A Bivariate Inverse Generalized Exponential Distribution and its Applications in Dependent Competing Risks Model

Fatemah A. Alqallaf [*] & Debasis Kundu [†]

Abstract

The aim of this paper is to ~~introduce a~~ consider the bivariate inverse generalized exponential distribution which has a singular component. The ~~proposed~~ bivariate inverse generalized exponential distribution can be used when the marginals have heavy tailed distributions, and they have non-monotone hazard functions. Due to presence of the singular component, it can be used quite effectively when there are ties in the data. Since it has four parameters, it is a very flexible bivariate distribution and it can be used quite effectively for analyzing various bivariate data sets. Several dependency properties and dependency measures have been obtained. The maximum likelihood estimators cannot be obtained in closed form, and it involves solving a four dimensional optimization problem. To avoid that we have proposed to use an EM algorithm and it involves solving only one non-linear equation at each 'E'-step. Hence, the implementation of the proposed EM algorithm is very straight forward in practice. Extensive simulation experiments and the analysis of one data set have been performed. We have observed that the ~~proposed~~ bivariate inverse generalized exponential distribution can be used for modeling dependent competing risks data. One data set has been analyzed to show the effectiveness of the ~~proposed~~ model.

Key Words and Phrases: Marshall-Olkin bivariate exponential distribution; Block and Basu bivariate distributions;maximum likelihood estimators; EM algorithm; competing risks.

AMS Subject Classifications: 62F10, 62F03, 62H12.

[*]Department of Statistics and Operation Research, Faculty of Science, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait

[†]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India. E-mail: kundu@iitk.ac.in, Phone no. 91-512-2597141, Fax no. 91-512-2597500.

1 Introduction

Generalized exponential (GE) distribution has received a considerable amount of attention since it has been introduced by Gupta and Kundu [14]. It is a very flexible two-parameter distribution and it has several properties which are very similar to that of the two-parameter gamma and Weibull distributions. All these three distributions are generalization of the exponential distribution but in different ways. They can have decreasing and unimodal probability density function (PDF)s and monotone hazard function (HF)s. Extensive work has been done in developing several properties and providing various inference procedures of the unknown parameters of a GE distribution. See for example the review article by Nadarajah [27] and the book length treatment by Al-Hussaini and Ahsanullah [1].

Recently, inverse GE (IGE) distribution has been introduced in the literature similar to the inverse Weibull distribution (IWE), see for example Oguntunde and Adejumo [29]. It has some interesting features which are very similar to the IWE distribution. It has an unimodal PDF, it always has a non-monotone hazard function and it can be heavy tailed also. Therefore, if the data indicate that they may be coming from a heavy tailed distribution then IGE or IWE can be used for analyzing the data set. Since both the distribution functions have explicit cumulative distribution functions (CDFs), both of them can be used quite effectively for analyzing censored data also.

It may be mentioned that the Marshall and Olkin [25] bivariate exponential (MOBE) distribution is one of the most popular bivariate exponential distributions with singular components. It has been used when there are ties in the data. It has received a significant attention in the statistical literature due to its simplicity and analytical tractability, see for example Arnold [2], Baxter and Rachev [4], Begum and Khan [5], Bemis et al. [6], Pena and Gupta [30], Boland [7] etc. But one major drawback of the MOBE distribution is that it has exponential marginals, and it may not be very useful if the marginals have non-monotone hazard functions. Due to this limitation, several other bivariate distributions have been introduced in the literature which have singular components and which have more flexible marginal distributions, for example the Marshall-Olkin bivariate Weibull (MOBW) distribution, see Lu [24], Kundu and Dey [16], Kundu and Gupta [19], Lai et al. [22], bivariate Kumaraswamy distribution by Barreto-Souza and Lemonte [3], generalized Marshall-Olkin bivariate distribution by Gupta et al. [13], bivariate generalized exponential distribution by Kundu and Gupta [20], Sarhan-Balakrishnan distribution by Sarhan and Balakrishnan [33], bivariate generalized linear failure rate distribution by Sarhan et al. [34], bivariate inverse Weibull (BIWE) distribution by Kundu and Gupta [21]. Although, several distributions have been introduced in the literature, other than the BIWE distribution none of the other distribution can have heavy tailed marginals.

The aim of this paper is two fold. First we ~~introduce~~ consider a bivariate IGE (BIGE) distribution whose marginals are IGE distributions. The joint cumulative distribution function (JCDF) has a singular component similar to the MOBE distribution. Therefore, if the preliminary data analysis indicates that the marginals may have heavy tailed distributions and there are ties in the bivariate data set, then the ~~proposed~~ BIGE distribution can be used for analyzing this data set. The BIGE distribution has four parameters, and the joint PDF can take variety of shapes. We derive different properties of the BIGE distributions. It is observed that it can be obtained from the Marshall-Olkin copula. As it has been mentioned before that the above construction is not new. Recently, Popović et al. [31] proposed a class of multivariate distributions with proportional hazard rate marginals in the same manner. The proposed BIGE distribution can be obtained as a special case. Popović et al. [31] provided several dependency properties of this general class of distributions, hence, those properties hold for the BIGE distribution also. It may be mentioned that the aim of the paper of Popović et al. [31] is quite different than the present one. Popović et al. [31] mainly provided different dependency properties for the general class of distributions, but in this present paper our aim is to provide an efficient estimation procedure for BIGE distribution and use this model for analyzing dependent competing risks data.

The ~~proposed~~ BIGE distribution has four parameters. The maximum likelihood estimators of the unknown parameters cannot be obtained in closed form. They have to be obtained by solving four non-linear equations simultaneously. Some numerical procedures like Newton-Raphson method or some of its variants may be used to solve these non-linear equations. But it needs very careful choice of the initial guesses, otherwise it may converge to a local maximum rather than a global maximum. To avoid that we have proposed to treat the maximization problem as a missing value problem and obtain the maximum likelihood estimates by using EM algorithm. It is observed that the proposed EM algorithm requires solving only one one-dimensional optimization problem at each step of the EM algorithm. Moreover, it does not need any integration calculation also like some of the existing EM algorithms available in the literature. Hence, the implementation of the proposed EM algorithm is quite straight forward. Moreover, the confidence intervals of the unknown parameters can be obtained based on the observed Fisher information matrix as it has been suggested by Louis [23] in a quite straight forward manner.

In reliability or in survival analysis, very often the experimental units are exposed to more than one cause of failures. In this case one observes the failure time as well as the cause of failure. The analysis of the lifetime distribution due to a particular cause in presence of other causes is an important problem and it is known as the competing risks problem in the statistical literature. Extensive work has been done in developing different competing risks models and providing inferential procedures. One may refer to the book by Crowder [10] in this respect. Several models are available to analyze competing risks data. One of the most popular ones is the latent failure model proposed by Cox [9]. In this case the failure time T can be written as $T = \min\{T_1, \ldots, T_m\}$, where T_i denotes the latent failure time due to i-th risk factor for $i = 1, \ldots, m$. Most of the analyses of the competing risks models have been performed based on the assumption that T_i's are independent, which may not be very reasonable. Recently there is a growing interest to develop different dependent competing risks models, see for example Feizjavadian and Hashemi [12], Cai et al. [8], Shen and Xu [35] and see the references cited therein.

The second aim of this paper is to develop a dependent competing risks model based on the BIGE distribution. We propose to use the BIGE distribution as a dependent competing risks model when latent failure distributions may be heavy tailed and it does not have monotone hazard functions, which has not been addressed in the literature. It may be mentioned MOBW distribution has been used to analyze dependent competing risks data. But it cannot be used if the data are of heavy tailed. Although, BIWE has heavy tailed marginals, it cannot be used very conveniently for analyzing competing risks data. This is the main motivation to use ~~the proposed~~ this model for analyzing dependent competing risks data. In this case also it is observed that the MLEs of the unknown parameters cannot be obtained in closed form and it involves solving a four dimensional optimization problem. But using the profile likelihood method it is observed that the MLEs can be obtained by solving only one one-dimensional optimization problem. Therefore, the MLEs can be obtained very conveniently. We have analyzed one competing risk data and the performances are quite satisfactory.

The rest of the paper is organized as follows. In Section 2, we have developed the BIGE distribution and develop its different properties. The maximum likelihood estimators are provided in Section 3. In Section 4 we have present simulation results and the analysis of a bivariate data set. The application of the BIGE to the dependent competing risks model has been provided in Section 5. Finally we conclude the paper in Section 6.

2 BIGE Distribution

2.1 Preliminaries

A random variable X is said to follow a GE distribution if the CDF of X is of the form:

$$F_{GE}(x;\alpha,\lambda) = \begin{cases} (1-e^{-\lambda x})^{\alpha} & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0. \end{cases}$$

Here, $\alpha > 0$ and $\lambda > 0$ are the shape and scale parameters, respectively. From now on it will be denoted by GE(α,λ). The corresponding PDF becomes

$$f_{GE}(x;\alpha,\lambda) = \begin{cases} \alpha\lambda e^{-\lambda x}(1-e^{-\lambda x})^{\alpha-1} & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0. \end{cases}$$

The PDF of GE(α,λ) is a decreasing function in x when $0 < \alpha \leq 1$ and it is an unimodal function when $1 < \alpha < \infty$, for all $\lambda > 0$. Similarly, the hazard function of GE(α,λ) is a decreasing function if $0 < \alpha \leq 1$ and it is an increasing function if $\alpha > 1$, for all possible values of λ. It has been observed by several authors that the GE distribution behaves very similar to the gamma and Weibull distributions.

A random variable X is said to follow a IGE distribution if the survival function (SF) of X has the following form:

$$S_{IGE}(x;\alpha,\lambda) = \begin{cases} (1-e^{-\frac{\lambda}{x}})^{\alpha} & \text{if} \quad x > 0 \\ 1 & \text{if} \quad x \leq 0. \end{cases}$$

Here also $\alpha > 0$ and $\lambda > 0$ are the shape and scale parameters, respectively. We will denote this by IGE(α,λ). The corresponding PDF becomes

$$f_{IGE}(x;\alpha,\lambda) = \begin{cases} \frac{\alpha\lambda}{x^2} e^{-\frac{\lambda}{x}}(1-e^{-\frac{\lambda}{x}})^{\alpha-1} & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0. \end{cases}$$

The PDF of IGE(α,λ) is an unimodal function of x for all $\alpha > 0$ and $\lambda > 0$.

The hazard function of IGE(α,λ) is as follows:

$$h_{IGE}(x;\alpha,\lambda) = \frac{f_{IGE}(x;\alpha,\lambda)}{S_{IGE}(x;\alpha,\lambda)} = \frac{\alpha\lambda}{x^2(e^{\frac{\lambda}{x}}-1)}. \tag{1}$$

It is clear that the shape of the hazard function does not depend on α and λ. Hence, for $\alpha = 1$ and $\lambda = 1$,

$$\ln h_{IGE}(x; 1, 1) = -2 \ln x - \ln(e^{\frac{1}{x}} - 1),$$

and

$$\frac{d}{dx} \ln h_{IGE}(x; 1, 1) = -\frac{1}{x} + \frac{1}{x^2(e^{\frac{1}{x}} - 1)} = 0$$

has a unique solution. Therefore, the hazard function of an IGE distribution is an unimodal function for all possible values of α and λ. It behaves very similarly to the inverse Weibull distribution. The PDFs and hazard functions of an IGE distribution for different parameters values are provided in Figure 1 and Figure 2, respectively. Now we will define BIGE distribution.

2.2 BIGE Distribution

Suppose U follows ($\sim$) IGE(α_1, λ), $V \sim$ GE(α_2, λ and $W \sim$ GE(α_3, λ), and they are independently distributed. If $X = \min\{U, W\}$ and $Y = \min\{V, W\}$, then (X, Y) is said to have bivariate IGE (BIGE) distribution with parameters α_1, α_2, α_3 and λ. We will denote this by BIGE($\alpha_1, \alpha_2, \alpha_3, \lambda$). It may be mentioned this BIGE distribution has the same interpretation as the shock model similar to the MOBE model. Here the shock appears following an IGE distribution, which can be heavy tailed.

The following results provide the joint SF and the joint PDF of BIGE($\alpha_1, \alpha_2, \alpha_3, \lambda$).

Theorem 2.1 Suppose $(X, Y) \sim$ BIGE($\alpha_1, \alpha_2, \alpha_3, \lambda$), then the joint SF of X and Y for $x > 0$ and $y > 0$ becomes

$$S_{X,Y}(x, y) = P(X > x, Y > y) = \begin{cases} (1 - e^{-\frac{\lambda}{x}})^{\alpha_1}(1 - e^{-\frac{\lambda}{y}})^{\alpha_2+\alpha_3} & \text{if} \quad 0 < x < y < \infty \\ (1 - e^{-\frac{\lambda}{x}})^{\alpha_1+\alpha_3}(1 - e^{-\frac{\lambda}{y}})^{\alpha_2} & \text{if} \quad 0 < y < x < \infty \\ (1 - e^{-\frac{\lambda}{x}})^{\alpha_1+\alpha_2+\alpha_3} & \text{if} \quad 0 < x = y < \infty \end{cases}$$

Proof: It is trivial and it is omitted.

We have the following unique decomposition of the joint SF of X and Y.

THEOREM 2.2 Suppose $(X, Y) \sim \text{BIGE}(\alpha_1, \alpha_2, \alpha_3, \lambda)$, then the joint SF of X and Y for $x > 0$ and $y > 0$ can be decomposed as follows:

$$S_{X,Y}(x, y) = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} S_{ac}(x, y) + \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} . S_{si}(x, y).$$

Here,

$$S_{si}(x, y) = (1 - e^{-\frac{\lambda}{x}})^{\alpha_1 + \alpha_2 + \alpha_3}$$

if $x = y$, zero, otherwise and

$$S_{ac}(x, y) = \frac{\alpha_1 + \alpha_2 + \alpha_3}{\alpha_1 + \alpha_2} (1 - e^{-\frac{\lambda}{x}})^{\alpha_1} (1 - e^{-\frac{\lambda}{y}})^{\alpha_2} (1 - e^{-\frac{\lambda}{z}})^{\alpha_3} - \frac{\alpha_3}{\alpha_1 + \alpha_2} (1 - e^{-\frac{\lambda}{x}})^{\alpha_1 + \alpha_2 + \alpha_3},$$

here $z = \max\{x, y\}$.

PROOF: The proof is not difficult and it is omitted.

From Theorem 2.2, we can immediately obtain the joint PDF of X and Y as follows:

$$f_{X,Y}(x, y) = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} f_{ac}(x, y) + \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} f_{si}(z),$$

where

$$f_{ac}(x, y) = \frac{\alpha_1 + \alpha_2 + \alpha_3}{\alpha_1 + \alpha_2} \times \begin{cases} f_{IGE}(x; \alpha_1, \lambda) f_{IGE}(y; \alpha_2 + \alpha_3, \lambda) & \text{if} \quad 0 < x < y < \infty \\ f_{IGE}(x; \alpha_1 + \alpha_3, \lambda) f_{IGE}(y; \alpha_2, \lambda) & \text{if} \quad 0 < y < x < \infty \end{cases}$$

and

$$f_{si}(z) = f_{IGE}(z; \alpha_1 + \alpha_2 + \alpha_3, \lambda).$$

In Figure 3, we have provided the surface plots of the absolutely continuous part of the joint PDF of $\text{BIGE}(\alpha_1, \alpha_2, \alpha_3, \lambda)$ distribution for different values of α_1, α_2 and α_3 keeping $\lambda = 1$. It is observed that for all values of α_1, α_2 and α_3, the joint PDF is an unimodal function.

It may be mentioned that the generation of a random sample from a given distribution is very important. It is needed to perform any simulation experiments. It is very simple to generate a random sample from a BIGE($\alpha_1, \alpha_2, \alpha_3, \lambda$). The following algorithm can be used for that purpose.

Algorithm 1:

Step 1: Generate v_1, v_2 and v_3 independently from uniform $(0, 1)$.

Step 2: Compute $u_1 = -\lambda(\ln(1 - (1 - v_1)^{1/\alpha_1}))^{-1}$, $u_2 = -\lambda(\ln(1 - (1 - v_2)^{1/\alpha_2}))^{-1}$ and $u_3 = -\lambda(\ln(1 - (1 - v_3)^{1/\alpha_3}))^{-1}$.

Step 3: Compute $x = \min\{u_1, u_3\}$ and $y = \min\{u_2, u_3\}$. Then (x, y) is the required sample.

2.3 Different Properties

In this section we provide different properties of the BIGE distribution. We provide the copula associated with the BIGE distribution. This can be useful for deriving several dependency measures and also dependency properties, see also Popović et al. [31] in this respect. It may be useful for developing inference procedures of the unknown parameters. This can have some independent interests also. The following results provide the marginals, the minimum and the stress-strength measure of a BIGE distribution.

Theorem 2.3: Suppose $(X, Y) \sim$ BIGE$(\alpha_1, \alpha_2, \alpha_3, \lambda)$, then

(a) $X \sim$ IGE$(\alpha_1 + \alpha_3, \lambda)$ and $Y \sim$ IGE$(\alpha_2 + \alpha_3, \lambda)$.

(b) $\min\{X, Y\} \sim$ IGE$(\alpha_1 + \alpha_2 + \alpha_3, \lambda)$.

(c) $P(X < Y) = \dfrac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3}$.

Proof: It is not very difficult to prove. The details are avoided.

Now we provide the copula associated with the BIGE distribution. In this case we work with the survival copula which will be easy to deal with, since the joint SF and also the marginal SFs are in convenient forms. It may be recalled that every bivariate survival function $S_{X,Y}(x,y)$, with continuous marginal SFs $S_X(x)$ and $S_Y(y)$, corresponds a unique $C : [0,1] \times [0,1] \to [0,1]$, called a copula such that for all $(x,y) \in (-\infty,\infty) \times (-\infty,\infty)$,

$$S_{X,Y}(x,y) = C(S_X(x), S_Y(y)),$$

see Nelsen [28]. Moreover, in this case the unique copula C can be obtained as

$$C(u,v) = S_{X,Y}(S_X^{-1}(u), S_Y^{-1}(v)),$$

for $0 < u,v < 1$. If $(X,Y) \sim \text{BIGE}(\alpha_1,\alpha_2,\alpha_3,\lambda)$, then the corresponding survival copula function for $0 < u,v < 1$ becomes

$$C(u,v) = \begin{cases} u^{1-\beta_1}v & \text{if} \quad u^{\beta_1} \geq v^{\beta_2} \\ uv^{1-\beta_2} & \text{if} \quad u^{\beta_1} > v^{\beta_2}, \end{cases}$$

here $\beta_1 = \dfrac{\alpha_3}{\alpha_1+\alpha_3}$ and $\beta_2 = \dfrac{\alpha_3}{\alpha_2+\alpha_3}$. The above copula is the well known Marshall-Olkin copula, see Nelsen [28]. Therefore, it immediately follows that for a $\text{BIGE}(\alpha_1,\alpha_2,\alpha_3,\lambda)$ distribution, the Kendall's τ and Spearman's ρ become $\dfrac{\beta_1\beta_2}{\beta_1-\beta_1\beta_2+\beta_2}$ and $\dfrac{3\beta_1\beta_2}{2\beta_1-\beta_1\beta_2+2\beta_2}$, respectively. Using the copula other dependency properties can be easily established.

3 Maximum Likelihood Estimators

In this section we develop the ML estimators of the unknown parameters based on a random sample of size n from a $\text{BIGE}(\alpha_1,\alpha_2,\alpha_3,\lambda)$. It is assumed that we have the following data

$$\mathcal{D} = \{(x_1,y_1),\ldots,(x_n,y_n)\}.$$

We use the following notations:

$$I_1 = \{i : x_i < y_i\}, \quad I_2 = \{i; x_i > y_i\}, \quad I_3 = \{i; x_i = y_i = v_i\}, \quad I = I_1 \cup I_2 \cup I_3,$$

$$|I_1| = n_1, \quad |I_2| = n_2, \quad |I_3| = n_3.$$

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

$$\begin{aligned}
l(\alpha_1, \alpha_2, \alpha_3, \lambda|\mathcal{D}) &= \sum_{i \in I_1} \{\ln f_{IGE}(x_i; \alpha_1, \lambda) + \ln f_{IGE}(y_i; \alpha_2 + \alpha_3, \lambda)\} \\
&\quad + \sum_{i \in I_2} \{\ln f_{IGE}(x_i; \alpha_1 + \alpha_3, \lambda) + \ln f_{IGE}(y_i; \alpha_2, \lambda)\} \\
&\quad + n_3(\ln \alpha_3 - \ln(\alpha_1 + \alpha_2 + \alpha_3)) + \sum_{i \in I_3} \ln f_{IGE}(v_i; \alpha_1 + \alpha_2 + \alpha_3, \lambda) \\
&= n \ln \lambda + n_1 \ln \alpha_1 + n_1 \ln(\alpha_2 + \alpha_3) + n_2 \ln(\alpha_1 + \alpha_3) + n_2 \ln \alpha_2 + n_3 \ln \alpha_3 \\
&\quad - \lambda \left\{ \sum_{i \in I_1 \cup I_2} (x_i^{-1} + y_i^{-1}) + \sum_{i \in I_3} v_i^{-1} \right\} \\
&\quad + \alpha_1 \left\{ \sum_{i \in I_1} \ln(1 - e^{-\frac{\lambda}{x_i}}) + \sum_{i \in I_2} \ln(1 - e^{-\frac{\lambda}{x_i}}) + \sum_{i \in I_3} \ln(1 - e^{-\frac{\lambda}{v_i}}) \right\} \\
&\quad + \alpha_2 \left\{ \sum_{i \in I_1} \ln(1 - e^{-\frac{\lambda}{y_i}}) + \sum_{i \in I_2} \ln(1 - e^{-\frac{\lambda}{y_i}}) + \sum_{i \in I_3} \ln(1 - e^{-\frac{\lambda}{v_i}}) \right\} \\
&\quad + \alpha_3 \left\{ \sum_{i \in I_1} \ln(1 - e^{-\frac{\lambda}{y_i}}) + \sum_{i \in I_2} \ln(1 - e^{-\frac{\lambda}{x_i}}) + \sum_{i \in I_3} \ln(1 - e^{-\frac{\lambda}{v_i}}) \right\} \\
&\quad - \sum_{i \in I_1 \cup I_2} \ln(1 - e^{-\frac{\lambda}{x_i}}) - \sum_{i \in I_1 \cup I_2} \ln(1 - e^{-\frac{\lambda}{y_i}}) - \sum_{i \in I_3} \ln(1 - e^{-\frac{\lambda}{v_i}}).
\end{aligned}$$

Therefore, the MLEs of the unknown parameters can be obtained by maximizing the log-likelihood function $l(\alpha_1, \alpha_2, \alpha_3, \lambda|\mathcal{D})$ with respect to the unknown parameters. It is a four dimensional optimization problem, and it can be obtained by solving four non-linear equations simultaneously. Clearly, it cannot be obtained in explicit forms. One needs to use some numerical procedure to obtain these solutions. Moreover, it will be iterative in nature. Therefore, very good initial estimates are needed for any iterative procedure to converge. To avoid that we propose to use EM algorithm to compute the MLEs. We treat this problem as a missing value problem. The basic idea is as follows.

Suppose instead of observing only (X, Y), we also observe a random vector (Δ_1, Δ_2) indicting the associated U, V and W, where U, V and W are same as defined before, as follows:

Case 1 ($X < Y$): $\Delta_1 = 1$ and

$$\Delta_2 = \begin{cases} 2 & \text{if} \quad Y = V \\ 3 & \text{if} \quad Y = W. \end{cases}$$

Case 2 ($X > Y$): $\Delta_2 = 2$ and

$$\Delta_1 = \begin{cases} 1 & \text{if} \quad X = U \\ 3 & \text{if} \quad X = W. \end{cases}$$

Case 3 ($X = Y$): $\Delta_1 = \Delta_2 = 3$.

Now we will provide the likelihood contribution of the complete data $(x, y, \delta_1, \delta_2)$ for different cases. Note that if $x < y$, then the log-likelihood contribution of $(x, y, 1, 2)$ is

$$\ln f_{IGE}(x; \alpha_1, \lambda) + \ln f_{IGE}(y; \alpha_2, \lambda) + \ln S_{IGE}(y; \alpha_3, \lambda)$$

and of $(x, y, 1, 3)$ is

$$\ln f_{IGE}(x; \alpha_1, \lambda) + \ln f_{IGE}(y; \alpha_3, \lambda) + \ln S_{IGE}(y; \alpha_2, \lambda).$$

If $x > y$, then the log-likelihood contribution of $(x, y, 1, 2)$ is

$$\ln f_{IGE}(x; \alpha_1, \lambda) + \ln f_{IGE}(y; \alpha_2, \lambda) + \ln S_{IGE}(x; \alpha_3, \lambda)$$

and of $(x, y, 3, 2)$ is

$$\ln f_{IGE}(x; \alpha_3, \lambda) + \ln f_{IGE}(y; \alpha_3, \lambda) + \ln S_{IGE}(x; \alpha_1, \lambda).$$

If $x = y$, then the log-likelihood contribution of $(x, x, 3, 3)$ is

$$\ln f_{IGE}(x; \alpha_3, \lambda) + \ln S_{IGE}(x; \alpha_1, \lambda) + \ln S_{IGE}(x; \alpha_2, \lambda).$$

It can be easily checked that if (Δ_1, Δ_2) is also known, the MLEs of the unknown parameters can be obtained by solving only one non-linear equation. Therefore, for the 'complete' data set $\{(x_i, y_i, \delta_{1i}, \delta_{2i}); i = 1, \ldots, n\}$, the MLEs of the unknown parameters can be obtained by solving one non-linear equation instead of solving four non-linear equations simultaneously as we have seen before, and that is the main motivation of the proposed EM algorithm.

We need the following discussions for further development. Let us consider Table 2 which provides different configuration of X, Y and the associate configuration of U, V and W. It will help us to understand which variables are missing and with what probability?

Table 1: Possible configuration and Missing Variables.

Set	Relation Between X and Y	Possible Configuration of U, V, W	Observed Variable	Conditional Probability	Missing
I_1	$X < Y$	$U < V < W$	$U = X, V = Y$	$\frac{\alpha_2}{(\alpha_2 + \alpha_3)}$	W
		$U < W < V$	$U = X, W = Y$	$\frac{\alpha_3}{(\alpha_2 + \alpha_3)}$	V
I_2	$Y < X$	$V < U < W$	$V = Y, U = X$	$\frac{\alpha_1}{(\alpha_1 + \alpha_3)}$	W
		$V < W < U$	$V = Y, W = X$	$\frac{\alpha_3}{(\alpha_1 + \alpha_3)}$	U
I_3	$X = Y$	$W < U < V$	$W = X = Y$	$\frac{\alpha_1}{(\alpha_1 + \alpha_2)}$	U, V
		$W < V < U$	$W = X = Y$	$\frac{\alpha_2}{(\alpha_1 + \alpha_2)}$	U, V

Now we can define the EM algorithm. Let us define the estimates of α_1, α_2, α_3 and λ at the k-th stage of the EM algorithm as $\alpha_1^{(k)}$, $\alpha_2^{(k)}$, $\alpha_3^{(k)}$ and $\lambda^{(k)}$, respectively. Let us also define

$$a^{(k)} = \frac{\alpha_2^{(k)}}{\alpha_2^{(k)} + \alpha_3^{(k)}} \quad \text{and} \quad b^{(k)} = \frac{\alpha_1^{(k)}}{\alpha_1^{(k)} + \alpha_3^{(k)}},$$

$\Theta = (\alpha_1, \alpha_2, \alpha_3, \lambda)$ and $\Theta^{(k)} = (\alpha_1^{(k)}, \ \alpha_2^{(k)}, \alpha_3^{(k)}, \lambda^{(k)})$. At the 'E'-step, the 'pseudo' log-likelihood function can be obtained as the expected log-likelihood function, and it can be written as

$$
\begin{aligned}
l_p(\Theta|\Theta^{(k)}) &= a^{(k)}\left[\sum_{i\in I_1}\ln f_{IGE}(x_i;\alpha_1,\lambda)+\sum_{i\in I_1}\ln f_{IGE}(y_i;\alpha_2,\lambda)+\sum_{i\in I_1}\ln S_{IGE}(y_i;\alpha_3,\lambda)\right]+\\
&\quad (1-a^{(k)})\left[\sum_{i\in I_1}\ln f_{IGE}(x_i;\alpha_1,\lambda)+\sum_{i\in I_1}\ln f_{IGE}(y_i;\alpha_3,\lambda)+\sum_{i\in I_1}\ln S_{IGE}(y_i;\alpha_2,\lambda)\right]+\\
&\quad b^{(k)}\left[\sum_{i\in I_2}\ln f_{IGE}(x_i;\alpha_1,\lambda)+\sum_{i\in I_2}\ln f_{IGE}(y_i;\alpha_2,\lambda)+\sum_{i\in I_2}\ln S_{IGE}(x_i;\alpha_3,\lambda)\right]+\\
&\quad (1-b^{(k)})\left[\sum_{i\in I_2}\ln f_{IGE}(x_i;\alpha_3,\lambda)+\sum_{i\in I_2}\ln f_{IGE}(y_i;\alpha_2,\lambda)+\sum_{i\in I_2}\ln S_{IGE}(x_i;\alpha_1,\lambda)\right]+\\
&\quad \left[\sum_{i\in I_3}\ln f_{IGE}(z_i;\alpha_3,\lambda)+\sum_{i\in I_3}\ln S_{IGE}(x_i;\alpha_1,\lambda)+\sum_{i\in I_3}\ln S_{IGE}(y_i;\alpha_2,\lambda)\right]\\
&= (n_1+b^{(k)}n_2)\ln\alpha_1+\alpha_1\sum_{i\in I}\ln(1-e^{-\frac{\lambda}{x_i}})+(n_2+a^{(k)}n_1)\ln\alpha_2+\\
&\quad \alpha_2\sum_{i\in I}\ln(1-e^{-\frac{\lambda}{y_i}})+(n_1(1-a^{(k)})+(1-b^{(k)})n_2+n_3)\ln\alpha_3+\\
&\quad \alpha_3\left[\sum_{i\in I_1}\ln(1-e^{-\frac{\lambda}{y_i}})+\sum_{i\in I_2}\ln(1-e^{-\frac{\lambda}{x_i}})+\sum_{i\in I_3}\ln(1-e^{-\frac{\lambda}{z_i}})\right]+\\
&\quad (2n_1+2n_2+n_3)\ln\lambda-\lambda\left[\sum_{i\in I_1\cup I_2}\left(\frac{1}{x_i}+\frac{1}{y_i}\right)+\sum_{i\in I_3}\frac{1}{z_i}\right]-\\
&\quad \left[\sum_{i\in I_1\cup I_2}\left(\ln(1-e^{-\frac{\lambda}{x_i}})+\ln(1-e^{-\frac{\lambda}{y_i}})\right)+\sum_{i\in I_3}\ln(1-e^{-\frac{\lambda}{z_i}})\right]. \qquad (2)
\end{aligned}
$$

At the 'M'-step $\Theta^{(k+1)}$ can be obtained by maximizing (2) with respect to Θ. Note that for a fixed λ,

$$
\alpha_1^{(k+1)}(\lambda)=\frac{n_1+b^{(k)}n_2}{\sum_{i\in I}\ln(1-e^{-\frac{\lambda}{x_i}})},\qquad \alpha_2^{(k+1)}(\lambda)=\frac{n_2+a^{(k)}n_1}{\sum_{i\in I}\ln(1-e^{-\frac{\lambda}{y_i}})},
$$

$$
\alpha_3^{(k+1)}(\lambda)=\frac{n_1(1-a^{(k)})+n_2(1-b^{(k)})+n_3}{\sum_{i\in I_2}\ln(1-e^{-\frac{\lambda}{x_i}})+\sum_{i\in I_1}\ln(1-e^{-\frac{\lambda}{y_i}})+\sum_{i\in I_3}\ln(1-e^{-\frac{\lambda}{z_i}})},
$$

maximize (2).

Hence, $\lambda^{(k+1)}$ which maximizes (2) can be obtained by $\lambda^{(k+1)} = \arg\max g(\lambda)$, where

$$
\begin{aligned}
g(\lambda) \;=\;& (n_1 + b^{(k)} n_2) \ln \alpha_1^{(k+1)}(\lambda) + (n_2 + a^{(k)} n_1) \ln \alpha_2^{(k+1)}(\lambda) + \\
& (n_1(1 - a^{(k)}) + (1 - b^{(k)}) n_2 + n_3) \ln \alpha_3^{(k+1)}(\lambda) + \\
& (2n_1 + 2n_2 + n_3) \ln \lambda - \lambda \left[\sum_{i \in I_1 \cup I_2} \left(\frac{1}{x_i} + \frac{1}{y_i} \right) + \sum_{i \in I_3} \frac{1}{z_i} \right] - \\
& \left[\sum_{i \in I_1 \cup I_2} \left(\ln(1 - e^{-\frac{\lambda}{x_i}}) + \ln(1 - e^{-\frac{\lambda}{y_i}}) \right) + \sum_{i \in I_3} \ln(1 - e^{-\frac{\lambda}{z_i}}) \right].
\end{aligned} \tag{3}
$$

Hence,

$$
\alpha_1^{(k+1)} = \alpha_1^{(k+1)}(\lambda^{(k+1)}), \quad \alpha_2^{(k+1)} = \alpha_2^{(k+1)}(\lambda^{(k+1)}), \quad \alpha_3^{(k+1)} = \alpha_3^{(k+1)}(\lambda^{(k+1)}). \tag{4}
$$

Therefore, the following algorithm can be used for the implementation of the EM algorithm at the k-th stage.

EM Algorithm

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

Step 2: Find the value of λ so that it maximizes (3), say $\lambda^{(k+1)}$.

Step 3: Compute $\alpha_1^{(k)}$, $\alpha_2^{(k)}$ and $\alpha_3^{(k)}$, using (4).

Step 4: Check the convergence, if it does not converge go back to Step 1 and replace k by $k + 1$.

∎

4 Simulations and Data Analysis

4.1 Simulations

In this section we present some simulation results to show how the proposed EM algorithm behaves in practice. We have used different sets of parameter values and different sample sizes. In all cases we have used the EM algorithm as proposed in the previous section, and we have stopped the iteration whenever the difference between two consecutive 'pseudo' log-likelihood values is less than 10^{-6}. We have used two different sets of parameter values namely: Set 1: $(\alpha_1, \alpha_2, \alpha_3, \lambda)$ = (2.0,2.0,2.0,1.0) and Set 2: $(\alpha_1, \alpha_2, \alpha_3, \lambda)$ = (0.5,0.5,0.5,1) and four different sample sizes: 25, 50, 75, 100. In each case we have calculated the estimates of the unknown parameters and replicate 1000 times, and report the average estimates and the square root of the mean squared errors (SMSE).

Table 2: Average estimates and SMSEs of the MLEs: Set 1.

n	α_1	α_2	α_3	λ
25	2.3530 (1.1295)	2.3140 (1.0927)	2.3364 (1.0210)	1.0596 (0.1814)
50	1.1802 (0.6618)	2.0519 (0.6638)	2.1570 (0.6096)	1.0308 (0.1226)
75	2.1172 (0.4954)	2.1065 (0.5323)	2.1054 (0.4482)	1.0193 (0.0952)
100	2.0867 (0.4190)	2.0761 (0.4278)	2.0776 (0.3757)	1.0143 (0.0803)

Some of the points are quite clear from Tables 2 and 3. In both the cases it is observed that the average biases and SMSE decrease as the sample size increases. It verifies the consistency properties of the MLEs. It is observed in all the cases the EM algorithm converges within 20 steps. It shows that the EM algorithm works quite well. We have also tried with different

Table 3: Average estimates and SMSEs of the MLEs: Set 2.

n	α_1	α_2	α_3	λ
25	0.5524 (0.2140)	0.5429 (0.2106)	0.5518 (0.1964)	1.0780 (0.2342)
50	0.5302 (0.1376)	0.5225 (0.1375)	0.5250 (0.1235)	1.0434 (0.1652)
75	0.5204 (0.1044)	0.5158 (0.1133)	0.5182 (0.0955)	1.0281 (0.1294)
100	0.5152 (0.0898)	0.5124 (0.0915)	0.5133 (0.0799)	1.0206 (0.1090)

initial guesses, and it is observed that it converges to the same point. It indicates that the EM algorithm converges to the global maximum rather than the local maximum.

4.2 Data Analysis

In this section we have analyzed one bivariate data set with ties. The data set was originally obtained from Meintanis [26] and it is presented in Table 1. The main purpose of this data analysis is to verify how the ~~proposed~~ model and the proposed EM algorithm work in practice. 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 represents the time in minutes of the first kick goal scored by any team and Y represents the first goal of any type scored by the home team. It may be noted that in this case all possibilities are open. For example $X < Y$ indicates that a kick goal has been scored by the opponent and it has been scored before the home team has scored any goal. Similarly, $Y < X$ means the home team has scored the first goal before any kick goal by either team has been scored. Similarly, $X = Y$ means the first goal scored by the home

Table 4: UEFA Champion's League data

2005-2006	X	Y	2004-2005	X	Y
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			

team is a kick goal and it has been scored before any kick goal (if at all) has been scored by the opponent. Meintanis [26] used Marshall-Olkin bivariate exponential model and showed that it does not fit the data set well.

We have fitted the BIGE distribution to this data set. We have used the EM algorithm proposed here with the initial guesses as $\alpha_1^0 = \alpha_2^0 = \alpha_3^0 = 1$. We have used the same stopping criterion as it has been used in the simulation study. We have provided the estimated values of α_1, α_2, α_3, λ and the corresponding log-likelihood (log-like) values at each iteration (Iter) in Table 5. It is observed that the 'pseudo' log-likelihood value increases in each iteration and the iteration stops at 18th iteration. The final estimates and the associated 95% confidence

Table 5: Estimates and log-likelihood values at each iteration

α_1	α_2	α_3	λ	log-like	Iter
0.4793	0.8472	0.7799	2.2040	-21.076	1
0.4116	0.8512	0.8372	2.2004	-19.859	2
0.3824	0.8456	0.8653	2.1962	-19.197	3
0.3691	0.8424	0.8785	2.1943	-18.864	4
0.3632	0.8412	0.8849	2.1943	-18.704	5
0.3605	0.8406	0.8878	2.1943	-18.629	6
0.3592	0.8404	0.8892	2.1944	-18.595	7
0.3587	0.8403	0.8897	2.1943	-18.579	8
0.3584	0.8402	0.8900	2.1944	-18.572	9
0.3583	0.8402	0.8901	2.1943	-18.568	10
0.3582	0.8402	0.8902	2.1943	-18.567	11
0.3582	0.8402	0.8903	2.1944	-18.566	12
0.3582	0.8402	0.8902	2.1943	-18.566	13
0.3582	0.8402	0.8902	2.1943	-18.566	14
0.3582	0.8401	0.8902	2.1942	-18.566	15
0.3582	0.8401	0.8902	2.1942	-18.566	16
0.3582	0.8402	0.8902	2.1943	-18.566	17
0.3582	0.8402	0.8902	2.1943	-18.566	18

intervals are as follows

$$\widehat{\alpha}_1 = 0.3582\ (\mp 0.1013), \qquad \widehat{\alpha}_2 = 0.8402\ (\mp 0.2331), \widehat{\alpha}_3 = 0.8902\ (\mp 0.2855)$$

$$\widehat{\lambda} = 2.1943\ (\mp 0.6573).$$

The associate log-likelihood value is -18.57.

Just to verify whether the convergence of the EM algorithm is to the global maximum or not, we have performed the maximization of the log-likelihood function using a four dimensional grid search method with grid size 10^{-6} and search the global maximum in the set $[0, 10] \times [0, 10] \times [0, 10] \times [0, 10]$. It converges to the same point as above but it took more than three hours in a Lenovo machine with Intel Core-i5 processor, where as in the same machine the EM algorithm converges in less than 15 seconds.

Now we would like to see whether the ~~proposed~~ BIGE distribution fits the above data set or not. It may be mentioned that although extensive work has been on the goodness of fit tests for univariate continuous distributions, see D'Agostino and Stephens [11], the same is not true in case of multivariate distributions, except for multivariate normal distribution, see for example Srivastava and Mudholkar [36]. Since we know that in case of BIGE distribution, the two marginals and the minimum follow IGE distributions, we test for those. Although, it is not sufficient, at least it is necessary.

~~The Kolmogorov-Smirnov distances and the corresponding p-values (reported within brackets) between the empirical distribution functions of X, Y, $\min\{X, Y\}$ and the fitted distribution functions based on the MLEs are 0.205 (0.145), 0.173 (0.215), 0.180 (0.178), respectively. Since the p values are quite high in all the three cases, we cannot reject the hypothesis that the marginals and the minimum follow IGE distribution.~~ The Kolmogorov-Smirnov distances between the empirical distribution functions of X, Y, $\min\{X, Y\}$ and the fitted distribution functions based on the MLEs are 0.205, 0.173, 0.180, respectively. Now as suggested by Ristić et al. [32], based on bootstrap method, we obtain the critical values at the 5% significance level as 0.245, 0.232, 0.238, respectively. Therefore, we cannot reject the null hypothesis at the 5% significance level that X, Y and $\min\{X, Y\}$ follow IGE distributions. It seems it is reasonable to use BIGE model to analyze the above data set.

For comparison purposes we have fitted bivariate GE (BVGE) and MOBE distributions also to the same data set, see for example Kundu and Gupta [20]. The MLEs of the unknown parameters $\alpha_1, \alpha_2, \alpha_3, \lambda$ in case of BVGE distribution, using the same notations as in Kundu and Gupta [20], are 1.445, 0.468,1.170,0.039, respectively, and the associated log-likelihood value is -20.59. In case of MOBE distribution, we have used the same estimates as Meintanis [26] and obtained the associated log-likelihood value as -44.57. Therefore, the AIC (BIC) values for BIGE, BVGE and MOBE are 45.14 (43.56), 49.18 (48.40), 95.14 (94.56), respectively. Hence, based on the AIC and BIC values BIGE provides a slightly better fit than BVGE and MOBE distributions for this particular data set.

But it may be mentioned that although for this data set BIGE provides a better fit in terms of AIC and BIC, but it does not mean it always provides a better fit than any other bivariate distribution. But at least it provides another option to a practitioner for analyzing bivariate data when there are ties. Moreover, we will show in the next section that it can be used quite effectively to analyze dependent competing risks data when there are two competing causes and failure may happen at a particular time due to both causes also.

5 Applications

The motivation of this application came from a data set of the Diabetic-Retinopathy Study (DRS) conducted by the National Eye Institute. The aim is to estimate the effect of the laser treatment in delaying the onset of blindness in patients with diabetic Retinopathy. At the beginning of the study, for each patient randomly one eye has been selected for laser treatment and the other eye way not given any treatment. The data are collected as follows: the minimum time to blindness (T), and the indicator denoting whether the treated eye, untreated eye or both eyes. Clearly time to blindness of two eyes are dependent and there are some ties also occur, i.e. both the eyes became blind simultaneously. So for a typical patient the observation is as follows (T, δ), here δ can take values 0, 1 or 2, indicates both eyes, treated eye and untreated eye, respectively.

Suppose the lifetime of the treated eye is X and that of the untreated eye is Y. It is assumed $(X, Y) \sim \text{BIGE}(\alpha_1, \alpha_2, \alpha_3, \lambda)$, and $T = \min\{X, Y\}$. With the abuse of notations we have used the following notations. The data set $\mathcal{D} = \{(t_1, \delta_1), \ldots, (t_n, \delta_n)\}$, the set I_1, I_2 and I_3 are as follows: $I_1 = \{i; \delta_i = 1\}$, $I_2 = \{i; \delta_i = 2\}$, $I_3 = \{i; \delta_i = 3\}$ and $|I_1| = n_1$,

$|I_2| = n_2$ and $|I_3| = n_3$. Our problem is to estimate the unknown parameters based on the data set $\mathcal{D}$. The log-likelihood function can be written as

$$\begin{aligned}
l(\alpha_1, \alpha_2, \alpha_3, \lambda|\mathcal{D}) &= \sum_{i \in I_1} \{\ln f_{IGE}(t_i; \alpha_1, \lambda) + \ln S_{IGE}(t_i; \alpha_2, \lambda) + \ln S_{IGE}(t_i; \alpha_3, \lambda)\} + \\
&\quad \sum_{i \in I_2} \{\ln f_{IGE}(t_i; \alpha_2, \lambda) + \ln S_{IGE}(t_i; \alpha_1, \lambda) + \ln S_{IGE}(t_i; \alpha_3, \lambda)\} + \\
&\quad \sum_{i \in I_3} \{\ln f_{IGE}(t_i; \alpha_3, \lambda) + \ln S_{IGE}(t_i; \alpha_1, \lambda) + \ln S_{IGE}(t_i; \alpha_2, \lambda)\} \\
&= \sum_{i \in I_1} \ln f_{IGE}(t_i; \alpha_1, \lambda) + \sum_{i \in I_2} \ln S_{IGE}(t_i; \alpha_1, \lambda) + \sum_{i \in I_3} \ln S_{IGE}(t_i; \alpha_1, \lambda) + \\
&\quad \sum_{i \in I_2} \ln f_{IGE}(t_i; \alpha_2, \lambda) + \sum_{i \in I_1} \ln S_{IGE}(t_i; \alpha_2, \lambda) + \sum_{i \in I_3} \ln S_{IGE}(t_i; \alpha_2, \lambda) + \\
&\quad \sum_{i \in I_3} \ln f_{IGE}(t_i; \alpha_3, \lambda) + \sum_{i \in I_1} \ln S_{IGE}(t_i; \alpha_3, \lambda) + \sum_{i \in I_2} \ln S_{IGE}(t_i; \alpha_3, \lambda).
\end{aligned}$$

It can be easily observed that for a given λ, the MLEs of α_1, α_2 and α_3 can be obtained as

$$\widehat{\alpha}_1(\lambda) = -\frac{n_1}{\sum_{i=1}^n \ln(1 - e^{-\frac{\lambda}{t_i}})}, \quad \widehat{\alpha}_2(\lambda) = -\frac{n_2}{\sum_{i=1}^n \ln(1 - e^{-\frac{\lambda}{t_i}})}, \quad \widehat{\alpha}_3(\lambda) = -\frac{n_3}{\sum_{i=1}^n \ln(1 - e^{-\frac{\lambda}{t_i}})},$$

and the MLE of λ can be obtained as maximizing the following function $h(\lambda)$ where

$$h(\lambda) = n_1 \ln(\widehat{\alpha}_1(\lambda)) + n_2 \ln(\widehat{\alpha}_2(\lambda)) + n_3 \ln(\widehat{\alpha}_3(\lambda)) + n \ln \lambda - \lambda \sum_{i=1}^n t_i^{-1} - \sum_{i=1}^n \ln(1 - e^{-\frac{\lambda}{t_i}}).$$

Hence, if $\widehat{\lambda}$ maximizes $h(\lambda)$, then the MLEs of α_1, α_2 and α_3 are $\widehat{\alpha}_1 = \widehat{\alpha}_1(\widehat{\lambda})$, $\widehat{\alpha}_2 = \widehat{\alpha}_2(\widehat{\lambda})$ and $\widehat{\alpha}_3 = \widehat{\alpha}_3(\widehat{\lambda})$, respectively. The associated confidence intervals can be obtained based on the observed Fisher information matrix.

To see how the proposed model works in practice we have analyzed a competing risks data using this model. The data come from the Diabetic Retinopathy (DR) Study conducted by the National Eye Institute to analyze the effect of laser treatment in delaying the onset of blindness in patients with DR symptoms. The randomized experiment is designed as follows. For each patient, at the beginning of the experiment, one eye was randomly selected for laser treatment and the other one was not given any treatment. The data set, the minimum time

Table 6: Minimum time to blindness in days and its cause for 71 patients with Diabetic Retinopathy

i	T	δ	i	T	δ	i	T	δ	i	T	δ	i	T	δ
1	266	1	16	125	2	31	717	2	46	663	3	61	503	1
2	91	2	17	777	2	32	642	1	47	567	2	62	423	2
3	154	2	18	306	1	33	141	2	48	966	3	63	285	2
4	285	3	19	415	1	34	407	1	49	203	3	64	315	2
5	583	1	20	307	2	35	356	1	50	84	1	65	727	2
6	547	2	21	637	2	36	1653	3	51	392	1	66	210	2
7	79	1	22	577	2	37	427	2	52	1140	2	67	409	2
8	622	3	23	178	1	38	699	1	53	901	1	68	584	1
9	707	2	24	517	2	39	36	2	54	1247	3	69	355	1
10	469	2	25	272	3	40	667	1	55	448	2	70	1302	1
11	93	1	26	1137	3	41	588	2	56	904	2	71	227	2
12	1313	2	27	1484	1	42	471	3	57	276	1			
13	805	1	28	315	1	43	126	1	58	520	1			
14	344	1	29	287	2	44	350	2	59	485	2			
15	790	2	30	1252	1	45	350	1	60	248	2			

to blindness (T) and the indicator specifying whether treated eye ($\delta = 1$), untreated eye ($\delta = 2$) or both eyes ($\delta = 3$) failed, is presented in Table 6. In this case the treatment or lack of treatment can be considered as the causes of blindness, hence, it can be considered as a competing risks data. Clearly, time to blindness of the two eyes are dependent, and there are some ties, hence, the proposed model has been used for analyzing this data set.

The preliminary data analysis indicates that the number of patients with $\delta = 1$, 2 and 3 are 28, 33 and 10, respectively. Because of the complicated nature of $h(\lambda)$ we could not prove that it is an unimodal function, but it is observed to be an unimodal function, see Figure 4. Finally we obtain the MLEs of λ, α_1, α_2 and α_3 and the associated 95% confidence intervals are

$$\widehat{\lambda} = 4.356682 \ (\mp 0.982365)$$

$$\widehat{\alpha}_1 = 0.747439 \ (\mp 0.187649), \quad \widehat{\alpha}_2 = 0.880911 \ (\mp 0.201578), \quad \widehat{\alpha}_2 = 0.266942 \ (\mp 0.053242).$$

Now one natural question is based on the observed sample (T, δ), how to test that the original (X, Y) follows BIGE distribution. It is not possible to test that assumption based on the observed sample (T, δ). But if (X, Y) follows BIGE, then T follows IGE distribution. It is observed that the Kolmogorov-Smirnov distance between the empirical distribution function of T and the fitted IGE distribution to T based on MLEs is 0.1389. ~~and the associated p values is 0.1290.~~ Based on the bootstrap method the critical value at the 5% significance level becomes 0.1723. Hence, at the 5% significance level, ~~Since p values is large,~~ we cannot reject the hypothesis that T follows IGE distribution. Moreover, since $\delta = 0$ with positive probability, therefore, it is clear that (X, Y) follows a bivariate distribution with a singular component. Hence, we have assumed that (X, Y) follows BIGE distribution.

6 Conclusions

In this paper we have ~~proposed a new bivariate distribution namely~~ considered the bivariate inverse generalized exponential distribution which has a singular components. This model can be used when there are ties and if it is found that the data are coming from a heavy tailed distribution. We have provided a very efficient EM algorithm to compute the MLEs of the unknown parameters. It is observed that the proposed EM algorithm works quite well in practice. Further we have used this model to analyze dependent competing risks data when there are two competing causes of failures. We have analyzed one competing risks data using this model, the results are quite satisfactory. It may be mentioned that although we have developed this model when there are only two competing causes of failures, it can be extended for more than two causes also. More work is needed along that direction.

Acknowledgements:

The authors would like to thank the reviewers and the Associate editor for their constructive comments, which have helped to improve the manuscript significantly.

References

[1] Al-Hussaini, E.K. and Ahsanullah, M. (2015), *Exponentiated Distributions*, Atlantis Press, Paris, France.

[2] Arnold, B. C. (1968), "Parameter estimation for a multivariate exponential distribution", *Journal of the American Statistical Association*, vol. 63, 848 – 852.

[3] Barreto-Souza, W. and Lemonte, A.J. (2013), "Bivariate Kumaraswamy distribution: properties and a new method to generate bivariate classes", *Statistics*, vol. 47, 1321 – 1342.

[4] Baxter, L.A. and Rachev, S.T. (1991), "The stability of a characterization of the bivariate Marshall-Olkin distribution", *Journal of Mathematical Analysis and Applications*, vol. 160, 563 – 573.

[5] Begum, A. A. and Khan, A. H. (2000), "Concomitants of order statistics from Marshall and Olkin's bivariate Weibull distribution", *Calcutta Statistical Association Bulletin*, vol. 50, 197 – 198.

[6] 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.

[7] Boland, P.J. (1998), "An arrangement increasing property of the Marshall-Olkin bivariate exponential", *Statistics and Probability Letters*, vol. 37, 167 – 170.

[8] Cai, J., Shi, Y., Liu, B. (2017), "Analysis of incomplete data in the presence of dependent competing risks from Marshall-Olkin bivariate Weibull distribution under progressive hybrid censoring", *Communications in Statistics - Theory and Methods*, vol. 46, 6497 – 6511.

[9] Cox, D.R. (1959), "The analysis of exponentially distributed lifetimes with two types of failures", *Journal of the Royal Statistical Society*, Ser. B, vol. 21, 411 - 421.

[10] Crowder, M.J. (2011), *Classical Competing Risks*, Chapman & Hall, Boca Raton, Florida, USA.

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

[12] Feizjavadian, S.H. and Hashemi, R. (2015), "Analysis of dependent competing risks in presence of progressive hybrid censoring using Marshall-Olkin bivariate Weibull distribution", *Computational Statistics and Data Analysis*, vol. 82, 19 – 34.

[13] Gupta, R. C., Kirmani, S.N.U.A., Balakrishnan, N (2013), "On a class of generalized Marshall-Olkin bivariate distributions and some reliability characteristics", *Probability Engineering and Information Science*, vol. 27, 261 – 275.

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

[15] Lu, Jye-Chyi (1989), "Weibull extension of the Freund and Marshall-Olkin bivariate exponential model", *IEEE Transactions on Reliability*, vol. 38, 615 – 619.

[16] Kundu, D. and Dey, A. K. (2009), "Estimating the Parameters of the Marshall Olkin Bivariate Weibull Distribution by EM Algorithm", *Computational Statistics and Data Analysis*, vol. 53, 956 – 965.

[17] Kundu, D. and Gupta, R.D. (2010), "Modified Sarhan-Balakrishnan singular bivariate distribution", *Journal of Statistical Planning and Inference*, vol. 140, 526 – 538.

[18] Kundu, D. and Gupta, R.D. (2010), "A Class of Bivariate Models with Proportional Reversed Hazard Marginals", *Sankhya*, Ser. B, vol. 72, 236 – 253.

[19] Kundu, D. and Gupta, A.K. (2013), "Bayes estimation for the Marshall-Olkin bivariate Weibull distribution", *Computational Statistics and Data Analysis*, vol. 57, 271 – 281.

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

[21] Kundu, D. and Gupta, A. (2017), "On bivariate inverse Weibull distribution", *Brazilian Journal of Probability and Statistics*, vol. 31, 275 – 302.

[22] Lai, C-D., Lin, G. D., Govindaraju, K. and Pirikahu, S. A. (2017) A simulation study on the correlation structure of Marshall-Olkin bivariate Weibull distribution", *Journal of Statistical Computation and Simulation*, vol. 87, 156 – 170.

[23] Louis, T. A. (1982), "Finding the observed information matrix when using the EM algorithm", *Journal of the Royal Statistical Society*, Series B 44, 226 – 233.

[24] Lu, Jye-Chyi (1989), "Weibull extension of the Freund and Marshall-Olkin bivariate exponential model", *IEEE Transactions on Reliability*, vol. 38, 615 - 619.

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

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

[27] Nadarajah, S. (2011), "The exponentiated exponential distribution; a survey", *Advances in Statistical Analysis*, vol. 95, 219 - 251

[28] Nelsen, R.B. (2006), An introduction to copulas, 2nd-edition, Springer, New York.

[29] Oguntunde, P.E. and Adejumo, A.O. (2015), "The generalized inverted generalized exponential distribution with an application to a censored Data", *Journal of Statistics Applications and Probability*, vol. 4, 223 - 230.

[30] Pena, E.A. and Gupta, A.K. (1990), “Bayes estimation for the MarshallÂ¡VOlkin exponential distribution”, *Journal of the Royal Statistical Society*, Ser. B, vol. 52, 379 – 389.

[31] Popović, B.V., Ristić, M.M., Genç, A.I. (2020), “Dependence properties of multivariate distributions with proportional hazard rate marginals”, *Applied Mathematical Modelling*, vol. 77, 182–198.

[32] Ristić, M.M., Popović, B.V., Zografos, K., Balakrishnan, N. (2018), “Discrimination among bivariate beta-generated distributions”, *Statistics*, vol. 52, 303 – 320.

[33] Sarhan, A.M. and Balakrishnan, N. (2007), “A new class of bivariate distribution and its mixture”, *Journal of Multivariate Analysis*, vol. 98, 1508 – 1527.

[34] Sarhan, A.M., Hamilton, D.C., Smith, B. and Kundu, D. (2011), “The bivariate generalized linear failure rate distribution and its multivariate extension”, *Computational Statistics and Data Analysis*, vol. 55, 644 – 654.

[35] Shen, Y. and Xu, A. (2018), “On the dependent competing risks using Marshall-Olkin bivariate Weibull model: parameter estimation with different methods”, *Communications in Statistics - Theory and Methods*, vol. 47, 5558 – 5572.

[36] 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, pp 869–956.

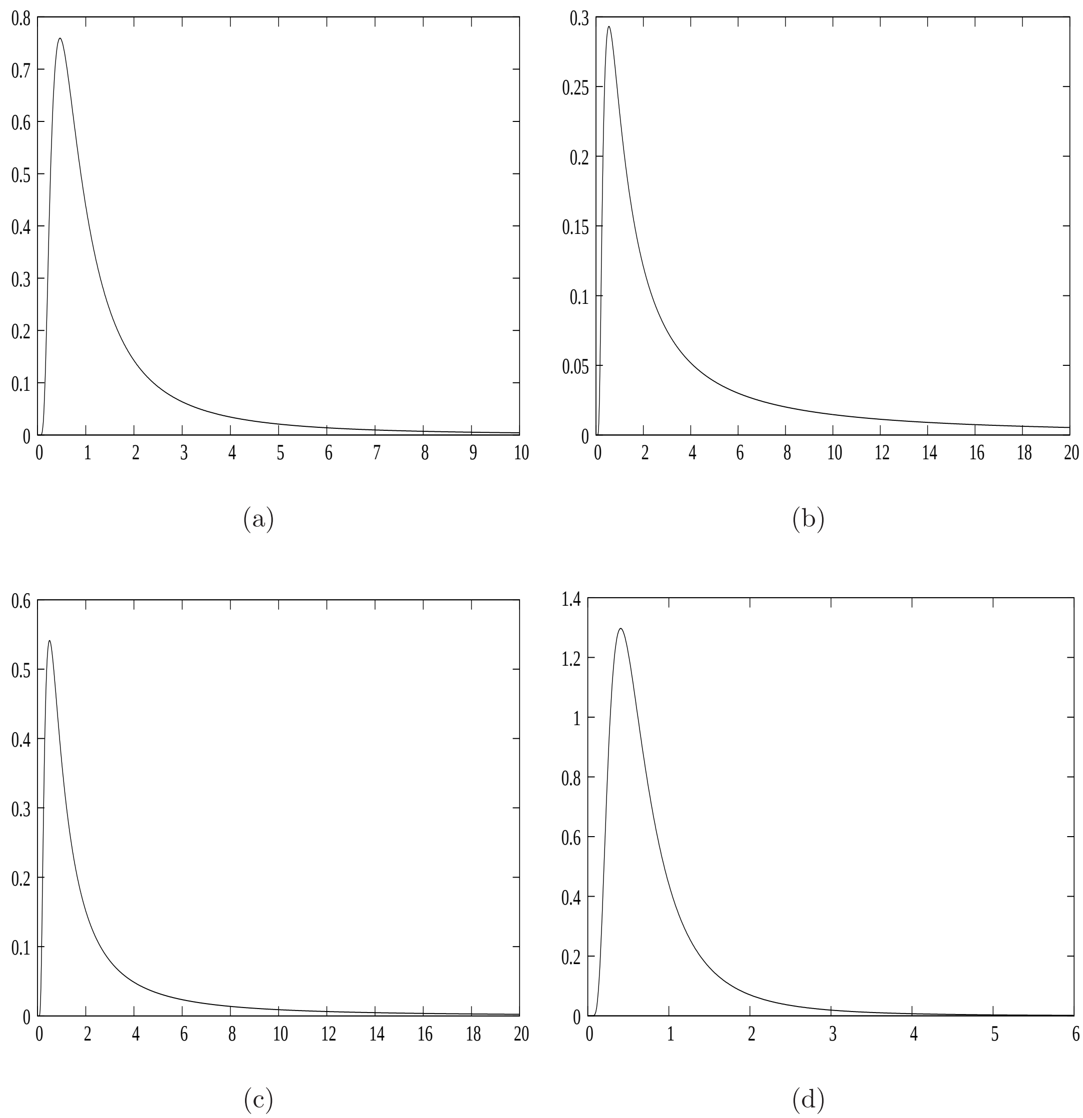

Figure 1: PDF plots of IGE($\alpha, 1$) distribution for different α values: (a) 1.5 (b) 0.5 (c) 1.0 (d) 3.0.

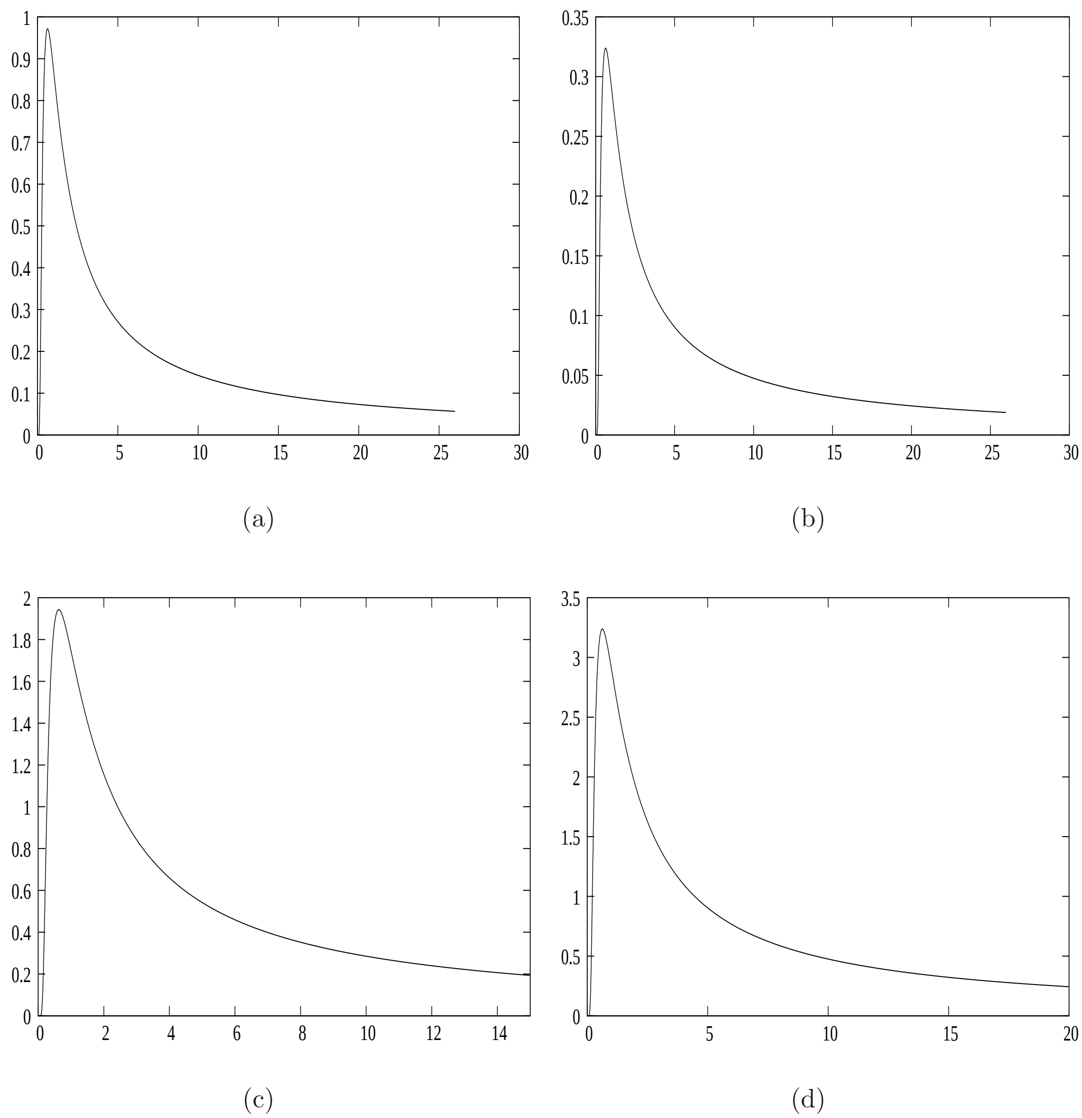

Figure 2: Hazard function plots of IGE($\alpha, 1$) distribution for different α values: (a) 1.5 (b) 0.5 (c) 3.0 (d) 5.0.

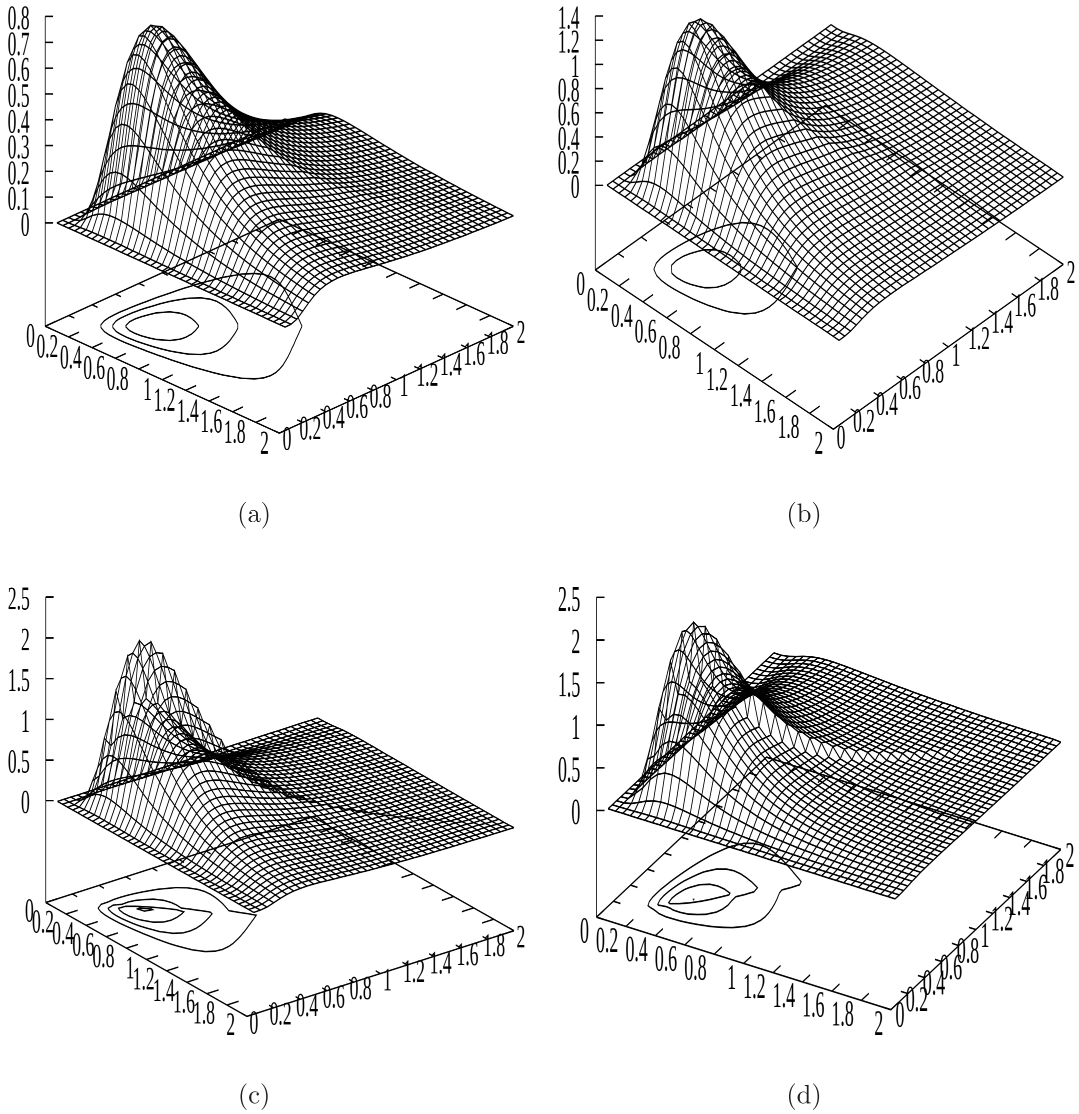

Figure 3: Bivariate surface plots of BIGE$(\alpha_1, \alpha_2, \alpha_3, \lambda)$ distribution for different α_1, α_2, α_3, λ values: (a) (1.0,1.0,1.0,1.0) (b) (1.0,1.0,2.0,1.0) (c) (1.0,2.0,1.0,1.0) (d) (2.0,1.0,2.0,1.0).

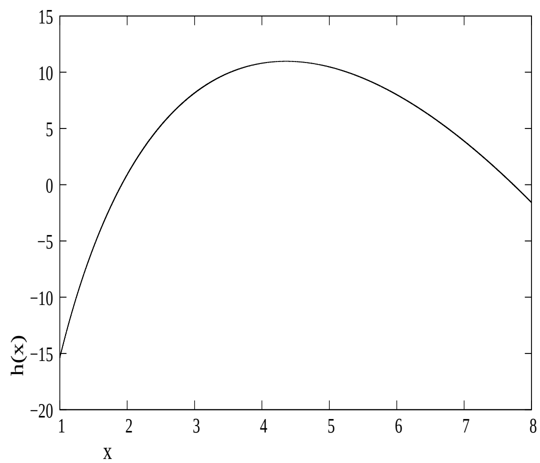

Figure 4: The plot of $h(x)$