x_1 \ne x_2$, and $P(X_1 \le x_1, X_2 \le x_2|A')$ is the absolute continuous part as its mixed partial derivative is a density function. ∎

Corollary 2.2: The joint PDF of X_1 and X_2 can be written as follows for $z = \min\{x_1,x_2\}$;

$$f_{X_1,X_2}(x_1,x_2) \quad = \quad \frac{\alpha_1+\alpha_2}{\alpha_1+\alpha_2+\alpha_3} f_a(x_1,x_2) + \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3} f_s(z),$$

where

$$f_a(x_1,x_2) = \frac{\alpha_1+\alpha_2+\alpha_3}{\alpha_1+\alpha_2} \times \begin{cases} f_{GE}(x_1;\alpha_1+\alpha_3) f_{GE}(x_2;\alpha_2) & \text{if} \quad x_1 < x_2 \\ f_{GE}(x_1;\alpha_1) f_{GE}(x_2;\alpha_2+\alpha_3) & \text{if} \quad x_1 > x_2 \end{cases}$$

and

$$f_s(x) = (\alpha_1+\alpha_2+\alpha_3)e^{-x}(1-e^{-x})^{\alpha_1+\alpha_2+\alpha_3-1} = f_{GE}(x;\alpha_1+\alpha_2+\alpha_3).$$

Clearly, here $f_a(x_1,x_2)$ and $f_s(z)$ are the absolute continuous part and singular part respectively. ∎

Comment 2.2: Using the joint PDF of X_1 and X_2, the different product moments $X_1^m X_2^n$ can be obtained in terms of infinite series similar to the one dimensional GE distribution, see Gupta and Kundu (1999).

From Theorem 2.3, it is clear that as $\alpha_3 \to 0$, $F_{X_1,X_2}(x_1,x_2) \to (1-e^{-x_1})^{\alpha_1}(1-e^{-x_2})^{\alpha_2}$, *i.e.*, X_1 and X_2 become independent. Moreover, since

$$A = (U_1 < U_3) \cap (U_2 < U_3) = \{\max\{U_1, U_2\} < U_3\} = \{X_1 = X_2\},$$

and $P(A) = \dfrac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}$, therefore, as $\alpha_3 \to \infty$, $P(A) = P(X_1 = X_2) \to 1$. It implies that for fixed α_1 and α_2, as α_3 varies from 0 to ∞, the correlation between X_1 and X_2 varies from 0 to 1.

The following theorem provides the marginal and the conditional results of the BVGE distribution.

THEOREM 2.4: If $(X_1, X_2) \sim \text{BVGE}(\alpha_1, \alpha_2, \alpha_3)$, then

(a) $X_1 \sim \text{GE}(\alpha_1+\alpha_3)$ and $X_2 \sim \text{GE}(\alpha_2+\alpha_3)$

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

$$F_{X_1|X_2=x_2}(x_1) = p_2 G_2(x_1) + (1-p_2) H_2(x_1),$$

where

$$G_2(x_1) = \frac{1}{p_2} \times \begin{cases} \frac{\alpha_2}{\alpha_2+\alpha_3}(1-e^{-x_2})^{-\alpha_3} \times (1-e^{-x_1})^{\alpha_1+\alpha_3} & \text{if} \quad x_1 < x_2 \\ (1-e^{-x_1})^{\alpha_1} - \frac{\alpha_3}{\alpha_2+\alpha_3}(1-e^{-x_2})^{\alpha_1} & \text{if} \quad x_1 > x_2, \end{cases}$$

$$H_2(x) = \begin{cases} 0 & \text{if} \quad x < x_2 \\ 1 & \text{if} \quad x \geq x_2 \end{cases}$$

and

$$p_2 = 1 - \frac{\alpha_3}{\alpha_2+\alpha_3}(1-e^{-x_2})^{\alpha_1}.$$

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

$$P(X_1 \leq x_1 | X_2 \leq x_2) = F_{X_1|X_2 \leq x_2}(x_1) = \begin{cases} (1-e^{-x_1})^{\alpha_1+\alpha_3}(1-e^{-x_2})^{-\alpha_3} & \text{if} \quad x_1 \leq x_2 \\ (1-e^{-x_1})^{\alpha_1} & \text{if} \quad x_1 > x_2 \end{cases}$$

PROOF: Trivial and therefore it is omitted. ∎

COMMENT 2.3: Since the joint survival function and the joint CDF have the following relation

$$S_{X_1,X_2}(x_1,x_2) = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{X_1,X_2}(x_1,x_2),$$

therefore, the joint survival function of X_1 and X_2 also can be expressed in a compact form.

COMMENT 2.4: Using Theorem 2.4, different moments of X_1, X_2, and conditional moments of $X_1|X_2 = x_2$ or $X_1|X_2 \leq x_2$ can be obtained in terms of infinite series.

An important property of the independent GE random variables X_1 and X_2 is that $\max\{X_1, X_2\}$ is also GE. If X_1 and X_2 are dependent but (X_1, X_2) is BVGE, then

$$P(\max\{X_1,X_2\} \leq x) = P(X_1 \leq x, X_2 \leq x) = P(U_1 \leq x, U_2 \leq x, U_3 \leq x) = (1-e^{-x})^{\alpha_1+\alpha_2+\alpha_3},$$

that is the maximum of X_1 and X_2 is also GE.

It is also interesting to observe that for all $0 < x_1, x_2 < \infty$,

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

Since

$$\bar{F}_{X_1,X_2}(x_1,x_2) - \bar{F}_{X_1}(x_1)\bar{F}_{X_2}(x_2) = F_{X_1,X_2}(x_1,x_2) - F_{X_1}(x_1)F_{X_2}(x_2),$$

therefore,

$$\bar{F}_{X_1,X_2}(x_1,x_2) \geq \bar{F}_{X_1}(x_1)\bar{F}_{X_2}(x_2).$$

Now we discuss the dependency properties of X_1 and X_2.

(i) Since $F_{X_1,X_2}(x_1,x_2) \geq F_{X_1}(x_1)F_{X_2}(x_2)$ for all x_1, x_2, therefore, they will be positive quadrant dependent, *i.e*, for every pair of increasing functions $h_1(\cdot)$ and $h_2(\cdot)$,

$$\text{Cov}(h_1(X_1), h_2(X_2)) \geq 0.$$

(ii) From part (b) of Theorem 2.4 it easily follows that for every x_1, $P(X_1 \leq x_1|X_2 = x_2)$ is a decreasing function of x_2, therefore X_2 is positive regression dependent of X_1. By symmetry it follows that X_1 is positive regression dependent of X_2.

(iii) From part (c) of Theorem 2.4 it easily follows that for every x_1, $P(X_1 \leq x_1|X_2 \leq x_2)$ is a decreasing function of x_2, therefore X_1 is left tail decreasing in X_2. By symmetry it follows that X_2 is left tail decreasing in X_1.

3 Maximum Likelihood Estimation

In this section we address the problem of computing the maximum likelihood estimators (MLEs) of the unknown parameters of BVGE distribution based on a random sample. The problem can be written as follows: Suppose $\{(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})\}$ is a random sample from BVGE$(\alpha_1, \alpha_2, \alpha_3, \lambda)$, the problem is to find the MLEs of the unknown parameters. We consider two cases separately, (a) α_3 is known, (b) α_3 is unknown. We use the following notation

$$I_1 = \{i; X_{1i} < X_{2i}\}, \quad I_2 = \{X_{1i} > X_{2i}\}, \quad I_0 = \{X_{1i} = X_{2i} = Y_i\}, \quad I = I_1 \cup I_2 \cup I_3,$$

$$|I_1| = n_1, \quad |I_2| = n_2, \quad |I_0| = n_0, \quad \text{and} \quad n_0 + n_1 + n_2 = n.$$

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

$$l(\alpha_1, \alpha_2, \alpha_3, \lambda) \quad = \quad n \ln \lambda + n_1 \ln(\alpha_1 + \alpha_3) + n_1 \ln \alpha_2 + (\alpha_1 + \alpha_3 - 1) \sum_{i \in I_1} \ln(1 - e^{-\lambda x_{1i}})$$

$$
\begin{aligned}
&+ \quad (\alpha_2 - 1) \sum_{i \in I_1} \ln(1 - e^{-\lambda x_{2i}}) + n_2 \ln \alpha_1 + n_2 \ln(\alpha_2 + \alpha_3) \\
&+ \quad (\alpha_1 - 1) \sum_{i \in I_2} \ln(1 - e^{-\lambda x_{1i}}) + (\alpha_2 + \alpha_3 - 1) \sum_{i \in I_2} \ln(1 - e^{-\lambda x_{2i}}) \\
&+ \quad n_0 \ln \alpha_3 + (\alpha_1 + \alpha_2 + \alpha_3 - 1) \sum_{i \in I_0} \ln(1 - e^{-\lambda y_i}) \\
&- \quad \lambda(\sum_{i \in I_0} y_i + \sum_{i \in I_1 \cup I_2} x_{1i} + \sum_{i \in I_2 \cup I_2} x_{2i}).
\end{aligned}
$$

CASE 1: α_3 is known.

In this case for fixed λ, the MLEs of α_1 and α_2, say $\widehat{\alpha}_1(\lambda)$ and $\widehat{\alpha}_2(\lambda)$ respectively, can be obtained as the solutions of the following equations;

$$
\frac{n_1}{(\alpha_1 + \alpha_3)} + \frac{n_2}{\alpha_1} = -\sum_{i \in I_0} \ln(1 - exp(-\lambda y_i)) - \sum_{i \in I_1 \cup I_2} \ln(1 - e^{-\lambda x_{1i}}) \tag{10}
$$

$$
\frac{n_2}{(\alpha_2 + \alpha_3)} + \frac{n_1}{\alpha_2} = -\sum_{i \in I_0} \ln(1 - exp(-\lambda y_i)) - \sum_{i \in I_1 \cup I_2} \ln(1 - e^{-\lambda x_{2i}}). \tag{11}
$$

It is not difficult to show that both the quadratic equations (10) and (11) have exactly one positive root each and they are

$$
\widehat{\alpha}_1(\lambda) \quad = \quad \frac{(-k_1\alpha_3 + n_1 + n_2) + \sqrt{(-k_1\alpha_3 + n_1 + n_2)^2 + 4k_1 n_2 \alpha_3}}{2k_1} \tag{12}
$$

$$
\widehat{\alpha}_2(\lambda) \quad = \quad \frac{(-k_2\alpha_3 + n_1 + n_2) + \sqrt{(-k_2\alpha_3 + n_1 + n_2)^2 + 4k_2 n_1 \alpha_3}}{2k_2}, \tag{13}
$$

where

$$
k_1 = -\left(\sum_{i \in I_0} \ln(1 - \exp(-\lambda y_i)) + \sum_{i \in I_1} \ln(1 - \exp(-\lambda x_{1i})) + \sum_{i \in I_2} \ln(1 - \exp(-\lambda x_{1i})) \right),
$$

$$
k_2 = -\left(\sum_{i \in I_0} \ln(1 - \exp(-\lambda y_i)) + \sum_{i \in I_1} \ln(1 - \exp(-\lambda x_{2i})) + \sum_{i \in I_2} \ln(1 - \exp(-\lambda x_{2i})) \right).
$$

Once $\widehat{\alpha}_1(\lambda)$ and $\widehat{\alpha}_2(\lambda)$ are obtained, the MLE of λ can be obtained by maximizing the profile log-likelihood of λ. It can be obtained as a solution of the following fixed point type equation;

$$
g(\lambda) = \lambda, \tag{14}
$$

where

$$
\begin{aligned}
g(\lambda) \;=\; & n\left[\sum_{i\in I_0} y_i + \sum_{i\in I_1\cup I_2} x_{1i} + \sum_{i\in I_1\cup I_2} x_{2i} - \widehat{\alpha}_1(\lambda)\sum_{i\in I_1}\frac{x_{1i}e^{-\lambda x_{1i}}}{(1-e^{-\lambda x_{1i}})}\right.\\
& -(\widehat{\alpha}_2(\lambda)-1)\sum_{i\in I_1}\frac{x_{2i}e^{-\lambda x_{2i}}}{(1-e^{-\lambda x_{2i}})} - (\widehat{\alpha}_1(\lambda)-1)\sum_{i\in I_2}\frac{x_{1i}e^{-\lambda x_{1i}}}{(1-e^{-\lambda x_{1i}})}\\
& \left.-\widehat{\alpha}_2(\lambda)\sum_{i\in I_2}\frac{x_{2i}e^{-\lambda x_{2i}}}{(1-e^{-\lambda x_{2i}})} - (\widehat{\alpha}_1(\lambda)+\widehat{\alpha}_2(\lambda))\sum_{i\in I_1}\frac{x_{1i}e^{-\lambda x_{1i}}}{(1-e^{-\lambda x_{1i}})}\right]^{-1}. \qquad (15)
\end{aligned}
$$

Simple iterative procedure as follows can be used to compute the MLEs. We start with the initial guess of λ as $\lambda^{(0)}$. Obtain $\widehat{\alpha}_1(\lambda_0)$ and $\widehat{\alpha}_2(\lambda_0)$ from (12) and (13). Compute $\lambda^{(1)} = g(\lambda^{(0)})$ using (15). Replace $\lambda^{(0)}$ by $\lambda^{(1)}$ and repeat the process. The process continues until $|\lambda^{(i)} - \lambda^{(i+1)}| < \epsilon$, where ϵ is some pre-assigned tolerance level.

Case 2: α_3 is also unknown.

In this case we suggest EM algorithm to compute the MLEs of the unknown parameters. We treat this as a missing value problem as follows. Assume that for a bivariate random vector (X_1, X_2), there is an associate random vector (Δ_1, Δ_2), $\Delta_1 = 1$ or 3, if $U_1 > U_3$ or $U_1 < U_3$ and similarly, $\Delta_2 = 2$ or 3, if $U_2 > U_3$ or $U_2 < U_3$ respectively. Therefore, if $X_1 = X_2$, then $\Delta_1 = \Delta_2 = 3$, but if $X_1 < X_2$ or $X_1 > X_2$, then (Δ_1, Δ_2) is missing. If $(x_1, x_2) \in I_1$, then the possible values of (Δ_1, Δ_2) are (1,2) and (3,2), similarly, if $(x_1, x_2) \in I_2$, then (Δ_1, Δ_2) can take (1,3) and (1,2) with non-zero probabilities.

Now we provide the 'E'-step and 'M'-step of the EM algorithm. In the 'E'-step, we treat the observations belong to I_0 as the complete observations. If the observation (x_1, x_2) belongs to either I_1 or I_2, we treat it as a missing observation. If $(x_1, x_2) \in I_1$, we form the 'pseudo observation' by fractioning (x_1, x_2) to two partially complete 'pseudo observation' of the form $(x_1, x_2, u_1(\gamma))$ and $(x_1, x_2, u_2(\gamma))$. Here $\gamma = (\alpha_1, \alpha_2, \alpha_3, \lambda)$, and the fractional mass $u_1(\gamma)$, $u_2(\gamma)$ assigned to the 'pseudo observation' (x_1, x_2) is the conditional probability that the random vector (Δ_1, Δ_2) takes the values (1,2) and (3,2) respectively, given that $X_1 < X_2$.

Similarly, if $(x_1, x_2) \in I_2$, we form the 'pseudo observation' of the form $(x_1, x_2, w_1(\gamma))$ and $(x_1, x_2, w_2(\gamma))$. Here the fractional mass $w_1(\gamma)$, $w_2(\gamma)$ assigned to the 'pseudo observation' (x_1, x_2) is the conditional probability that the random vector (Δ_1, Δ_2) takes the values (1,2) and (1,3) respectively, given that $X_1 > X_2$. Since

$$P(U_3 < U_1 < U_2) = \frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_3)(\alpha_1+\alpha_2+\alpha_3)}, \quad P(U_1 < U_3 < U_2) = \frac{\alpha_2\alpha_3}{(\alpha_1+\alpha_3)(\alpha_1+\alpha_2+\alpha_3)},$$

therefore,

$$u_1(\gamma) = \frac{\alpha_1}{\alpha_1+\alpha_3} \quad \text{and} \quad u_2(\gamma) = \frac{\alpha_3}{\alpha_1+\alpha_3}.$$

Similarly,

$$w_1(\gamma) = \frac{\alpha_2}{\alpha_2+\alpha_3} \quad \text{and} \quad w_2(\gamma) = \frac{\alpha_3}{\alpha_2+\alpha_3}.$$

From now on we write $u_1(\gamma)$, $u_2(\gamma)$, $w_1(\gamma)$, $w_2(\gamma)$ as u_1, u_2, w_1, w_2 respectively. The log-likelihood function of the 'pseudo data' can be written as

$$\begin{aligned}
l_{pseudo}(\alpha_1,\alpha_2,\alpha_3,\lambda) &= n_0 \ln\alpha_3 + n_0\ln\lambda + (\alpha_1+\alpha_2+\alpha_3-1)\sum_{i\in I_0}\ln(1-e^{-\lambda y_i}) - \lambda\sum_{i\in I_0} x_{1i} \\
&+ u_1\left[n_1\ln\alpha_1 + 2n_1\ln\lambda - \lambda\sum_{i\in I_1} x_{1i} + (\alpha_1+\alpha_3-1)\sum_{i\in I_1}\ln(1-e^{-\lambda x_{1i}})\right] \\
&+ u_2\left[n_1\ln\alpha_3 + 2n_1\ln\lambda - \lambda\sum_{i\in I_1} x_{1i} + (\alpha_1+\alpha_3-1)\sum_{i\in I_1}\ln(1-e^{-\lambda x_{1i}})\right] \\
&+ n_1\ln\alpha_2 - \lambda\sum_{i\in I_1} x_{2i} + (\alpha_2-1)\sum_{i\in I_1}\ln(1-e^{-\lambda x_{2i}}) \\
&+ w_1\left[n_2\ln\alpha_2 + 2n_2\ln\lambda - \lambda\sum_{i\in I_2} x_{2i} + (\alpha_2+\alpha_3-1)\sum_{i\in I_2}\ln(1-e^{-\lambda x_{2i}})\right] \\
&+ w_2\left[n_2\ln\alpha_3 + 2n_2\ln\lambda - \lambda\sum_{i\in I_2} x_{2i} + (\alpha_2+\alpha_3-1)\sum_{i\in I_2}\ln(1-e^{-\lambda x_{2i}})\right] \\
&+ n_2\ln\alpha_1 - \lambda\sum_{i\in I_2} x_{1i} + (\alpha_1-1)\sum_{i\in I_2}\ln(1-e^{-\lambda x_{1i}}) \\
&= n_0\ln\alpha_3 + \ln\lambda(n_0 + 2(n_1+n_2)) + (\alpha_1+\alpha_2+\alpha_3-1)\sum_{i\in I_0}\ln(1-e^{-\lambda x_{1i}}) \\
&- \lambda(\sum_{i\in I_0} y_i + \sum_{i\in I_1\cup I_2} x_{1i} + \sum_{i\in I_1\cup I_2} x_{2i}) + \ln\alpha_1(u_1n_1+n_2) + \ln\alpha_2(w_1n_2+n_1) \\
&+ (\alpha_1+\alpha_3-1)\sum_{i\in I_1}\ln(1-e^{-\lambda x_{1i}}) + (\alpha_2+\alpha_3-1)\sum_{i\in I_2}\ln(1-e^{-\lambda x_{2i}})
\end{aligned}$$

$$+ \quad (\alpha_2 - 1) \sum_{i \in I_1} \ln(1 - e^{-\lambda x_{2i}}) + (\alpha_1 - 1) \sum_{i \in I_2} \ln(1 - e^{-\lambda x_{1i}}).$$

Now the 'M' step involves the maximization of the $l_{pseudo}(\alpha_1, \alpha_2, \alpha_3, \lambda)$ with respect to α_1, α_2, α_3 and λ at each step. For fixed λ, the maximization of $l_{pseudo}(\alpha_1, \alpha_2, \alpha_3, \lambda)$ occurs at

$$\begin{aligned}
\widehat{\alpha}_1(\lambda) &= \frac{n_1 u_1 + u_2}{\sum_{i \in I_0} \ln(1 - e^{-\lambda y_i}) + \sum_{i \in I_1 \cup I_2} \ln(1 - e^{-\lambda x_{1i}})} \\
\widehat{\alpha}_2(\lambda) &= \frac{n_1 + w_1 n_2}{\sum_{i \in I_0} \ln(1 - e^{-\lambda y_i}) + \sum_{i \in I_1 \cup I_2} \ln(1 - e^{-\lambda x_{2i}})} \\
\widehat{\alpha}_3(\lambda) &= \frac{n_0 + n_1 u_2 + n_2 w_2}{\sum_{i \in I_0} \ln(1 - e^{-\lambda y_i}) + \sum_{i \in I_1} \ln(1 - e^{-\lambda x_{1i}}) + \sum_{i \in I_2} \ln(1 - e^{-\lambda x_{2i}})},
\end{aligned}$$

and $\widehat{\lambda}$, which maximizes $l_{pseudo}(\alpha_1, \alpha_2, \alpha_3, \lambda)$ can be obtained as a solution of the following fixed point equation;

$$g(\lambda) = \lambda, \tag{16}$$

where

$$\begin{aligned}
g(\lambda) = & \Bigg[\sum_{i \in I_0} y_i + \sum_{i \in I_1 \cup I_2} x_{1i} + \sum_{i \in I_1 \cup I_2} x_{2i} - (\widehat{\alpha}_1(\lambda) + \widehat{\alpha}_2(\lambda) + \widehat{\alpha}_3(\lambda) - 1) \sum_{i \in I_0} \frac{y_i e^{-\lambda y_i}}{(1 - e^{-\lambda y_i})} \\
& -(\widehat{\alpha}_1(\lambda) + \widehat{\alpha}_3(\lambda) - 1) \sum_{i \in I_1} \frac{x_{1i} e^{-\lambda x_{1i}}}{(1 - e^{-\lambda x_{1i}})} - (\widehat{\alpha}_2(\lambda) + \widehat{\alpha}_3(\lambda) - 1) \sum_{i \in I_2} \frac{x_{2i} e^{-\lambda x_{2i}}}{(1 - e^{-\lambda x_{2i}})} \\
& -(\widehat{\alpha}_2(\lambda) - 1) \sum_{i \in I_1} \frac{x_{2i} e^{-\lambda x_{2i}}}{(1 - e^{-\lambda x_{2i}})} - (\widehat{\alpha}_1(\lambda) - 1) \sum_{i \in I_2} \frac{x_{1i} e^{-\lambda x_{1i}}}{(1 - e^{-\lambda x_{1i}})} \Bigg] (n_0 + 2n_1 + 2n_2).
\end{aligned}$$

Now we describe how to compute $(i+1)$-th step from the i-th step in the EM algorithm.

- Step 1: Suppose at the i-th step the estimates of α_1, α_2, α_3 and λ are $\alpha_1^{(i)}$, $\alpha_2^{(i)}$, $\alpha_3^{(i)}$ and $\lambda^{(i)}$ respectively.

- Step 2: Compute u_1, u_2, w_1, w_2 using $\alpha_1^{(i)}$, $\alpha_2^{(i)}$, $\alpha_3^{(i)}$ and $\lambda^{(i)}$.

- Step 3: Find $\lambda^{(i+1)}$ by solving (16) similarly as (14).

- Step 4: Once $\lambda^{(i+1)}$ is obtained, compute $\alpha_1^{(i+1)} = \widehat{\alpha}_1(\lambda^{(i+1)})$, $\alpha_2^{(i+1)} = \widehat{\alpha}_2(\lambda^{(i+1)})$, $\alpha_3^{(i+1)} = \widehat{\alpha}_3(\lambda^{(i+1)})$.

4 Data Analysis

For illustrative purposes we have analyzed one data set to see how the proposed model and the EM algorithm works in practice. The data set has been obtained from Meintanis (2007) and it is presented in Table 1. The data represent the football (soccer) data where at least one goal scored by the home team and at least one goal scored directly from a penalty kick, foul kick or any other direct kick (all of them together will be called as *kick* goal) by any team have been considered. Here X_1 represents the time in minutes of the first *kick* goal scored by any team and X_2 represents the first goal of any type scored by the home team. In this case all possibilities are open, for example $X_1 < X_2$, or $X_1 > X_2$ or $X_1 = X_2 = Y$ (say). Meintanis (2007) used the Marshal-Olkin distribution to analyze the data. We would like to analyze the data using BVGE model.

Before going to analyze the data using BVGE model, we fit the GE distribution to X_1 and X_2 separately. The MLEs of the shape and scale parameters of the respective GE distribution for X_1 and X_2 are (3.121, 0.0449) and (1.678, 0.0413) respectively. The Kolmogorov-Smirnov distances between the fitted distribution and the empirical distribution function and the corresponding p values (in brackets) for X_1 and X_2 are 0.119 (0.667) and 0.121(0.654) respectively. Based on the p values GE distribution can not be rejected for the marginals.

First we try to fit the model under the assumption that U_3 is exponentially distributed, *i.e.* $\alpha_3 = 1$. In this case using the iterative algorithm (15), the MLEs of α_1, α_2 and λ are 1.385, 0.477 and 0.0373 respectively. Here, we started the iteration with the initial guess of

2005-2006	X_1	X_2	2004-2005	X1	X2
Lyon-Real Madrid	26	20	Internazionale-Bremen	34	34
Milan-Fenerbahce	63	18	Real Madrid-Roma	53	39
Chelsea-Anderlecht	19	19	Man. United-Fenerbahce	54	7
Club Brugge-Juventus	66	85	Bayern-Ajax	51	28
Fenerbahce-PSV	40	40	Moscow-PSG	76	64
Internazionale-Rangers	49	49	Barcelona-Shakhtar	64	15
Panathinaikos-Bremen	8	8	Leverkusen-Roma	26	48
Ajax-Arsenal	69	71	Arsenal-Panathinaikos	16	16
Man. United-Benfica	39	39	Dynamo Kyiv-Real Madrid	44	13
Real Madrid-Rosenborg	82	48	Man. United-Sparta	25	14
Villarreal-Benfica	72	72	Bayern-M. TelAviv	55	11
Juventus-Bayern	66	62	Bremen-Internazionale	49	49
Club Brugge-Rapid	25	9	Anderlecht-Valencia	24	24
Olympiacos-Lyon	41	3	Panathinaikos-PSV	44	30
Internazionale-Porto	16	75	Arsenal-Rosenborg	42	3
Schalke-PSV	18	18	Liverpool-Olympiacos	27	47
Barcelona-Bremen	22	14	M. Tel-Aviv-Juventus	28	28
Milan-Schalke	42	42	Bremen-Panathinaikos	2	2
Rapid-Juventus	36	52			

Table 1: UEFA Champion's League data

α as one and the iteration converges in 6 steps.

Now we fit the BVGE model under the assumptions that all the four parameters are unknown. Although we have some ideas about the values of $\alpha_1 + \alpha_3$ and $\alpha_2 + \alpha_3$, but we do not know about their individual values. We have an idea about the value of λ from the marginal λ's. For $\lambda = (0.0449 + 0.0413)/2 = 0.0431$, we plot the profile log-likelihood function of α_3 in Figure 1 and it is clear that the approximate value of α_3 should be close to one. Therefore, we get initial guesses of α_1 and α_2 also. We start the EM algorithm with the initial guesses of α_1, α_2, α_3 and λ as 2.0, 0.5, 1.0 and 0.04 respectively. The EM algorithm converges in 6-steps and MLEs of α_1, α_2, α_3 and λ are 1.445, 0.468, 1.170 and 0.0390. Since Max$\{X_1, X_2\}$ also follows GE$(\alpha_1 + \alpha_2 + \alpha_3)$, we can obtain the initial guesses as follows. We fit GE distributions to X_1, X_2 and to max$\{X_1, X_2\}$ and take the initial estimates of λ

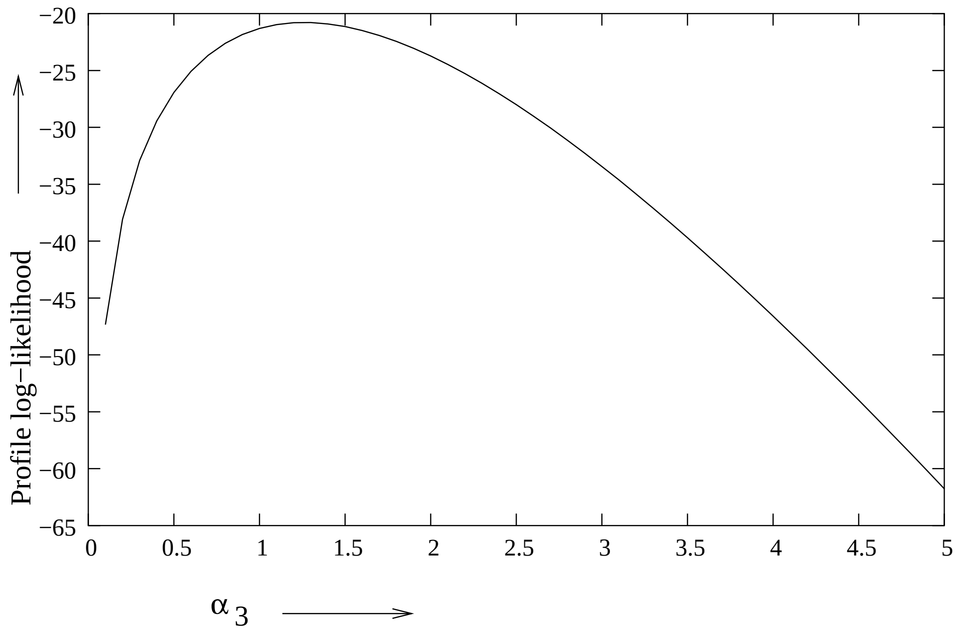

Figure 1: Profile log-likelihood function of α_3.

as the average of the three estimates. Once we get the estimates of λ we can obtain initial estimates of α_1, α_2 and α_3 from three linear equations. We obtain the initial estimates of λ in this case as 0.043 and using this value of λ we obtain the initial estimates of α_1, α_2 and α_3 as 2.55, 0.35 and 1.37 respectively. Using these initial values the EM algorithm converges to the same values after 11 iterations. We have computed the MLEs using direct maximization also (using grid search method) and we obtained the same estimates. Therefore, the EM algorithm works quite well in this case.

The corresponding 95% confidence intervals are obtained from the EM algorithm as suggested by Louis (1982) and they are as follows: (0.657, 2.233), (0.167, 0.769), (0.651, 1.689) and (0.028, 0.050) for α_1, α_2, α_3 and λ respectively. We have computed the K-S distance and the corresponding p values also (reported in brackets) between the fitted the GE(1.445+1.170=2.615, 0.0390) and GE(0.468+1.170=1.638, 0.0390) to the empirical distribution functions of X_1 and X_2 respectively. They are 0.103 (0.824) and 0.100 (0.852) respectively. Therefore, based on the marginals we can say that the BVGE distribution can be used quite effectively in this case.

Now we try to test whether BVGE or Marshal-Olkin (MO) provides better fit to the above data set. It may be mentioned that the MO model can not be obtained as a sub model from BVGE model. Therefore, the standard chi-square test can not be applied. We use the AIC and BIC to check the model validity. In case of BVGE, based on the above estimates the log-likelihood value is -20.59 and in case of MO model, using the estimates obtained by Meintanis (2007), the log-likelihood value becomes -44.57. The corresponding AIC (BIC) values are 49.18 (48.40) and 95.14 (94.56) respectively. Therefore, both the criteria suggest that BVGE provides a better fit than the MO model.

To see further why BVGE provides a better fit than the MO model, we look at the scaled TTT plot as suggested by Aarset (1987), which provides an idea of the shape of the hazard function of a distribution. For a family with the survival function $S(y) = 1 - F(y)$, the scaled TTT transform, with $H^{-1}(u) = \int_0^{F^{-1}(u)} S(y)dy$ defined for $0 < u < 1$ is $g(u) = H^{-1}(u)/H^{-1}(1)$. The corresponding empirical version of the scaled TTT transform is given by $g_n(r/n) = H_n^{-1}(r/n)/H_n^{-1}(1) = [(\sum_{i=1}^{r} y_{i:n}) + (n-r)y_{r:n}]/(\sum_{i=1}^{n} y_{i:n})$, where $r = 1, \ldots, n$ and $y_{i:n}, i = 1, \ldots, n$ represent the order statistics of the sample. It has been shown by Aarset (1987) that the scaled TTT transform is convex (concave) if the hazard rate is decreasing (increasing), and for bathtub (unimodal) hazard rates, the scaled TTT transform is first convex (concave) and then concave (convex). We plot the scaled TTT transform for X_1 and X_2 separately in Figure 2. From the Figure 2 it is quite apparent that both X_1 and X_2 have increasing hazard function and that also explains why BVGE, which may have increasing hazard functions for the marginals, provides better fit than MO model, which has only constant hazard functions for the marginals.

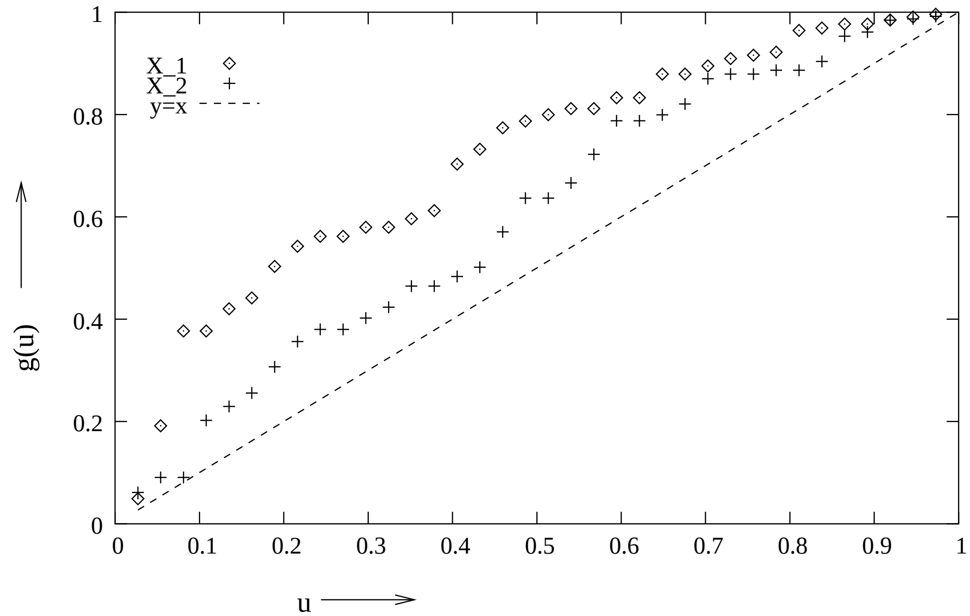

Figure 2: Scaled TTT transform for X_1 and X_2.

5 Conclusions

In this paper we have proposed bivariate generalized exponential distribution function whose marginals are generalized exponential distributions. It is observed that the BVGE distribution is a singular distribution and it has an absolute continuous part and a singular part. Since the joint distribution function and the joint density function are in closed forms, therefore this distribution can be used in practice for non-negative and positively correlated random variables. Although the maximum likelihood estimators of the unknown parameters can not be obtained in closed form but the EM algorithm works quite well and it can be effectively used to compute the MLEs. It may be mentioned that along the same line as Block and Basu (1974) bivariate exponential model, an absolute continuous version of the BVGE also can be obtained. Work is in progress in this direction and it will be reported else where.

Acknowledgements:

The authors would like to thank the two referees for several constructive suggestions.

Appendix

Expected Fisher Information Matrix

Let the Fisher Information matrix be;

$$I = E\begin{bmatrix} \frac{\partial^2 L}{\partial \alpha_1^2} & \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_2} & \frac{\partial^2 L}{\partial \alpha_1 \partial \alpha_3} & \frac{\partial^2 L}{\partial \alpha_1 \partial \lambda} \\ \frac{\partial^2 L}{\partial \alpha_2 \partial \alpha_1} & \frac{\partial^2 L}{\partial \alpha_2^2} & \frac{\partial^2 L}{\partial \alpha_2 \partial \alpha_3} & \frac{\partial^2 L}{\partial \alpha_2 \partial \lambda} \\ \frac{\partial^2 L}{\partial \alpha_3 \partial \alpha_1} & \frac{\partial^2 L}{\partial \alpha_3 \partial \alpha_2} & \frac{\partial^2 L}{\partial \alpha_3^2} & \frac{\partial^2 L}{\partial \alpha_3 \partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda \partial \alpha_1} & \frac{\partial^2 L}{\partial \lambda \partial \alpha_2} & \frac{\partial^2 L}{\partial \lambda \partial \alpha_3} & \frac{\partial^2 L}{\partial \lambda^2} \end{bmatrix} \tag{17}$$

Before providing the all the elements of the Fisher information matrix, we introduce the following notation. If $Z \sim \text{GE}(\alpha, \lambda)$, then

$$\begin{aligned} \xi(\alpha) &= E\left[\frac{Z^2 e^{-\lambda Z}}{(1-e^{-\lambda Z})^2}\right] \\ &= \frac{\alpha}{(\alpha-2)\lambda^2}\left[\psi'(1) - \psi'(\alpha-1) + (\psi(\alpha)-\psi(1))^2\right] \\ &\quad + \frac{\alpha}{(\alpha-1)\lambda^2}\left[\psi'(1) - \psi(\alpha) + (\psi(\alpha)-\psi(1))^2\right] \quad \text{if} \quad \alpha > 2 \\ &= \alpha\lambda \int_0^\infty z^2 e^{-2\lambda z}(1-e^{-\lambda z})^{\alpha-3} dz \quad \text{if} \quad 0 < \alpha \le 2, \end{aligned}$$

$$\begin{aligned} \eta(\alpha) &= E\left[\frac{Z e^{-\lambda Z}}{(1-e^{-\lambda Z})}\right] \\ &= \frac{1}{\lambda}\left[\frac{\alpha}{\alpha-1}(\psi(\alpha)-\psi(1)) - (\psi(\alpha+1)-\psi(1))\right] \quad \text{if} \quad \alpha > 2 \\ &= \alpha\lambda \int_0^\infty z e^{-2\lambda z}(1-e^{-\lambda z})^{\alpha-2} dz \quad \text{if} \quad 0 < \alpha \le 1, \end{aligned}$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma function and its derivative respectively, see Gupta and Kundu (1999) for details. Suppose $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$ is a random sample from

BVGE($\alpha_1, \alpha_2, \alpha_3, \lambda$) and n_0, n_1, n_2, I_0, I_1 and I_2 are same as defined in section 4. For brevity we further denote $\widetilde{\alpha} = \alpha_1 + \alpha_2 + \alpha_3$. We need the following results;

$$E(n_1) = nP(X_1 < X_2) = n\frac{\alpha_2}{\widetilde{\alpha}}, \quad E(n_2) = nP(X_2 < X_1) = n\frac{\alpha_1}{\widetilde{\alpha}}, \quad E(n_0) = nP(X_1 = X_2) = n\frac{\alpha_3}{\widetilde{\alpha}}.$$

Lemma A.1: Let $V_0 \sim \text{GE}(\widetilde{\alpha}, \lambda)$, $V_1 \sim \text{GE}(\alpha_1 + \alpha_3, \lambda)$ and $V_2 \sim \text{GE}(\alpha_2 + \alpha_3, \lambda)$ be three independent random variables and $g(\cdot)$ is a Borel measurable function, then

$$\begin{aligned}
E(g(X_i)|i \in I_1) &= E(g(V_1)) - \frac{\alpha_1 + \alpha_3}{\widetilde{\alpha}} E(g(V_0)) \\
E(g(X_i)|i \in I_2) &= \frac{\alpha_1}{\widetilde{\alpha}} E(g(V_0)) \\
E(g(X_i)|i \in I_0) &= \frac{\alpha_3}{\widetilde{\alpha}} E(g(V_0)) \\
E(g(Y_i)|i \in I_1) &= \frac{\alpha_2}{\widetilde{\alpha}} E(g(V_0)) \\
E(g(Y_i)|i \in I_2) &= E(g(V_2)) - \frac{\alpha_2 + \alpha_3}{\widetilde{\alpha}} E(g(V_0)).
\end{aligned}$$

Proof of Lemma A.1: Note that

$$\begin{aligned}
E(g(X_i)|i \in I_1) &= (\alpha_1 + \alpha_3)\alpha_2 \int_0^\infty \int_x^\infty g(x)(1 - e^{-\lambda x})^{\alpha_1 + \alpha_3 - 1}(1 - e^{-\lambda y})^{\alpha_2 - 1} e^{-\lambda x} e^{-\lambda y} dy dx \\
&= (\alpha_1 + \alpha_3) \int_0^\infty g(x)(1 - e^{-\lambda x})^{\alpha_1 + \alpha_3 - 1} e^{-\lambda x} \left[1 - (1 - e^{-\lambda x})^{\alpha_2}\right] dx \\
&= E(g(V_1)) - \frac{\alpha_1 + \alpha_3}{\widetilde{\alpha}} E(g(V_0)).
\end{aligned}$$

The others also can be obtained similarly. ∎

Now we obtain

$$\begin{aligned}
E\left[\frac{\partial^2 L}{\partial \alpha_1^2}\right] &= -E\left[\frac{n_1}{(\alpha_1 + \alpha_3)^2} + \frac{n_2}{\alpha_1^2}\right] = -\frac{n}{\widetilde{\alpha}}\left[\frac{\alpha_2}{(\alpha_1 + \alpha_3)^2} + \frac{1}{\alpha_1}\right] \\
E\left[\frac{\partial^2 L}{\partial \alpha_2^2}\right] &= -E\left[\frac{n_2}{(\alpha_2 + \alpha_3)^2} + \frac{n_1}{\alpha_2^2}\right] = -\frac{n}{\widetilde{\alpha}}\left[\frac{\alpha_1}{(\alpha_2 + \alpha_3)^2} + \frac{1}{\alpha_2}\right] \\
E\left[\frac{\partial^2 L}{\partial \alpha_3^2}\right] &= -E\left[\frac{n_1}{(\alpha_1 + \alpha_3)^2} + \frac{n_2}{(\alpha_2 + \alpha_3)^2} + \frac{n_0}{\alpha_3^2}\right] = -\frac{n}{\widetilde{\alpha}}\left[\frac{\alpha_2}{(\alpha_1 + \alpha_3)^2} + \frac{\alpha_1}{(\alpha_2 + \alpha_3)^2} + \frac{1}{\alpha_3}\right] \\
E\left[\frac{\partial^2 L}{\partial \lambda^2}\right] &= -E\left[\frac{1}{\lambda^2} + (\alpha_1 + \alpha_3 - 1)\sum_{i \in I_1} \frac{X_{1i}^2 e^{-\lambda X_{1i}}}{(1 - e^{-\lambda X_{1i}})^2} + (\alpha_2 - 1)\sum_{i \in I_1} \frac{X_{2i}^2 e^{-\lambda X_{2i}}}{(1 - e^{-\lambda X_{2i}})^2}\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad +(\alpha_1 - 1)\sum_{i\in I_2}\frac{X_{1i}^2 e^{-\lambda X_{1i}}}{(1-e^{-\lambda X_{1i}})^2} + (\alpha_2+\alpha_3-1)\sum_{i\in I_2}\frac{X_{2i}^2 e^{-\lambda X_{2i}}}{(1-e^{-\lambda X_{2i}})^2} \\
&\quad +(\widetilde{\alpha}-1)\sum_{i\in I_0}\frac{Y_i^2 e^{-\lambda Y_i}}{(1-e^{-\lambda Y_i})^2}\Bigg] \\
&= -n\Bigg[\frac{1}{\lambda^2} + \frac{\alpha_2(\alpha_1+\alpha_3-1)}{\widetilde{\alpha}}\left[\xi(\alpha_1+\alpha_3) - \frac{\alpha_1+\alpha_3}{\widetilde{\alpha}}\xi(\widetilde{\alpha})\right] + (\alpha_2-1)\left(\frac{\alpha_2}{\widetilde{\alpha}}\right)^2\xi(\widetilde{\alpha}) \\
&\quad +(\alpha_1-1)\left(\frac{\alpha_1}{\widetilde{\alpha}}\right)^2\xi(\widetilde{\alpha}) + \frac{\alpha_1(\alpha_2+\alpha_3-1)}{\widetilde{\alpha}}\left[\xi(\alpha_2+\alpha_3) - \frac{\alpha_2+\alpha_3}{\widetilde{\alpha}}\xi(\widetilde{\alpha})\right] \\
&\quad + (\widetilde{\alpha}-1)\left(\frac{\alpha_3}{\widetilde{\alpha}}\right)^2\xi(\widetilde{\alpha})\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial^2 L}{\partial\alpha_1\partial\lambda}\right] &= E\left[\sum_{i\in I_0}\frac{Y_i e^{-\lambda Y_i}}{(1-e^{-\lambda Y_i})}\right] + E\left[\sum_{i\in\cup I_1\cup I_2}\frac{X_{1i}e^{-\lambda X_{1i}}}{(1-e^{-\lambda X_{1i}})}\right] \\
&= n\left[\eta(\alpha_1+\alpha_3) - \frac{\alpha_1+\alpha_3}{\widetilde{\alpha}}\eta(\widetilde{\alpha}) + \frac{\alpha_1}{\widetilde{\alpha}}\eta(\widetilde{\alpha}) + \frac{\alpha_3}{\widetilde{\alpha}}\eta(\widetilde{\alpha})\right] = n\eta(\alpha_1+\alpha_3) \\
E\left[\frac{\partial^2 L}{\partial\alpha_2\partial\lambda}\right] &= E\left[\sum_{i\in I_1\cup I_2}\frac{X_{2i}e^{-\lambda X_{2i}}}{(1-e^{-\lambda X_{2i}})} + \sum_{i\in I_0}\frac{Y_i e^{-\lambda Y_i}}{(1-e^{-\lambda Y_i})}\right] \\
&= n\left[\frac{\alpha_2}{\widetilde{\alpha}}\eta(\widetilde{\alpha}) + \eta(\alpha_2+\alpha_3) - \frac{\alpha_2+\alpha_3}{\widetilde{\alpha}}\eta(\widetilde{\alpha}) + \frac{\alpha_3}{\widetilde{\alpha}}\eta(\widetilde{\alpha})\right] = n\eta(\alpha_2+\alpha_3) \\
E\left[\frac{\partial^2 L}{\partial\alpha_3\partial\lambda}\right] &= E\left[\sum_{i\in I_0}\frac{Y_i e^{-\lambda Y_i}}{(1-e^{-\lambda Y_i})} + \sum_{iI_1}\frac{X_{1i}e^{-\lambda X_{1i}}}{(1-e^{-\lambda X_{1i}})} + \sum_{i\in I_2}\frac{X_{2i}e^{-\lambda X_{2i}}}{(1-e^{-\lambda X_{2i}})}\right] \\
&= n\left[\eta(\alpha_2+\alpha_3) - \frac{\alpha_2+\alpha_3}{\widetilde{\alpha}}\eta(\widetilde{\alpha}) + \eta(\alpha_1+\alpha_3) - \frac{\alpha_1+\alpha_3}{\widetilde{\alpha}}\eta(\widetilde{\alpha}) + \frac{\alpha_3}{\widetilde{\alpha}}\eta(\widetilde{\alpha})\right] \\
&= n(\eta(\alpha_1+\alpha_3) + n\eta(\alpha_2+\alpha_3)) \\
E\left[\frac{\partial^2 L}{\partial\alpha_1\partial\alpha_2}\right] &= E\left[\frac{\partial^2 L}{\partial\alpha_1\partial\alpha_3}\right] = E\left[\frac{\partial^2 L}{\partial\alpha_2\partial\alpha_3}\right] = 0.
\end{aligned}
$$

Observed Fisher Information Matrix

For convenience we just present the observed Fisher information matrix obtained from the EM algorithm using the idea of Louis (1982). Using the same notation as Louis (1982), the observed Fisher information matrix can be written

$$F_{obs} = B - SS^T,$$

here B is the negative of the second derivative of the log-likelihood function and S is the derivative vector. We just provide the elements of the matrix B and the vector S. We use the following notation for brevity;

$$a_0 = \sum_{i \in I_0} \ln(1 - e^{-\widehat{\lambda} y_i}), a_{11} = \sum_{i \in I_1} \ln(1 - e^{-\widehat{\lambda} x_{1i}}), \;\; a_{12} = \sum_{i \in I_2} \ln(1 - e^{-\widehat{\lambda} x_{1i}}), \;\; a_{21} = \sum_{i \in I_1} \ln(1 - e^{-\widehat{\lambda} x_{2i}}),$$

$$a_{22} = \sum_{i \in I_2} \ln(1 - e^{-\widehat{\lambda} x_{2i}}), \quad b_0 = \sum_{i \in I_0} \frac{y_i e^{-\widehat{\lambda} y_i}}{1 - e^{-\widehat{\lambda} y_i}}, \quad b_{11} = \sum_{i \in I_1} \frac{x_{1i} e^{-\widehat{\lambda} x_{1i}}}{1 - e^{-\widehat{\lambda} x_{1i}}}, \quad b_{12} = \sum_{i \in I_2} \frac{x_{1i} e^{-\widehat{\lambda} x_{1i}}}{1 - e^{-\widehat{\lambda} x_{1i}}},$$

$$b_{21} = \sum_{i \in I_1} \frac{x_{2i} e^{-\widehat{\lambda} x_{2i}}}{1 - e^{-\widehat{\lambda} x_{2i}}}, \quad b_{22} = \sum_{i \in I_2} \frac{x_{2i} e^{-\widehat{\lambda} x_{2i}}}{1 - e^{-\widehat{\lambda} x_{2i}}}, \quad c_0 = \sum_{i \in I_0} \frac{y_i^2 e^{-\widehat{\lambda} y_i}}{1 - e^{-\widehat{\lambda} y_i}}, \quad c_{11} = \sum_{i \in I_1} \frac{x_{1i}^2 e^{-\widehat{\lambda} x_{1i}}}{1 - e^{-\widehat{\lambda} x_{1i}}},$$

$$c_{12} = \sum_{i \in I_2} \frac{x_{1i}^2 e^{-\widehat{\lambda} x_{1i}}}{1 - e^{-\widehat{\lambda} x_{1i}}}, \;\; c_{21} = \sum_{i \in I_1} \frac{x_{2i}^2 e^{-\widehat{\lambda} x_{2i}}}{1 - e^{-\widehat{\lambda} x_{2i}}}, \;\; c_{22} = \sum_{i \in I_2} \frac{x_{2i}^2 e^{-\widehat{\lambda} x_{2i}}}{1 - e^{-\widehat{\lambda} x_{2i}}}, \;\; d_0 = \sum_{i \in I_0} y_i, \;\; d_{11} = \sum_{i \in I_1} x_{1i},$$

$$d_{12} = \sum_{i \in I_2} x_{1i}, \quad d_{21} = \sum_{i \in I_1} x_{2i}, \quad d_{22} = \sum_{i \in I_2} x_{2i}.$$

Using the above notations we obtain;

$$S(1) = a_0 + \frac{n_1 u_1 + n_2}{\widehat{\alpha}_1} + a_{11} + a_{12}, \quad S(2) = a_0 + \frac{w_1 n_2 + n_1}{\widehat{\alpha}_2} + a_{21} + a_{22},$$

$$S(3) = \frac{1}{\widehat{\alpha}_3}(n_0 + n_1 u_1 + n_2 w_2) + a_0 + a_{11} + a_{22}$$

$$S(4) = \frac{1}{\widehat{\lambda}}(n_0 + 2n_1 + 2n_2) + b_0(\widehat{\alpha}_1 + \widehat{\alpha}_2 + \widehat{\alpha}_3) + (d_0 + d_{11} + d_{12} + d_{21} + d_{22}) + b_{11}(\widehat{\alpha}_1 + \widehat{\alpha}_3 - 1)) +$$

$$b_{22}(\widehat{\alpha}_1 + \widehat{\alpha}_3 - 1) + b_{21}(\widehat{\alpha}_1 - 1) + b_{12}(\widehat{\alpha}_2 - 1),$$

and

$$B(1,1) = \frac{(n_1 u_1 + n_2)}{\widehat{\alpha}_1^2}, \quad B(2,2) = \frac{(w_1 n_2 + n_1)}{\widehat{\alpha}_2^2}, \quad B(3,3) = \frac{(n_0 + n_1 u_1 + n_2 w_2)}{\widehat{\alpha}_3^2},$$

$$B(4,4) = \frac{1}{\widehat{\lambda}^2} + c_0(\widehat{\alpha}_1 + \widehat{\alpha}_2 + \widehat{\alpha}_3 - 1) + c_{11}(\widehat{\alpha}_1 + \widehat{\alpha}_3 - 1) + c_{22}(\widehat{\alpha}_2 + \widehat{\alpha}_3 - 1) + c_{21}(\widehat{\alpha}_1 - 1) + c_{12}(\widehat{\alpha}_2 - 1)$$

$$B(1,4) = B(4,1) = b_0 + b_{11} + b_{21}, \quad B(2,4) = B(4,2) = b_0 + b_{12} + b_{22},$$

$$B(3,4) = B(4,3) = b_0 + b_{11} + b_{22}, B(1,2) = B(2,1) = B(1,3) = b(3,1) = B(2,3) = B(3,2) = 0.$$

References

[1] Aarset, M.V. (1987), “How to identify a bathtub hazard rate?”, *IEEE Transactions on Reliability*, vol. 36, 106 -108.

[2] Arnold, B. (1967), "A note on multivariate distributions with specified marginals", *Journal of the American Statistical Association*, vol. 62 1460–1461.

[3] 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.

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

[5] Gupta, R. D. and Kundu, D. (1999), "Generalized exponential distributions", *Australian and New Zealand Journal of Statistics*, vol. 41, 173 - 188.

[6] Gupta, R. D. and Kundu, D. (2007), "Generalized exponential distributions: existing results and some recent developments", *Journal of Statistical Planning and Inference*, vol. 137, 3525 - 3536.

[7] Louis, T. A. (1982), "Finding the observed information matrix when using the EM algorithm", *Journal of the Royal Statistical Society*, Series B 44, 2, 226233.

[8] Marshall, A.W. and Olkin, I. (1967), "A multivariate exponential distribution", *Journal of the American Statistical Association*, vol. 62, 30 - 44.

[9] Meintanis, S.G. (2007), "Test of fit for Marshall-Olkin distributions with applications", *Journal of Statistical Planning and inference*, vol. 137, 3954-3963.

[10] Sarhan, A. and Balakrishnan, N. (2007), "A new class of bivariate distribution and its mixture", *Journal of the Multivariate Analysis*, vol. 98, 1508 - 1527.