Bivariate Distributions with Singular Components

Debasis Kundu *

Abstract

In this paper we mainly discuss classes of bivariate distributions with singular components. It is observed that there are mainly two different ways of defining bivariate distributions with singular components, when the marginals are absolutely continuous. Most of the bivariate distributions available in the literature can be obtained from these two general classes. A connection between the two approaches can be established based on their copulas. It is observed that under certain restrictions both these classes have very similar copulas. Several properties can be established of these proposed classes. It is observed that the maximum likelihood estimators (MLEs) may not always exist, whenever they exist they cannot be obtained in closed forms. Numerical techniques are needed to compute the MLEs of the unknown parameters. Alternatively, very efficient expectation maximization (EM) algorithm can be used to compute the MLEs. The corresponding observed Fisher information matrix also can be obtained quite conveniently at the last stage of the EM algorithm, and it can be used to construct confidence intervals of the unknown parameters. The analysis of one data set has been performed to see the effectiveness of the EM algorithm. We discuss different generalizations, propose several open problems and finally conclude the paper.

Key Words and Phrases: Absolute continuous distribution; singular distribution; Fisher information matrix; EM algorithm; joint probability distribution function; joint probability density function.

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

*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

Bivariate continuous distributions occur quite naturally in practice. An extensive amount of work has been done on different bivariate continuous distributions in the statistical literature. Some of the well known absolutely continuous bivariate continuous distributions are bivariate normal, bivariate-$t$, bivariate log-normal, bivariate gamma, bivariate extreme value, bivariate Birnbaum-Saunders distributions, bivariate skew normal distribution, bivariate geometric skew normal distribution etc., see for example the books by Balakrishnan and Lai [7], Kotz et al. [27] on different bivariate and multivariate distributions, the recent review article by Balakrishnan and Kundu [6] on bivariate and multivariate Birnbaum-Saunders distributions, the article by Azzalini and Dalla Valle [4] and the monograph by Azzalini and Capitanio [3] on multivariate skew-normal distribution, the recent article by Kundu [29] on multivariate geometric skew-normal distribution, and the references cited therein. The main purpose of any bivariate distribution is to model the two marginals and also to find association between the two marginals.

It may be mentioned that although there are numerous absolutely continuous bivariate distributions available in the literature, if there are ties in the data set, then these absolutely continuous bivariate distributions cannot be used to analyze this data set. Sometimes, the ties may occur due to truncation, but in many situations the ties may occur naturally and with a positive probability. To analyze a bivariate data set with ties, one needs a bivariate model with a singular component. These class of bivariate distributions assign a positive probability on $X = Y$, where $X$ and $Y$ denote the two marginals, and both are assumed to be absolutely continuous.

Marshall and Olkin [44] first proposed a bivariate distribution such that its both the marginals $X$ and $Y$ have exponential distributions, and $P(X = Y) > 0$. From now on we
call this distribution as the Marshall-Olkin bivariate exponential (MOBE) distribution, and popularly it is also known as the shock model. Since its inception, an extensive amount of work has been done related to this distribution. Several properties have been established, its characterizations, and both classical and Bayesian inferential procedures have been developed, see for example Arnold [2], Baxter and Rachev [8], Boland [14], Muliere and Scarsini [50], Pena and Gupta [54], Ryu [57], the review article by Nadarajah [51], and the references cited therein.

Lu [41] provided the Weibull extension of the MOBE model. Since then quite a bit of work has been done on this and some related distributions mainly developing inference procedures under complete sample and under various sampling schemes, and develop the properties of the order statistics. This model has been used quite successfully to analyze dependent competing risks data. See for example, Begum and Khan [10], Cai et al. [15], Feizjavadian and Hashemi [18], Jose et al. [25], Kundu and Dey [30], Kundu and Gupta [32], Lai et al. [39] and see the references cited therein.

Some of the other bivariate distributions with singular components which can be found in the literature are bivariate Kumaraswamy (BVK), bivariate Pareto (BVP), bivariate double generalized exponential (BDGE), bivariate exponentiated Frechet (BEF), bivariate Gumbel (BVG) distributions, see for example Barreto-Souza and Lemonte [9], bivariate generalized exponential (BGE) model of Kundu and Gupta [36], Sarhan-Balakrishnan’s bivariate (SBB) distribution introduced by Sarhan and Balakrishnan [58], modified Sarhan-Balakrishnan’s bivariate (MSBB) distribution introduced by Kundu and Gupta [37], bivariate model with proportional reversed hazard marginals proposed by Kundu and Gupta [38], bivariate generalized linear failure rate model introduced by Sarhan et al. [59], the generalized Marshall-Olkin bivariate distributions introduced by Gupta et al. [21], bivariate inverse Weibull distribution as proposed by Muhammed [49] and Kundu and Gupta [34] and see the references
cited therein.

In many situations although the data are continuous in nature, say time, pressure etc., they often measured in discrete units. In a situation like this, we often get ties in a bivariate data set. But we will provide few examples where ties occur naturally.

**SHOCK MODEL:** It was originally proposed by Marshall and Olkin [44], and it is considered to be the most classical model of a bivariate distribution with a singular component. Suppose, there are two components of a system, and there are three shocks which can affect the two components. The shocks appear randomly and they affect the systems. The Shock 1 affects the Component 1, Shock 2 affects the Component 2, and Shock 3 affects both the components. The component fails as soon as it receives a shock. The failure times of the both the components are observed as a bivariate random variable. In this case clearly there is a positive probability that the failure times of the two components become equal.

**STRESS MODEL:** It was originally proposed by Kundu and Gupta [36] and it can be described as follows. Suppose a system has two components, and each component is subject to individual stress, say $V_1$ and $V_2$. Other than the individual stresses, the system has a overall stress $V_3$ which has been propagated to both the components equally irrespective of their individual stresses. Therefore, the observed stress at the two components are $X = \max\{V_1, V_3\}$ and $Y = \max\{V_2, V_3\}$, respectively.

**SOCCER MODEL:** Suppose the first component of a bivariate data represents the time of the first kick goal scored by any team, and second component represents the time of the first goal of any type by the home team. In this case also if the first goal is scored by the home team and it is a kick goal, then there is a tie of the two components and it happens with a positive probability.

**MAINTENANCE MODEL:** Suppose a system has two components, say Component 1 and
Component 2, and it is assumed that both components have been maintained independently, and also there is an overall maintenance to both the components. It is assumed that due to component maintenance, suppose the lifetime of Component $i$ is increased by $U_i$ amount, for $i = 1$ and 2, and because of the overall maintenance, the lifetime of each component is increased by $U_3$ amount. Hence, the increased lifetimes of the two components are $X_1 = \max\{U_1, U_3\}$ and $X_2 = \max\{U_2, U_3\}$, respectively, see for example Kundu and Gupta [36] in this respect.

Note that most of the bivariate distributions with singular components are based on two different approaches, namely minimization approach and maximization approach. The main aim of this manuscript is to put both the methods under the same framework. It may be mentioned that any bivariate distribution is characterized by its marginals and the copula function. It is observed that based on the copula many properties can be derived for any class of bivariate distribution functions. We derive some basic properties in both the cases, and it is observed that under certain restrictions they can be generated from very similar copulas. Some specific examples have been provided.

The maximum likelihood estimators (MLEs) of the unknown parameters may not always exist, and even if they exist, they cannot be obtained in explicit forms. One needs to solve a higher dimensional optimization problem to compute the MLEs. To avoid that we have proposed to use this problem as a missing value problem, and we have used a very efficient EM algorithm to compute the MLEs of the unknown parameters. It avoids solving a higher dimensional optimization problem. Moreover, the observed Fisher information matrix also can be obtained quite conveniently at the last step of the EM algorithm, and it can be used to construct confidence intervals of the unknown parameters. The analysis of one real life data set has been performed to see the effectiveness of the EM algorithm.

Finally, we provide few examples, and they cannot be obtained directly using the two
methods which we have mentioned above. In all the cases we have provided explicit expres-
sions of the joint PDF and have mentioned the estimation procedures of the unknown para-
eters in each case. We have further mentioned few bivariate distributions with singular com-
ponents which cannot be obtained by the above two methods, and finally we conclude the paper.

The rest of the paper is organized as follows. In Section 2, we provide some preliminaries and in Section 3 we describe the two main approaches to produce bivariate distribution with a singular component. Some special cases are presented in Section 4 and the MLEs are provided in Section 5. In Section 6 we have provided the analysis of a data set. We have provided examples of few bivariate distributions which cannot be obtained by the proposed methods in Section 7, finally we presented several open problems and conclude the paper in Section 8.

2 Preliminaries

In this section we discuss two important class of distribution functions namely (i) proportional hazard class and (ii) proportional reversed hazard class of distribution functions. We will also discuss briefly about the copula function, and three important class of distribution functions which will be used quite extensively later.

2.1 Proportional Hazard Class

Suppose \( F_B(t; \theta) \) is a distribution with the support on the positive real axis as mentioned before and \( S_B(t, \theta) = 1 - F_B(t; \theta) \) is the corresponding survival function. Let us consider the class of distribution functions which has the survival function of the following form;

\[
S_\text{PHM}(t; \alpha, \theta) = [S_B(t; \theta)]^\alpha; \quad t > 0, \quad (1)
\]
with parameters $\theta$, $\alpha > 0$, and zero otherwise. Here $\theta$ can be vector valued and $S_B(t; \theta)$ is called as the baseline survival function. In this case the class of distribution functions defined by (1) is known as the proportional hazard model (PHM). In this case the PDF of the PHM becomes

$$f_{PHM}(t; \alpha, \theta) = \alpha f_B(t; \theta)[S_B(t; \theta)]^{\alpha-1}; \quad t \geq 0,$$

and zero otherwise. The proportional hazard model was originally proposed by Cox [16] as a regression model in the life-table data analysis. The class of distribution functions defined through the survival function (1) is called the proportional hazard class because if the hazard function of $f_B(t; \theta)$ is

$$h_B(t; \theta) = \frac{f_B(t; \theta)}{S_B(t; \theta)},$$

then the hazard function of $f_{PHM}(t; \alpha, \theta)$ becomes

$$h_{PHM}(t; \alpha, \theta) = \frac{f_{PHM}(t; \alpha, \theta)}{S_{PHM}(t; \alpha, \theta)} = \alpha \frac{f_B(t; \theta)}{S_B(t; \theta)} = \alpha h_B(t; \theta).$$

Hence, in this case the hazard function of any member of the proportional hazard class is proportional to the baseline hazard function. Since the inception of the model by Cox [16], an extensive amount of work has been done related to Cox’s PHMs. Most of the standard statistical books on survival analysis discuss this model in details, see for example Cox and Oakes [17], Therneau and Grambsch [63] and the references cited therein.

### 2.2 Proportional Reversed Hazard Class

Suppose $F_B(t; \theta)$ is a distribution with the support on the positive real axis. Then consider the class of distribution functions of the form

$$F_{PRHM}(t; \alpha, \theta) = [F_B(t; \theta)]^{\alpha}; \quad t > 0,$$

with parameters $\alpha > 0$ and $\theta$ (may be a vector valued), and baseline distribution function $F_B(t; \theta)$. This class of distribution functions is known as the proportional reversed hazard
model (PRHM). If $F_B(t; \theta)$ admits the PDF $f_B(t; \theta)$, then the PRHM has a PDF

$$f_{PRHM}(t; \alpha, \theta) = \alpha [F_B(t; \theta)]^{\alpha-1} f_B(t; \theta); \quad t \geq 0.$$ 

Lehmann [43] first proposed this model in the context of hypotheses testing. It is known as the proportional reversed hazard class because if the baseline distribution function $F_B(t; \theta)$ has the reversed hazard function

$$r_B(t; \theta) = \frac{f_B(t; \theta)}{F_B(t; \theta)},$$

then $F_{PRHM}(t; \alpha, \theta)$ has the reversed hazard function

$$r_{PRHM}(t; \alpha, \theta) = \frac{f_{PRHM}(t; \alpha, \theta)}{F_{PRHM}(t; \alpha, \theta)} = \alpha \frac{f_B(t; \theta)}{F_B(t; \theta)} = \alpha r_B(t; \theta).$$

Hence, the reversed hazard function of any member of the proportional reversed hazard class is proportional to the baseline reversed hazard function. For a detailed discussion on this issue one is referred to Block et al. [13]. An extensive amount of work has been done on different proportional reversed hazard classes, see for example exponentiated Weibull distribution of Mudholkar et al. [48], generalized exponential distribution of Gupta and Kundu [22], exponentiated Rayleigh of Surles and Padgett [62], generalized linear failure rate model of Sarhan and Kundu [60], see also Kundu and Gupta [35] and the references cited therein.

### 2.3 COPULA

The dependence between two random variables, say $X$ and $Y$, is completely described by the joint distribution function $F_{X,Y}(x, y)$. The main idea of separating $F_{X,Y}(x, y)$ in two parts: the one which describes the dependence structure, and the other one which describes the marginal behavior, leads to the concept of copula. To every bivariate distribution function $F_{X,Y}(x, y)$, with continuous marginals $F_X(x)$ and $F_Y(y)$, corresponds to a unique function
$C: [0, 1] \times [0, 1] \to [0, 1]$, called a copula function such that

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)); \quad \text{for } (x, y) \in (-\infty, \infty) \times (-\infty, \infty).$$

Note that $C(u, v)$ is a proper distribution function on $[0, 1] \times [0, 1]$. Moreover, from Sklar’s theorem, see for example Nelsen [53], it follows that if $F_{X,Y}(\cdot, \cdot)$ is a joint distribution function with continuous marginals $F_X(\cdot)$, $F_Y(\cdot)$, and if $F_X^{-1}(\cdot)$, $F_Y^{-1}(\cdot)$ are the inverse functions of $F_X(\cdot)$, $F_Y(\cdot)$, respectively, then there exists a unique copula $C$ in $[0, 1] \times [0, 1]$, such that

$$C(u, v) = F_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)); \quad \text{for } (u, v) \in [0, 1] \times [0, 1].$$

Moreover, if $S_{X,Y}(x, y)$ is the joint survival function of $X$ and $Y$, and $S_X(x)$ and $S_Y(y)$ are survival functions of $X$ and $Y$, respectively, then there exists unique function $\overline{C}: [0, 1] \times [0, 1] \to [0, 1]$, called a (survival) copula function such that

$$S_{X,Y}(x, y) = \overline{C}(S_X(x), S_Y(y)); \quad \text{for } (x, y) \in (-\infty, \infty) \times (-\infty, \infty).$$

In this case

$$\overline{C}(u, v) = S_{X,Y}(S_X^{-1}(u), S_Y^{-1}(v)); \quad \text{for } (u, v) \in [0, 1] \times [0, 1].$$

Moreover,

$$\overline{C}(u, v) = u + v - 1 + C(1-u, 1-v); \quad \text{for } (u, v) \in [0, 1] \times [0, 1].$$

It should be pointed out that the survival copula is also a copula, i.e. $\overline{C}(u, v)$ is also a proper distribution function on $[0, 1] \times [0, 1]$. It is well known that many dependence properties of a bivariate distribution are copula properties, and therefore, can be obtained by studying the corresponding copula. These properties do not depend on the marginals.

### 2.4 Three Important Distributions

In this section we discuss three important distribution functions which will be used quite extensively in our future development.
Exponential Distribution

A random variable $X$ is said to have an exponential distribution with the parameter $\lambda > 0$, if the CDF of $X$ is as follows:

$$F_X(x; \lambda) = P(X \leq x) = 1 - e^{-\lambda x}; \quad x > 0,$$

and zero, otherwise. The corresponding PDF of $X$ becomes

$$f_X(x; \lambda) = \lambda e^{-\lambda x}; \quad x > 0,$$

and zero, otherwise. From now on we will denote it by Exp($\lambda$). The PDF of an exponential distribution is always a decreasing function and it has a constant hazard function for all values of $\lambda$. The exponential distribution is the most used distribution in lifetime data analysis. It has several interesting properties including the lack of memory property and it belongs to the PHM class. Interested readers are referred to Balakrishnan and Basu [5] for a detailed discussions on exponential distribution.

Weibull Distribution

A random variable $X$ is said to have a two-parameter Weibull distribution if it has the following CDF

$$F_X(x; \alpha, \lambda) = 1 - e^{-\lambda x^\alpha}; \quad x > 0,$$

and zero, otherwise. Here, $\alpha > 0$ is called the shape parameter and $\lambda > 0$ as the scale parameter. The corresponding PDF becomes

$$f_X(x; \alpha, \lambda) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha}; \quad x > 0,$$

and zero, otherwise. From now on it will be denoted by WE($\alpha, \lambda$).

A two-parameter Weibull distribution is more flexible than a one-parameter exponential distribution. The shape of the PDF and the hazard function depend on the shape parameter.
The PDF can be a decreasing or an unimodal function if $\alpha \leq 1$ or $\alpha > 1$, respectively. Similarly, for $\alpha \leq 1$, the hazard function is a decreasing function and for $\alpha > 1$ the hazard function is an increasing function. Because of its flexibility it has been used quite extensively in reliability and in survival analysis. It is a PHM. An excellent hand book on Weibull distribution is by Rinne [56]. Interested readers are referred to that hand book for further reading.

**Generalized Exponential Distribution**

A random variable $X$ is said to have a two-parameter Generalized Exponential distribution if it has the following CDF

$$F_X(x; \alpha, \lambda) = (1 - e^{-\lambda x})^\alpha; \quad \text{for } x > 0,$$

and zero, otherwise. Here also $\alpha > 0$ is called the shape parameter and $\lambda > 0$ as the scale parameter. The corresponding PDF becomes

$$f_X(x; \alpha, \lambda) = \alpha \lambda e^{-\lambda x}(1 - e^{-\lambda x})^{\alpha - 1}; \quad \text{for } x > 0,$$

and zero, otherwise. From now on we will denote it by $\text{GE}(\alpha, \lambda)$.

A two-parameter GE distribution was first introduced by Gupta and Kundu [22] as an alternative to the two-parameter Weibull and gamma distribution. Since it has been introduced it has been used quite extensively in analyzing different lifetime data. It may be mentioned that it is a PRHM. Interested readers are refereed to the review article by Nadarajah [52] or a book length treatment by Al-Hussaini and Ahsanullah [1] for different developments on the GE distribution till date.
3 TWO MAIN APPROACHES

In this section we provide the two main approaches namely the minimization and maximization approaches, to construct a bivariate distribution with a singular component. We provide both the methods briefly and discuss several common properties of the general class of distribution functions. It is assumed throughout that $F_B(t; \theta)$ is an absolutely continuous distribution function with the support on the positive real axis and $S_B(t; \theta) = 1 - F_B(t; \theta)$. Moreover, the PDF of $F_B(t; \theta)$ is $f_B(t; \theta)$ for $t > 0$ and zero, otherwise. Here $\theta$ can be vector valued also as mentioned in the previous section.

3.1 MINIMIZATION APPROACH (MODEL 1)

In this section we provide the bivariate distributions with singular components which are based on minimum. Suppose $U_1, U_2$ and $U_3$ are three independent non-negative random variables with survival functions $S_1(t; \alpha_1, \theta) = [S_B(t; \theta)]^{\alpha_1}, S_2(t; \alpha_2, \theta) = [S_B(t; \theta)]^{\alpha_2}, S_3(t; \alpha_3, \theta) = [S_B(t; \theta)]^{\alpha_3}$, respectively for $\alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0$. Now we define a new bivariate random variable $(X, Y)$ as follows;

$$X = \min\{U_1, U_3\} \quad \text{and} \quad Y = \min\{U_2, U_3\}.$$  \hspace{1cm} (4)

Note that although $U_1, U_2$ and $U_3$ are independent, due to presence of $U_3$ in both $X$ and $Y$, $X$ and $Y$ are dependent. We would like to obtain the joint cumulative distribution function (JCDF) and the joint probability density function (JPDF) of $X$ and $Y$. But before that let us observe the following facts. Since $X$ is defined as in (4), the survival function of $X$ becomes

$$P(X > x) = S_X(x; \alpha_1, \alpha_3, \theta) = P(U_1 > x, U_3 > x) = P(U_1 > x)P(U_3 > x) = [S_B(x; \theta)]^{\alpha_1 + \alpha_3}.$$  

Hence, the survival function and the CDF of $X$ depends on $\theta$ and $\alpha_1 + \alpha_3$. Similarly, the survival function of $Y$ becomes $P(Y > y) = [S_B(y; \theta)]^{\alpha_2 + \alpha_3}$. Hence, if $U_1$, $U_2$ and $U_3$
are absolutely continuous random variables, then $X$ and $Y$ are also absolutely continuous random variables. Moreover, in this case

$$P(X = Y) = P(U_3 < U_1, U_3 < U_2)$$

$$= \alpha_3 \int_0^\infty f_B(t; \theta)[S_B(t; \theta)]^{\alpha_3-1}[S_B(t; \theta)]^{\alpha_1}[S_B(x; \theta)]^{\alpha_2} dt$$

$$= \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} \int_0^\infty (\alpha_1 + \alpha_2 + \alpha_3)f_B(t; \theta)[S_B(t; \theta)]^{\alpha_1+\alpha_2+\alpha_3-1}$$

$$= \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} > 0. \quad (5)$$

Hence, (5) indicates that $X = Y$ has a positive probability and for fixed $\alpha_1$ and $\alpha_2$, \lim_{\alpha_3 \to 0} P(X = Y) = 0, \lim_{\alpha_3 \to \infty} P(X = Y) = 1$. Along the same line it can be easily obtained that \[ P(X < Y) = \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3} \quad \text{and} \quad P(Y < X) = \frac{\alpha_2}{\alpha_1 + \alpha_2 + \alpha_3}. \quad (6) \]

Now we will provide the joint survival function of $X$ and $Y$ and also derive the joint PDF of $X$ and $Y$. The joint survival function of $X$ and $Y$ can be written as

$$S_{X,Y}(x, y) = P(X > x, Y > y)$$

$$= P(U_1 > x, U_2 > y, U_3 > \max\{x, y\})$$

$$= [S_B(x, \theta)]^{\alpha_1}[S_B(y, \theta)]^{\alpha_2}[S_B(z, \theta)]^{\alpha_3}, \quad (7)$$

here $z = \max\{x, y\}$. Equivalently, (7) can be written as follows;

$$S_{X,Y}(x, y) = \begin{cases} 
[S_B(x, \theta)]^{\alpha_1}[S_B(y, \theta)]^{\alpha_2+\alpha_3} & \text{if } x < y \\
[S_B(x, \theta)]^{\alpha_1+\alpha_3}[S_B(y, \theta)]^{\alpha_2} & \text{if } x \geq y.
\end{cases} \quad (8)$$

Now we will show that the joint survival function (7) or (8) is not an absolutely continuous survival function. Let us recall that a joint survival function $S(x, y)$ is said to be absolute continuous if there exists a $f(x, y) \geq 0$, such that

$$S(x, y) = \int_x^\infty \int_y^\infty f(u, v) dudv \quad \text{for all } x > 0, y > 0.$$
In that case \(f(x, y)\) can be recovered from \(S(x, y)\) as

\[
f(x, y) = \frac{\partial^2}{\partial x \partial y} S(x, y).
\]

Let us denote \(f(x, y) = \frac{\partial^2}{\partial x \partial y} S_{X,Y}(x, y)\). Hence, from (8),

\[
f(x, y) = \begin{cases} 
\alpha_1(\alpha_2 + \alpha_3)f_B(x, \theta)[S_B(x, \theta)]^{\alpha_1-1}f_B(y, \theta)[S_B(y, \theta)]^{\alpha_2+\alpha_3-1} & \text{if } x < y \\
\alpha_2(\alpha_1 + \alpha_3)f_B(x, \theta)[S_B(x, \theta)]^{\alpha_1+\alpha_3-1}f_B(y, \theta)[S_B(y, \theta)]^{\alpha_2-1} & \text{if } x > y.
\end{cases}
\]

Now it can be easily observed that

\[
\int_0^\infty \int_0^\infty f(x, y)dydx = \int_0^\infty \int_y^\infty f(x, y)dydx + \int_0^\infty \int_x^\infty f(x, y)dydx = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} < 1.
\]

Hence, clearly \(S_{X,Y}(x, y)\) is not an absolutely continuous survival function. Moreover,

\[
\int_0^\infty \int_0^\infty f_{X,Y}(x, y)dydx + P(X = Y) = 1.
\]

Let us denote for \(x > 0\) and \(y > 0\)

\[
f_{ac}(x, y) = \frac{\alpha_1 + \alpha_2 + \alpha_3}{\alpha_1 + \alpha_2} f(x, y).
\]

Clearly, \(f_{ac}(x, y)\) is a bivariate density function on the positive quadrant. Observe that for \(x > 0, y > 0\) and for \(z = \max\{x, y\}\),

\[
S_{X,Y}(x, y) = P(X > x, Y > y) = \int_x^\infty \int_y^\infty f(u, v)dudv + P(X = Y > z)
\]

\[
= \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} \int_x^\infty \int_y^\infty f_{ac}(u, v)dudv + \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} \int_z^\infty f_s(u)du. \tag{10}
\]

Here \(f_{ac}(u, v)\) is same as defined in (9) and

\[
f_s(u) = (\alpha_1 + \alpha_2 + \alpha_3)f_B(u, \theta)[S_B(u, \theta)]^{\alpha_1+\alpha_2+\alpha_3-1},
\]

and it is a probability density function on the positive real axis. Based on (10), for \(x > 0, y > 0\) and for \(z = \max\{x, y\}\), the random variable \((X, Y)\) has the joint PDF \(f_{X,Y}(x, y)\) of the form;

\[
f_{X,Y}(x, y) = \begin{cases} 
\alpha_1(\alpha_2 + \alpha_3)f_B(x, \theta)[S_B(x, \theta)]^{\alpha_1-1}f_B(y, \theta)[S_B(y, \theta)]^{\alpha_2+\alpha_3-1} & \text{if } x < y \\
\alpha_2(\alpha_1 + \alpha_3)f_B(x, \theta)[S_B(x, \theta)]^{\alpha_1+\alpha_3-1}f_B(y, \theta)[S_B(y, \theta)]^{\alpha_2-1} & \text{if } x > y \\
\alpha_3f_B(x, \theta)[S_B(x, \theta)]^{\alpha_1+\alpha_2+\alpha_3-1} & \text{if } x = y.
\end{cases}
\]

(11)
In this case the random variable \((X, Y)\) has an absolute continuous part and a singular part. The function \(f_{X,Y}(x, y)\) is considered to be a density function of \((X, Y)\), if it is understood that the first two terms are densities with respect to a two-dimensional Lebesgue measure and the third term is a density function with respect to a one-dimensional Lebesgue measure, see for example Bemis, Bain and Higgins [11]. It simply means that

\[
P(X > x, Y > y) = \int_x^\infty \int_y^\infty f_{X,Y}(u, v) dv du + \int_{\max\{x,y\}}^\infty f_{X,Y}(v, v) dv.
\]

From (10) it is immediate that

\[
S_{X,Y}(x, y) = \frac{\alpha_1 + \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} S_{ac}(x, y) + \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} S_{si}(x, y), \tag{12}
\]

here \(S_{ac}(x, y)\) is the absolutely continuous part of the survival function,

\[
S_{ac}(x, y) = \int_x^\infty \int_y^\infty f_{ac}(u, v) dudv,
\]

and \(S_{si}(x, y)\) is the singular component of \(S_{X,Y}(x, y)\), and it can be written for \(z = \max\{x, y\}\), as

\[
S_{si}(x, y) = [S_B(z; \theta)]^{\alpha_1 + \alpha_2 + \alpha_3}.
\]

Now we would like to find the copula associated with \(S_{X,Y}(x, y)\). Since \(X\) and \(Y\) have the survival functions as \([S_B(x; \theta)]^{\alpha_1 + \alpha_3}\) and \([S_B(y; \theta)]^{\alpha_2 + \alpha_3}\), respectively, therefore

\[
\overline{C}(u, v) = \begin{cases} u^{\alpha_1 + \alpha_3} v & \text{if } u < v^{\alpha_1 + \alpha_3} \\ u^{\alpha_2 + \alpha_3} v & \text{if } u \geq v^{\alpha_2 + \alpha_3}. \end{cases}
\]

If we write \(\beta = \alpha_3/(\alpha_1 + \alpha_3)\) and \(\delta = \alpha_3/(\alpha_2 + \alpha_3)\), then

\[
\overline{C}(u, v) = \begin{cases} u^{1-\beta} v & \text{if } u^{\beta} v^{\delta} < v^{\beta} \\ u v^{1-\delta} & \text{if } u^{\beta} v^{\delta} \geq v^{\beta}. \end{cases}
\]

If we consider a special case \(\alpha_1 = \alpha_2\), and \(\eta = \alpha_3/(\alpha_1 + \alpha_3) = \alpha_3/(\alpha_2 + \alpha_3)\), then

\[
\overline{C}(u, v) = \begin{cases} u^{1-\eta} v & \text{if } u < v \\ u v^{1-\eta} & \text{if } u \geq v. \end{cases}
\]
3.2 Maximization Approach (Model 2)

In the last section we had provided the bivariate distributions with singular components which are based on minimum and in this section we provide the bivariate distributions with singular components which are based on maximum. Let us assume that $V_1, V_2$ and $V_3$ are three independent non-negative random variables with distribution functions $F_1(t; \beta_1, \lambda) = [F_B(t; \lambda)]^{\beta_1}, F_2(t; \beta_2, \lambda) = [F_B(t; \lambda)]^{\beta_2}$ and $F_3(t; \beta_3, \lambda) = [F_B(t; \lambda)]^{\beta_3}$ respectively, for $\beta_1 > 0, \beta_2 > 0$ and $\beta_3 > 0$. Let us define a new bivariate random variable $(X, Y)$ as follows:

$$X = \max\{V_1, V_3\} \quad \text{and} \quad Y = \max\{V_2, V_3\}.$$  

(13)

In this case also similarly as before, $X$ and $Y$ will be dependent random variables due to the presence of $V_3$. The following results can be easily obtained following the same line as the previous section. The CDFs of $X$ and $Y$ become

$$P(X \leq x) = F_X(x; \beta_1, \beta_3, \lambda) = [F_B(x; \lambda)]^{\beta_1+\beta_3}$$

and

$$P(Y \leq y) = F_Y(x; \beta_2, \beta_3, \lambda) = [F_B(x; \lambda)]^{\beta_2+\beta_3}.$$  

$$P(X = Y) = \frac{\beta_3}{\beta_1+\beta_2+\beta_3}, \quad P(X < Y) = \frac{\beta_1}{\beta_1+\beta_2+\beta_3}, \quad P(X > Y) = \frac{\beta_2}{\beta_1+\beta_2+\beta_3}.$$  

The joint CDF of $X$ and $Y$ for $z = \min\{x, y\}$, becomes

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

$$= P(V_1 \leq x, V_2 \leq y, V_3 \leq \min\{x, y\})$$

$$= [F_B(x, \lambda)]^{\beta_1}[F_B(y, \lambda)]^{\beta_2}[F_B(z, \lambda)]^{\beta_3}$$

$$= \begin{cases} 
[F_B(x, \lambda)]^{\beta_1+\beta_3}[F_B(y, \lambda)]^{\beta_2} & \text{if } x < y \\
[F_B(x, \lambda)]^{\beta_2+\beta_3}[F_B(x, \lambda)]^{\beta_1} & \text{if } x \geq y
\end{cases}.$$  

(14)

Following the same approach as before, the joint PDF of $X$ and $Y$ can be obtained as

$$f_{X,Y}(x,y) = \begin{cases} 
\beta_2(\beta_1 + \beta_3)f_B(x, \lambda)[F_B(x, \lambda)]^{\beta_1+\beta_3-1}f_B(y, \lambda)[F_B(y, \lambda)]^{\beta_2-1} & \text{if } x < y \\
\beta_1(\beta_2 + \beta_3)f_B(x, \theta)[S_B(x, \theta)]^{\beta_1-1}f_B(y, \theta)[F_B(y, \theta)]^{\beta_2+\beta_3-1} & \text{if } x > y \\
\beta_3f_B(x, \lambda)[F_B(x, \lambda)]^{\beta_1+\beta_2+\beta_3-1} & \text{if } x = y
\end{cases}.$$  

(15)
In this case also the joint CDF of the random variable \((X, Y)\) has an absolute continuous part and a singular part and the function (15) is considered to be the joint PDF of \(X\) and \(Y\) in the same sense as before. Moreover, the copula associate with the joint CDF \(F_{X,Y}(x, y)\) becomes

\[
C(u, v) = \begin{cases} 
    uv^{\frac{\beta_3}{\beta_1+\beta_3}} & \text{if } u < v^{\frac{\beta_1+\beta_3}{\beta_1+\beta_3}} \\
    u^{\frac{\beta_1}{\beta_1+\beta_3}} v & \text{if } u \geq v^{\frac{\beta_1+\beta_3}{\beta_1+\beta_3}}
\end{cases}
\] (16)

Therefore, if we write as before that \(\beta = \beta_3/(\beta_1 + \beta_3)\) and \(\delta = \beta_3/(\beta_2 + \beta_3)\), then (16) becomes

\[
C(u, v) = \begin{cases} 
    uv^{1-\delta} & \text{if } u^{\delta} < v^\delta \\
    u^{1-\delta} v & \text{if } u^{\delta} \geq v^\delta
\end{cases}.
\]

Hence, for the special case \(\beta_1 = \beta_2\), and for \(\eta = \beta_3/(\beta_1 + \beta_3) = \beta_3/(\beta_2 + \beta_3)\), the copula \(C(u, v)\) becomes

\[
C(u, v) = \begin{cases} 
    uv^{1-\eta} & \text{if } u < v \\
    u^{1-\eta} v & \text{if } u \geq v
\end{cases}.
\]

4 Some Special Cases

In this section we provide some special cases based on these two approaches. Different special cases have been considered in details in the literature. In this section our main aim is to provide those special cases and mention relevant references associate with those models. We also provide some new bivariate models where more work can be done.

4.1 Model 1

Marshall-Olkin Bivariate Exponential Distribution

Marshall-Olkin bivariate exponential (MOBE) distribution seems to be the most popular bivariate distribution with a singular component and it was originally introduced by Marshall and Olkin [44]. In this case the survival function of the base line distribution namely \(S_B(t, \theta) = e^{-t}\), for \(t > 0\), and zero, otherwise. Hence, \(U_1, U_2\) and \(U_3\) as defined in Section
follow \( \text{Exp}(\alpha_1) \), \( \text{Exp}(\alpha_2) \) and \( \text{Exp}(\alpha_3) \), respectively. Hence, the joint PDF of \( X \) and \( Y \) becomes

\[
f_{X,Y}(x, y) = \begin{cases} 
\alpha_1(\alpha_2 + \alpha_3)e^{-\alpha_1 x}e^{-(\alpha_2 + \alpha_3)y} & \text{if } x < y \\
\alpha_2(\alpha_1 + \alpha_3)e^{-(\alpha_1 + \alpha_3)x}e^{-\alpha_2 y} & \text{if } x > y \\
\alpha_3 e^{-(\alpha_1 + \alpha_2 + \alpha_3)x} & \text{if } x = y,
\end{cases}
\]

and it will be denoted by \( \text{MOBE}(\alpha_1, \alpha_2, \alpha_3) \). The absolute continuous part of the PDF of \( \text{MOBE}(\alpha_1, \alpha_2, \alpha_3) \) for different values of \( \alpha_1, \alpha_2 \) and \( \alpha_3 \) are provided in Figure 1. It is clear from the figures that for all parameter values the maximum occurs at \((0,0)\).

Note that the marginals of the MOBE distribution are exponential distributions. In this case \( X \) and \( Y \) follow \( \text{Exp}(\alpha_1 + \alpha_3) \) and \( \text{Exp}(\alpha_2 + \alpha_3) \), respectively. This model is also known as the shock model since it has been introduced as modeling shocks to a parallel system and it has an interesting connection with the homogeneous Poisson process. Interested readers are referred to the original paper of Marshall and Olkin [44] in this respect. An extensive amount of work has been done dealing with different aspects of the MOBE model. Arnold [2] discussed the existence of the maximum likelihood estimators of the unknown parameters. Bemis, Bain and Higgins [11] and Bhattacharyya and Johnson [12] discussed different properties of the MOBE distribution. Pena and Gupta [54] developed the Bayesian inference of the unknown parameters based on a very flexible Beta-Gamma priors and Karlis [26] provided a very efficient EM algorithm to compute the maximum likelihood estimators of the unknown parameters.

Marshall-Olkin Bivariate Weibull Distribution

It can be seen that the MOBE has exponential marginals, and due to this reason it has some serious limitations. For example if the data indicate that the marginals have unimodal PDFs, then clearly MOBE may not be used. Due to this reason, in the same paper Marshall and Olkin [44] introduced the Marshall Olkin bivariate Weibull (MOBW) model by replacing the exponential distribution with the Weibull distribution. In this case
the base line survival function is a Weibull distribution and it is taken as \( S_B(t, \theta) = e^{-t^\theta} \), for \( t > 0 \), and zero otherwise. Hence, the base line distribution is a Weibull distribution with the shape parameter \( \theta \) and scale parameter 1. Using the same notations as in Section 3.1, it can be easily seen that \( U_1, U_2 \) and \( U_3 \) follow Weibull distribution with the same shape parameter \( \theta \), and having scale parameter \( \alpha_1, \alpha_2 \) and \( \alpha_3 \), respectively. Hence, the joint PDF of \( X \) and \( Y \) becomes

\[
f_{X,Y}(x, y) = \begin{cases} 
\theta^2 \alpha_1 (\alpha_2 + \alpha_3) x^{\theta-1} y^{\theta-1} e^{-\alpha_1 x^\theta} e^{-(\alpha_2 + \alpha_3) y^\theta} & \text{if } x < y \\
\theta^2 \alpha_2 (\alpha_1 + \alpha_3) x^{\theta-1} y^{\theta-1} e^{-\alpha_1 x^\theta} e^{-(\alpha_1 + \alpha_3) y^\theta} & \text{if } x > y \\
\theta \alpha_3 x^{\theta-1} e^{-(\alpha_1 + \alpha_2 + \alpha_3) x^\theta} & \text{if } x = y,
\end{cases}
\] (17)

and it will be denoted by MOBW\((\alpha_1, \alpha_2, \alpha_3, \theta)\). The absolute continuous part of the PDF of MOBW\((\alpha_1, \alpha_2, \alpha_3, \theta)\) for different values of \( \alpha_1, \alpha_2, \alpha_3 \) and \( \theta \) are provided in Figure 2. It is clear from the figures that for all parameter values the maximum occurs at \((0, 0)\) if the common shape parameter \( 0 < \theta \leq 1 \), otherwise it is always unimodal.

Note that the marginals of the MOBW distributions are Weibull distributions with the same shape parameter, namely \( X \) follows WE\((\theta, \alpha_1 + \alpha_3)\) and \( Y \) follows WE\((\theta, \alpha_2 + \alpha_3)\). This model has several real life applications, and it has an interesting connection with the renewal process. An extensive amount of work has been done mainly related to the estimation of the unknown parameters both for classical and Bayesian methods. It may be noted that the maximum likelihood estimators of the unknown parameters cannot be obtained in closed form. It needs solving a four dimensional optimization problem. Kundu and Dey [30] developed a very efficient EM algorithm to compute the maximum likelihood estimators of the unknown parameters which needs solving only one one-dimensional optimization problem. Kundu and Gupta [33] developed a very efficient Bayesian inference of the unknown based on a very flexible priors. Different methods have been evolved for analyzing censored data also. See for example Lu [41, 42]. Recently, Feizjavadian and Hashemi [18] and Shen and Xu [61] used MOBW distribution for analyzing dependent competing risks data. They have developed very efficient EM algorithm to compute the known parameters of the model.
It will be interesting to develop Bayesian inference of the unknown parameters in this case also.

**Weighted Marshall-Olkin Bivariate Exponential Distribution**

Jamalizadeh and Kundu [24] introduced the weighted Marshall-Olkin bivariate exponential distribution as an alternative to the MOBW distribution. It is also a bivariate singular distribution and it can be a very flexible distribution similar to the MOBW distribution. It may be recalled that a random variable $X$ is said to have a weighted exponential (WEE) distribution with parameters $\alpha > 0$ and $\lambda > 0$, if the PDF of $X$ is of the form:

$$f_{\text{WEE}}(x; \alpha, \lambda) = \frac{\alpha + 1}{\alpha} \lambda e^{-\lambda x} \left(1 - e^{-\lambda x}\right); \quad x > 0,$$

and 0, otherwise. The WEE distribution was originally introduced by Gupta and Kundu [23] as an alternative to the two-parameter Weibull, gamma or generalized exponential distributions. The PDFs and the hazard functions of the WEE distribution can take variety of shapes similar to the Weibull, gamma or generalized exponential distributions. The weighted Marshall-Olkin bivariate exponential (BWEE) distribution introduced by Jamalizadeh and Kundu [24] has the following PDF

$$f_{X,Y}(x, y) = \begin{cases} \frac{\alpha + \lambda}{\alpha} \lambda_1 e^{-\lambda_1 x} (\lambda_2 + \lambda_3) e^{-(\lambda_2 + \lambda_3) y} (1 - e^{-\alpha x}) & \text{if } x < y \\ \frac{\alpha + \lambda}{\alpha} (\lambda_1 + \lambda_3) e^{-(\lambda_1 + \lambda_3) x} \lambda_2 e^{-\lambda_2 y} (1 - e^{-\alpha y}) & \text{if } x > y \\ \frac{\alpha + \lambda}{\alpha} \lambda_3 e^{-\lambda x} (1 - e^{-\alpha x}) & \text{if } x = y, \end{cases}$$

for $\lambda = \lambda_1 + \lambda_2 + \lambda_3$, and it will be denoted by BWEE$(\lambda_1, \lambda_2, \lambda_3, \alpha)$. The absolute continuous part of the PDF of BWEE$(\lambda_1, \lambda_2, \lambda_3, \alpha)$ for different values of $\lambda_1, \lambda_2, \lambda_3$ and $\alpha$ are provided in Figure 3. It is clear from the figures that for all parameter values the joint PDF is unimodal.

The marginals of the BWEE distribution are WEE distributions. Jamalizadeh and Kundu [24] established different properties of a BWEE distribution. The maximum likelihood estimators of the unknown parameters cannot be obtained in explicit forms. Efficient EM algorithm has been proposed by Jamalizadeh and Kundu [24] to compute the maximum likelihood estimators of the unknown parameters.
Bivariate Kumaraswamy Distribution

Barreto-Souza and Lemonte [9] considered the bivariate Kumaraswamy (BVK) distribution whose marginals are Kumaraswamy distribution. Since Kumaraswamy distribution has the support on \([0,1]\), the BVK distribution has the support \([0,1] \times [0,1]\). It may be recalled that a random variable \(X\) is said to have a Kumaraswamy distribution with parameters \(\alpha > 0\) and \(\beta > 0\), if it has the following CDF and PDF, respectively

\[
F_K(x; \alpha, \beta) = 1 - (1 - x^\beta)^\alpha \quad \text{and} \quad f_K(x; \alpha, \beta) = \alpha\beta x^{\beta-1}(1 - x^\beta)^{\alpha-1},
\]

for \(x > 0\). Hence, a BVK distribution with parameters \(\alpha_1, \alpha_2, \alpha_3\) and \(\beta\) has the following PDF

\[
f_{X,Y}(x, y) = \begin{cases} 
\alpha_1(\alpha_2 + \alpha_3)\beta^2x^{\beta-1}(1 - x^\beta)^{\alpha_1-1}y^{\beta-1}(1 - y^\beta)^{\alpha_2+\alpha_3-1} & \text{if } x < y \\
(\alpha_1 + \alpha_3)\alpha_2\beta^2x^{\beta-1}(1 - x^\beta)^{\alpha_1+\alpha_3-1}y^{\beta-1}(1 - y^\beta)^{\alpha_2-1} & \text{if } x > y \\
\theta\alpha_3\beta x^{\beta-1}(1 - x^\beta)^{\alpha_1+\alpha_2+\alpha_3-1} & \text{if } x = y,
\end{cases}
\]  

(19)

and it will be denoted by BVK(\(\alpha_1, \alpha_2, \alpha_3, \beta\)).

The PDF of the absolute continuous part of BVK(\(\alpha_1, \alpha_2, \alpha_3, \beta\)) distribution for different \(\alpha_1, \alpha_2, \alpha_3, \beta\) are provided in Figure 4. It is clear that it has bounded support on \([0,1] \times [0,1]\), and it can take variety of shapes depending on the parameter values.

Barreto-Souza and Lemonte [9] developed several properties of the BVK distribution and also provided a very efficient EM algorithm to compute the maximum likelihood estimators of the unknown parameters. They have used this model to analyze a bivariate singular data set with bounded support.

4.2 Model 2

Bivariate Proportional Reversed Hazard Distribution

Kundu and Gupta [36] introduced the bivariate generalized exponential (BVGE) distribution as a stress model and as a maintenance model. The idea is very similar to the MOBE
or the MOBW distribution, but in this case the authors considered the maximization approach than the minimization method. It is assumed that the base line distribution function is an exponential distribution with the scale parameter \( \lambda \), i.e. \( F_B(t; \lambda) = (1 - e^{-\lambda t}) \), for \( t > 0 \) and zero, otherwise. Let us assume that \( V_1, V_2 \) and \( V_3 \) have the CDFs \( (1 - e^{-\lambda t})^{\beta_1}, (1 - e^{-\lambda t})^{\beta_2}, (1 - e^{-\lambda t})^{\beta_3} \) and they are independently distributed. Here \( \beta_1 > 0, \beta_2 > 0 \) and \( \beta_3 > 0 \). It is clear that \( V_1, V_2 \) and \( V_3 \) have \( GE(\beta_1, \lambda), GE(\beta_2, \lambda) \) and \( GE(\beta_3, \lambda) \), respectively. Hence, the joint PDF of \((X, Y)\) in this case becomes

\[
  f_{X,Y}(x, y) = \begin{cases} 
    (\beta_1 + \beta_3)\beta_2(1 - e^{-\lambda x})^{\beta_1+\beta_3-1}(1 - e^{-\lambda y})^{\beta_2-1}e^{-\lambda(x+y)} & \text{if } 0 < x < y < \infty \\
    (\beta_2 + \beta_3)\beta_1(1 - e^{-\lambda x})^{\beta_1-1}(1 - e^{-\lambda y})^{\beta_2+\beta_3-1}e^{-\lambda(x+y)} & \text{if } 0 < y < x < \infty \\
    \beta_3(1 - e^{-\lambda x})^{\beta_1+\beta_2+\beta_3-1}e^{-\lambda x} & \text{if } 0 < y = x < \infty, 
  \end{cases}
\]

and it will be denoted by \( BVGE(\beta_1, \beta_2, \beta_3, \lambda) \). The absolute continuous part of the PDF of \( BVGE(\beta_1, \beta_2, \beta_3, \lambda) \) for different values of \( \beta_1, \beta_2, \beta_3 \) and \( \lambda \) are provided in Figure 5. It is clear that the joint PDF of a BVGE is very similar to the joint PDF of a MOBW distribution for different parameter values. It is clear that if \( 0 < \beta_1 + \beta_3 < 1 \) and \( 0 < \beta_1 + \beta_2 < 1 \), then the maximum occurs at \((0, 0)\), otherwise it is unimodal.

It is observed that the BVGE distribution is also quite flexible like the BVWE distribution, and the marginals of the BVGE distribution follow generalized exponential distributions. Because of the presence of the four parameters, the BVGE distribution can be used quite effectively in analyzing various bivariate data sets. Kundu and Gupta [36] provided a very effective EM algorithm in computing the maximum likelihood estimators of the unknown parameters. It will be interesting to see how the EM algorithm can be modified for analyzing censored data also. No work has been done in developing Bayesian inference of the unknown parameters. It may be mentioned that as MOBW has been used for analyzing dependent competing risks data, BVGE distribution may be used for analyzing dependent complementary risks data, see for example Mondal and Kundu [47] in this respect.

Kundu and Gupta [38] extended the BVGE distribution to a more general class of bi-
variate proportional reversed hazard (BVPRH) distribution. In that paper the authors introduced three other classes of bivariate distributions namely (a) bivariate exponentiated Weibull (BVEW) distribution, (b) bivariate exponentiated Rayleigh (BVER) distribution and (c) bivariate generalized linear failure rate (BVGLF) distribution. The BVEW distribution has been obtained by taking the base line distribution as $F_B(t; \alpha, \lambda) = (1 - e^{-\lambda t^\alpha})$, i.e. a Weibull distribution with the scale parameter $\lambda$ and the shape parameter $\alpha$. The BVER distribution can be obtained by taking the base line distribution as a Rayleigh distribution, i.e. $F_B(t; \lambda) = (1 - e^{-\lambda t^2})$. Similarly, the BVGLF distribution can be obtained by taking $F_B(t; \lambda, \theta) = (1 - e^{-(\lambda t + \theta t^2)})$. Sarhan et al. [59] provided the detailed analysis of the BVGLF distribution. They obtained various properties and developed classical inference of the unknown parameters. It will be interesting to develop Bayesian inferences in all the above cases.

5 Classical Inference

In this section we present the classical inferences of the unknown parameters for both the classes of models. It may be mentioned that Arnold [2] first considered the MLEs of the unknown parameters for the Marshall-Olkin bivariate and multivariate normal distributions. Karlis [26] proposed the EM algorithm to compute the MLEs of the unknown parameters of the MOBE distribution. Kundu and Dey [30] extended the result of Karlis [26] to the case of MOBW distribution. In a subsequent paper, Kundu and Gupta [36] provided an EM algorithm to compute the MLEs of the unknown parameters for the modified Sarhan-Balakrishnan singular bivariate distribution. In this section we provide a general EM algorithm which can be used to compute the MLEs of the unknown parameters of bivariate distributions with singular components which can be obtained either by minimization or maximization approach.
5.1 EM Algorithm: Model 1

In this section we provide the EM algorithm to compute the MLEs of the unknown parameters when the data are coming from the joint PDF (11). It is assumed that we have the following data:

\[ D = \{(x_1, y_1), \ldots, (x_n, y_n)\}. \]

It is assumed that \( \alpha_1, \alpha_2, \alpha_3 \) and \( \theta \) are unknown parameters and the form of \( S_B(t; \cdot) \) is known. It may be mentioned that \( \theta \) may be a vector valued also, but in this case we assume it to be a scalar for simplicity. Our method can be easily modified for the vector valued \( \theta \) also. We further use the following notations:

\[ I_0 = \{i : x_i = y_i\}, \quad I_1 = \{i : x_i < y_i\}, \quad I_2 = \{i : x_i > y_i\}, \quad I = I_0 \cup I_1 \cup I_2, \]

\[ |I_0| = n_0, \quad |I_1| = n_1, \quad |I_2| = n_2, \]

here \( |I_j| \) for \( j = 0, 1, 2 \) denotes the number of elements in the set \( I_j \). The log-likelihood function can be written as

\[
l(\alpha_1, \alpha_2, \alpha_3, \theta) = n_0 \ln \alpha_3 + n_1 \ln \alpha_1 + n_1 \ln(\alpha_2 + \alpha_3) + n_2 \ln \alpha_2 + n_2 \ln(\alpha_1 + \alpha_3) + \sum_{i \in I} \ln f_B(x_i, \theta) + \sum_{i \in I_1} \ln f_B(y_i, \theta) + (\alpha_1 + \alpha_2 + \alpha_3 - 1) \sum_{i \in I_0} \ln S_B(x_i, \theta) + (\alpha_1 - 1) \sum_{i \in I_1} \ln S_B(x_i, \theta) + (\alpha_2 + \alpha_3 - 1) \sum_{i \in I_1} \ln S_B(y_i, \theta) \]

\[
(\alpha_1 + \alpha_3 - 1) \sum_{i \in I_1} \ln S_B(x_i, \theta) + (\alpha_2 - 1) \sum_{i \in I_1} \ln S_B(y_i, \theta). \quad (21)
\]

Hence, the MLEs of the unknown parameters can be obtained by maximizing (21) with respect to the unknown parameters. It is known that even in special cases i.e. in case of exponential and Weibull distribution, see for example Bemis et al. [11] and Kundu and Dey [30] that the MLEs do not exist if \( n_1n_2n_3 = 0 \). If \( n_1n_2n_3 > 0 \), then the MLEs exist, but they cannot be obtained in explicit forms. The MLEs have to be obtained by solving
non-linear equations, and it becomes a non-trivial problem. Note that in case of exponential one needs to solve a three dimensional optimization problem and in case of Weibull it is a four dimensional optimization problem. Due to this reason several approximations and alternative estimators have been proposed in the literature, see for example Arnold [2] and Proschan and Sullo [55] in this respect.

We propose an EM algorithm which can be used to compute the MLEs of the unknown parameters, and it is an extension of the EM algorithm originally proposed by Karlis [26] for the exponential distribution. The basic idea of the proposed EM algorithm is quite simple. It may be observed that if instead of \((X, Y)\), \((U_1, U_2, U_3)\) are known then the MLEs of the unknown parameters can be obtained quite conveniently. Suppose we have the following complete data

\[
\mathcal{D}^c = \{(u_{i1}, u_{i2}, u_{i3}), \ldots, (u_{n1}, u_{n2}, u_{n3})\},
\]

then the log-likelihood function of the complete data can be written as

\[
l_c(\alpha_1, \alpha_2, \alpha_3, \theta) = n \ln \alpha_1 + \alpha_1 \sum_{i=1}^{n} \ln S_B(u_{i1}, \theta) + \sum_{i=1}^{n} (\ln f_B(u_{i1}, \theta) + \ln S_B(u_{i1}, \theta))
\]

\[
+ n \ln \alpha_2 + \alpha_2 \sum_{i=1}^{n} \ln S_B(u_{i2}, \theta) + \sum_{i=1}^{n} (\ln f_B(u_{i2}, \theta) + \ln S_B(u_{i2}, \theta))
\]

\[
+ n \ln \alpha_3 + \alpha_3 \sum_{i=1}^{n} \ln S_B(u_{i3}, \theta) + \sum_{i=1}^{n} (\ln f_B(u_{i3}, \theta) + \ln S_B(u_{i3}, \theta))
\].

Hence, for a fixed \(\theta\), the MLEs of \(\alpha_1, \alpha_2\) and \(\alpha_3\) can be obtained as

\[
\hat{\alpha}_1(\theta) = -\frac{n}{\sum_{i=1}^{n} \ln S_B(u_{i1}, \theta)}, \quad \hat{\alpha}_2(\theta) = -\frac{n}{\sum_{i=1}^{n} \ln S_B(u_{i2}, \theta)}, \quad \hat{\alpha}_3(\theta) = -\frac{n}{\sum_{i=1}^{n} \ln S_B(u_{i3}, \theta)}.
\]

Once, \(\hat{\alpha}_1(\theta), \hat{\alpha}_2(\theta)\) and \(\hat{\alpha}_3(\theta)\) are obtained, the MLE of \(\theta\) can be obtained by maximizing the profile log-likelihood function \(l_c(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \theta)\) with respect to \(\theta\). Therefore, instead of solving a four dimensional optimization problem, one needs to solve only a one dimensional optimization problem in this case. Due to this reason we treat this problem as a missing
value problem. The following table will be useful in identifying the missing $U_i$'s in different cases and the associated probabilities.

<table>
<thead>
<tr>
<th>Different cases</th>
<th>$X &amp; Y$</th>
<th>Set</th>
<th>Value of $U_i$'s</th>
<th>Missing $U_i$'s</th>
<th>Probability of Missing $U_i$'s</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U_3 &lt; \min{U_1, U_2}$</td>
<td>$X = Y$</td>
<td>$I_0$</td>
<td>$U_3 = X = Y$</td>
<td>$U_1 &amp; U_2$</td>
<td>1</td>
</tr>
<tr>
<td>$U_1 &lt; \min{U_2, U_3}$</td>
<td>$X &lt; Y$</td>
<td>$I_1$</td>
<td>$X = U_1, Y = U_2$</td>
<td>$U_3$</td>
<td>$\frac{\alpha_1}{\alpha_2 + \alpha_3}$</td>
</tr>
<tr>
<td>$U_1 &lt; \min{U_2, U_3}$</td>
<td>$X &lt; Y$</td>
<td>$I_1$</td>
<td>$X = U_1, Y = U_3$</td>
<td>$U_2$</td>
<td>$\frac{\alpha_2}{\alpha_2 + \alpha_3}$</td>
</tr>
<tr>
<td>$U_2 &lt; \min{U_1, U_3}$</td>
<td>$Y &lt; X$</td>
<td>$I_2$</td>
<td>$X = U_1, Y = U_2$</td>
<td>$U_3$</td>
<td>$\frac{\alpha_1}{\alpha_1 + \alpha_3}$</td>
</tr>
<tr>
<td>$U_2 &lt; \min{U_1, U_3}$</td>
<td>$Y &lt; X$</td>
<td>$I_2$</td>
<td>$X = U_3, Y = U_2$</td>
<td>$U_1$</td>
<td>$\frac{\alpha_3}{\alpha_1 + \alpha_3}$</td>
</tr>
</tbody>
</table>

Table 1: Different cases and missing $U_i$'s

We need the following derivations for developing the EM algorithm for $i, j = 1, 2, 3$ and $i \neq j$.

\begin{align*}
E(U_j | U_j > u) &= \frac{1}{[S_B(u, \theta)]^{\alpha_j}} \int_u^{\infty} \alpha_j t f_B(t; \theta) [S_B(t; \theta)]^{\alpha_j - 1} dt \quad (25) \\
E(U_j \mid \min\{U_i, U_j\} = u) &= \frac{\alpha_j}{\alpha_i + \alpha_j} u + \frac{\alpha_i}{\alpha_i + \alpha_j} \times \frac{1}{[S_B(u, \theta)]^{\alpha_j}} \int_u^{\infty} \alpha_j t f_B(t; \theta) [S_B(t; \theta)]^{\alpha_j - 1} dt. \quad (26)
\end{align*}

Using (25) and (26), we estimate the missing $U_i$'s by its expected value. The expectation either can be performed by a direct numerical integration or by Monte Carlo simulations. Therefore, the following EM algorithm can be used to compute the MLEs of the unknown parameters in this case. Start with an initial guess of $\alpha_1^{(0)}$, $\alpha_2^{(0)}$, $\alpha_3^{(0)}$ and $\theta^{(0)}$, respectively. We provide the explicit method how the $(k + 1)$-th iterate can be obtained from the $k$-th iterate. Suppose at the $k$-th iterate the values of $\alpha_1$, $\alpha_2$, $\alpha_3$ and $\theta$ are $\alpha_1^{(k)}$, $\alpha_2^{(k)}$, $\alpha_3^{(k)}$ and $\theta^{(k)}$, respectively. Then if a data point $(x_i, x_i) \in I_0$, clearly, $u_{i3} = x_i$, and the missing $u_{i1}$ and $u_{i2}$ can be obtained as

\begin{align*}
u_{i1}^{(k)} &= \frac{1}{[S_B(x_i, \theta^{(k)})]^{\alpha_1^{(k)}}} \int_{x_i}^{\infty} \alpha_1^{(k)} t f_B(t; \theta^{(k)}) [S_B(t; \theta^{(k)})]^{\alpha_1^{(k)} - 1} dt, \quad \text{and} \\
u_{i2}^{(k)} &= \frac{1}{[S_B(x_i, \theta^{(k)})]^{\alpha_2^{(k)}}} \int_{x_i}^{\infty} \alpha_2^{(k)} t f_B(t; \theta^{(k)}) [S_B(t; \theta^{(k)})]^{\alpha_2^{(k)} - 1} dt.
\end{align*}
Similarly, if \((x_i, y_i) \in I_1\), then \(u_{i1} = x_i\) and the missing \(u_{i2}\) and \(u_{i3}\) can be obtained as

\[
\begin{align*}
u_{i2}^{(k)} &= \frac{y_i \alpha_2^{(k)}}{\alpha_2^{(k)} + \alpha_3^{(k)}} + \frac{\alpha_2^{(k)}}{\alpha_2^{(k)} + \alpha_3^{(k)}} \times \frac{1}{[S_B(x_i, \theta^{(k)})]^{\alpha_3^{(k)}}} \int_{y_i}^{\infty} \alpha_2^{(k)} t f_B(t; \theta^{(k)}) [S_B(t; \theta^{(k)})]^{\alpha_3^{(k)} - 1} dt, \\
u_{i3}^{(k)} &= \frac{y_i \alpha_3^{(k)}}{\alpha_2^{(k)} + \alpha_3^{(k)}} + \frac{\alpha_3^{(k)}}{\alpha_2^{(k)} + \alpha_3^{(k)}} \times \frac{1}{[S_B(x_i, \theta^{(k)})]^{\alpha_3^{(k)}}} \int_{y_i}^{\infty} \alpha_3^{(k)} t f_B(t; \theta^{(k)}) [S_B(t; \theta^{(k)})]^{\alpha_3^{(k)} - 1} dt.
\end{align*}
\]

If \((x_i, y_i) \in I_2\), then \(u_{i2} = y_i\) and the missing \(u_{i1}\) and \(u_{i3}\) can be obtained as

\[
\begin{align*}
u_{i1}^{(k)} &= \frac{x_i \alpha_1^{(k)}}{\alpha_1^{(k)} + \alpha_3^{(k)}} + \frac{\alpha_1^{(k)}}{\alpha_1^{(k)} + \alpha_3^{(k)}} \times \frac{1}{[S_B(y_i, \theta^{(k)})]^{\alpha_1^{(k)}}} \int_{x_i}^{\infty} \alpha_1^{(k)} t f_B(t; \theta^{(k)}) [S_B(t; \theta^{(k)})]^{\alpha_1^{(k)} - 1} dt, \\
u_{i3}^{(k)} &= \frac{x_i \alpha_3^{(k)}}{\alpha_1^{(k)} + \alpha_3^{(k)}} + \frac{\alpha_3^{(k)}}{\alpha_1^{(k)} + \alpha_3^{(k)}} \times \frac{1}{[S_B(y_i, \theta^{(k)})]^{\alpha_3^{(k)}}} \int_{x_i}^{\infty} \alpha_3^{(k)} t f_B(t; \theta^{(k)}) [S_B(t; \theta^{(k)})]^{\alpha_3^{(k)} - 1} dt.
\end{align*}
\]

Therefore, we can obtain \(\hat{\alpha}_{1c}^{(k+1)}(\theta)\), \(\hat{\alpha}_{2c}^{(k+1)}(\theta)\) and \(\hat{\alpha}_{3c}^{(k+1)}(\theta)\) from the equation (24) by replacing: (i) \(u_{i1}^{(k)}\) by \(u_{i1}^{(k)}\) if \(i \in I_0 \cup I_2\), (ii) \(u_{i2}^{(k)}\) by \(u_{i2}^{(k)}\) if \(i \in I_0 \cup I_1\) and (iii) \(u_{i3}^{(k)}\) by \(u_{i3}^{(k)}\) if \(i \in I_1 \cup I_2\).

Obtain \(\theta^{(k+1)}\) as

\[
\theta^{(k+1)} = \arg \max_{\theta} l_c(\hat{\alpha}_{1c}^{(k+1)}(\theta), \hat{\alpha}_{2c}^{(k+1)}(\theta), \hat{\alpha}_{3c}^{(k+1)}(\theta), \theta).
\]

Here the function \(l_c(\cdot, \cdot, \cdot, \cdot)\) is the log-likelihood function of the complete data set as defined by (23). Once we obtain \(\theta^{(k+1)}\), \(\alpha_1^{(k+1)}\), \(\alpha_2^{(k+1)}\) and \(\alpha_3^{(k+1)}\) can be obtained as

\[
\alpha_1^{(k+1)} = \alpha_{1c}^{(k+1)}(\theta^{(k+1)}), \quad \alpha_2^{(k+1)} = \alpha_{2c}^{(k+1)}(\theta^{(k+1)}), \quad \alpha_3^{(k+1)} = \alpha_{3c}^{(k+1)}(\theta^{(k+1)}).
\]

The process continues unless convergence takes place.

5.2 EM Algorithm: Model 2

In this section we provide the EM algorithm when the data are coming from a bivariate distribution with the joint PDF (15). In this case we have assumed that the data are of the same form as in the previous case and we have the sets \(I_0\), \(I_1\) and \(I_2\) as defined before. In this case the unknown parameters are \(\beta_1\), \(\beta_2\) and \(\beta_3\) and \(\lambda\). In general \(\lambda\) can be vector valued.
also, but for simplicity it has been assumed to be a scalar valued. If it is assumed that the complete data are of the form:

\[ D^c = \{(v_{11}, v_{12}, v_{13}), \ldots, (v_{n1}, v_{n2}, v_{n3})\}, \]

then the log-likelihood function based on the complete data can be written as

\[
l_c(\beta_1, \beta_2, \beta_3, \lambda) = n \ln \beta_1 + \beta_1 \sum_{i=1}^{n} \ln F_B(v_{i1}, \lambda) + \sum_{i=1}^{n} (\ln f_B(v_{i1}, \theta) + \ln F_B(v_{i1}, \theta))
\]

\[
- n \ln \beta_2 + \beta_2 \sum_{i=1}^{n} \ln F_B(v_{i2}, \theta) + \sum_{i=1}^{n} (\ln f_B(v_{i2}, \theta) + \ln F_B(v_{i2}, \theta))
\]

\[
- n \ln \beta_3 + \beta_3 \sum_{i=1}^{n} \ln F_B(v_{i3}, \theta) + \sum_{i=1}^{n} (\ln f_B(v_{i3}, \theta) + \ln F_B(v_{i3}, \theta)). \tag{27}
\]

Hence, for a fixed \( \lambda \), the MLEs of \( \beta_1, \beta_2 \) and \( \beta_3 \) can be obtained as

\[
\hat{\beta}_1(\lambda) = -\frac{1}{\sum_{i=1}^{n} \ln F_B(v_{i1}, \lambda)}, \quad \hat{\beta}_2(\lambda) = -\frac{1}{\sum_{i=1}^{n} \ln F_B(v_{i2}, \lambda)}, \quad \hat{\beta}_3(\lambda) = -\frac{1}{\sum_{i=1}^{n} \ln F_B(v_{i3}, \lambda)}. \tag{28}
\]

As before, once, \( \hat{\beta}_1(\lambda), \hat{\beta}_2(\lambda) \) and \( \hat{\beta}_3(\lambda) \) are obtained, the MLE of \( \lambda \) can be obtained by maximizing the profile log-likelihood function \( l_c(\hat{\beta}_1(\lambda), \hat{\beta}_2(\lambda), \hat{\beta}_3(\lambda), \lambda) \) with respect to \( \lambda \).

We have a similar table as Table 1 identifying the missing \( V_i \)'s in different cases. Similar to

<table>
<thead>
<tr>
<th>Different cases</th>
<th>( X \ &amp; Y )</th>
<th>Set</th>
<th>Value of ( V_i )'s</th>
<th>Missing ( V_i )'s</th>
<th>Probability of Missing ( V_i )'s</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_3 \ &gt; \max{V_1, V_2} )</td>
<td>( X = Y )</td>
<td>( I_0 )</td>
<td>( V_3 = X = Y )</td>
<td>( V_1 \ &amp; V_2 )</td>
<td>1</td>
</tr>
<tr>
<td>( V_2 \ &gt; \max{V_1, V_3} )</td>
<td>( X &lt; Y )</td>
<td>( I_1 )</td>
<td>( X = V_1, Y = V_2 )</td>
<td>( V_3 )</td>
<td>( \frac{\beta_1}{\beta_1 + \beta_3} )</td>
</tr>
<tr>
<td>( V_2 \ &gt; \max{V_1, V_3} )</td>
<td>( X &lt; Y )</td>
<td>( I_1 )</td>
<td>( X = V_3, Y = V_2 )</td>
<td>( V_1 )</td>
<td>( \frac{\beta_2}{\beta_2 + \beta_3} )</td>
</tr>
<tr>
<td>( V_1 \ &gt; \max{V_2, V_3} )</td>
<td>( Y &lt; X )</td>
<td>( I_2 )</td>
<td>( X = V_1, Y = V_2 )</td>
<td>( V_3 )</td>
<td>( \frac{\beta_1}{\beta_1 + \beta_3} )</td>
</tr>
<tr>
<td>( V_1 \ &gt; \max{V_2, V_3} )</td>
<td>( Y &lt; X )</td>
<td>( I_2 )</td>
<td>( X = V_1, Y = V_3 )</td>
<td>( V_2 )</td>
<td>( \frac{\beta_2}{\beta_2 + \beta_3} )</td>
</tr>
</tbody>
</table>

Table 2: Different cases and missing \( V_i \)'s

(25) and (26), we need the following expressions in this case for developing the EM algorithm for \( i, j = 1, 2, 3 \) and \( i \neq j \).

\[
E(V_j | V_j < u) = \frac{1}{[F_B(u, \lambda)]^{\beta_j}} \int_0^u \beta_j f_B(t; \lambda)[F_B(t; \lambda)]^{\beta_j-1} dt \tag{29}
\]
\[
E(V_j | \max \{V_i, V_j\} = u) = \frac{\beta_j}{\beta_i + \beta_j} u + \frac{\beta_i}{\beta_i + \beta_j} \frac{1}{[F_B(u, \lambda)]^{\beta_j}} \int_0^u \beta_j f_B(t; \lambda) [F_B(t; \lambda)]^{\beta_j - 1} dt.
\]

(30)

Now we provide the explicit method in this case how the \( (k + 1)-\text{th iterate} \) can be obtained from the \( k \)-th iterate. Let us assume that at the \( k \)-th iterate the values of \( \beta_1, \beta_2, \beta_3 \) and \( \lambda \) are \( \beta_1^{(k)}, \beta_2^{(k)}, \beta_3^{(k)} \) and \( \lambda^{(k)} \), respectively. Then if a data point \( (x_i, x_i) \in I_0 \), clearly, \( u_3 = x_i \), and the missing \( v_{i1} \) and \( v_{i2} \) can be obtained as

\[
v_{i1}^{(k)} = \frac{1}{[F_B(x_i, \lambda^{(k)})]^{\beta_1^{(k)}}} \int_0^{x_i} \beta_1^{(k)} f_B(t; \lambda^{(k)}) [F_B(t; \lambda^{(k)})]^{\beta_1^{(k)} - 1} dt, \quad \text{and}
\]

\[
v_{i2}^{(k)} = \frac{1}{[F_B(x_i, \lambda^{(k)})]^{\beta_2^{(k)}}} \int_0^{x_i} \beta_2^{(k)} f_B(t; \lambda^{(k)}) [F_B(t; \lambda^{(k)})]^{\beta_2^{(k)} - 1} dt.
\]

Similarly, if \( (x_i, y_i) \in I_1 \), then \( v_{i2} = y_i \) and the missing \( v_{i1} \) and \( v_{i3} \) can be obtained as

\[
v_{i1}^{(k)} = \frac{x_i \beta_1^{(k)}}{\beta_1^{(k)} + \beta_3^{(k)}} + \frac{\beta_3^{(k)}}{\beta_1^{(k)} + \beta_3^{(k)}} \times \frac{1}{[F_B(x_i, \lambda^{(k)})]^{\beta_3^{(k)}}} \int_0^{x_i} \beta_3^{(k)} f_B(t; \lambda^{(k)}) [F_B(t; \lambda^{(k)})]^{\beta_3^{(k)} - 1} dt,
\]

\[
v_{i3}^{(k)} = \frac{x_i \beta_3^{(k)}}{\beta_1^{(k)} + \beta_3^{(k)}} + \frac{\beta_3^{(k)}}{\beta_1^{(k)} + \beta_3^{(k)}} \times \frac{1}{[F_B(x_i, \lambda^{(k)})]^{\beta_3^{(k)}}} \int_0^{x_i} \beta_3^{(k)} f_B(t; \lambda^{(k)}) [S_B(t; \theta^{(k)})]^{\beta_3^{(k)} - 1} dt.
\]

If \( (x_i, y_i) \in I_2 \), then \( v_{i1} = x_i \) and the missing \( v_{i2} \) and \( v_{i3} \) can be obtained as

\[
v_{i2}^{(k)} = \frac{y_i \beta_2^{(k)}}{\beta_2^{(k)} + \beta_3^{(k)}} + \frac{\beta_3^{(k)}}{\beta_2^{(k)} + \beta_3^{(k)}} \times \frac{1}{[F_B(y_i, \lambda^{(k)})]^{\beta_3^{(k)}}} \int_0^{y_i} \beta_3^{(k)} f_B(t; \lambda^{(k)}) [F_B(t; \lambda^{(k)})]^{\beta_3^{(k)} - 1} dt,
\]

\[
v_{i3}^{(k)} = \frac{y_i \beta_3^{(k)}}{\beta_2^{(k)} + \beta_3^{(k)}} + \frac{\beta_3^{(k)}}{\beta_2^{(k)} + \beta_3^{(k)}} \times \frac{1}{[F_B(y_i, \lambda^{(k)})]^{\beta_3^{(k)}}} \int_0^{y_i} \beta_3^{(k)} f_B(t; \lambda^{(k)}) [F_B(t; \lambda^{(k)})]^{\beta_3^{(k)} - 1} dt.
\]

In this case also similarly as before, we can obtain \( \hat{\beta}^{(k+1)}_{1c}(\lambda), \hat{\beta}^{(k+1)}_{2c}(\lambda) \) and \( \hat{\beta}^{(k+1)}_{3c}(\lambda) \) from the equation (28) by replacing: (i) \( v_{i1} \) by \( v_{i1}^{(k)} \) if \( i \in I_0 \cup I_1 \), (ii) \( v_{i2} \) by \( v_{i2}^{(k)} \) if \( i \in I_0 \cup I_2 \) and (iii) \( v_{i3} \) by \( v_{i3}^{(k)} \) if \( i \in I_1 \cup I_2 \). Obtain \( \lambda^{(k+1)} \) as

\[
\lambda^{(k+1)} = \arg \max \ l_c(\hat{\beta}^{(k+1)}_{1c}(\lambda), \hat{\beta}^{(k+1)}_{2c}(\lambda), \hat{\beta}^{(k+1)}_{3c}(\lambda), \lambda).
\]

Here the function \( l_c(\cdot, \cdot, \cdot, \cdot) \) is the log-likelihood function of the complete data set as defined by (27). Once we obtain \( \lambda^{(k+1)}, \beta_1^{(k+1)}, \beta_2^{(k+1)} \) and \( \beta_3^{(k+1)} \) can be obtained as

\[
\beta_1^{(k+1)} = \hat{\beta}^{(k+1)}_{1c}(\lambda^{(k+1)}), \quad \beta_2^{(k+1)} = \hat{\beta}^{(k+1)}_{2c}(\lambda^{(k+1)}), \quad \beta_3^{(k+1)} = \hat{\beta}^{(k+1)}_{3c}(\lambda^{(k+1)}).
\]
The process continues unless convergence takes place.

In both the cases the associated Fisher information matrix can be easily obtained at the last step of the EM algorithm, and it can be used to construct confidence intervals of the unknown parameters.

6 Data Analysis

In this section we provide the analysis of one data set mainly to show how the proposed EM algorithms work in practice. It is observed that it might be possible to provide some physical justification also in certain cases under some simplified assumptions.

This data set is originally available in Meintanis [46] and it is presented in Table 3. This

<table>
<thead>
<tr>
<th>2005-2006</th>
<th>X</th>
<th>Y</th>
<th>2004-2005</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyon-Real Madrid</td>
<td>26</td>
<td>20</td>
<td>Internazionale-Bremen</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Milan-Fenerbahce</td>
<td>63</td>
<td>18</td>
<td>Real Madrid-Roma</td>
<td>53</td>
<td>39</td>
</tr>
<tr>
<td>Chensea-Anderlecht</td>
<td>19</td>
<td>19</td>
<td>Man. United-Fenerbahce</td>
<td>54</td>
<td>7</td>
</tr>
<tr>
<td>Club Brugge-Juventus</td>
<td>66</td>
<td>85</td>
<td>Bayern-Ajax</td>
<td>51</td>
<td>28</td>
</tr>
<tr>
<td>Fenerbahce-PSV</td>
<td>40</td>
<td>40</td>
<td>Moscos-PSG</td>
<td>76</td>
<td>64</td>
</tr>
<tr>
<td>Internazionale-Rangers</td>
<td>49</td>
<td>49</td>
<td>Barcelona-Shakhtar</td>
<td>64</td>
<td>15</td>
</tr>
<tr>
<td>Panathinaikos-Bremen</td>
<td>8</td>
<td>8</td>
<td>Leverkusen-Roma</td>
<td>26</td>
<td>48</td>
</tr>
<tr>
<td>Ajax-Arsenal</td>
<td>69</td>
<td>71</td>
<td>Arsenal-Panathinaikos</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Man. United-Benfica</td>
<td>39</td>
<td>39</td>
<td>Bayern-M. TelAviv</td>
<td>55</td>
<td>11</td>
</tr>
<tr>
<td>Juventus-Bayern</td>
<td>66</td>
<td>62</td>
<td>Bremen-Internazionale</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Club Brugge-Rapid</td>
<td>25</td>
<td>9</td>
<td>Anderlecht-Valencia</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Olympiacos-Lyon</td>
<td>41</td>
<td>3</td>
<td>Panathinaikos-PSV</td>
<td>44</td>
<td>30</td>
</tr>
<tr>
<td>Internazionale-Porto</td>
<td>16</td>
<td>75</td>
<td>Arsenal-Rosenborg</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>Schalke-PSV</td>
<td>18</td>
<td>18</td>
<td>Liverpool-Olympiacos</td>
<td>27</td>
<td>47</td>
</tr>
<tr>
<td>Barcelona-Bremen</td>
<td>22</td>
<td>14</td>
<td>M. TelAviv-Juventus</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Milan-Schalke</td>
<td>42</td>
<td>42</td>
<td>Bremen-Panathinaikos</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3: UEFA Champion League Data
data set represents the soccer data where at least one goal is scored by the home team and at least one goal is scored directly from a penalty kick, foul kick or any other direct kick by any team. All these direct goals are usually called as kick goals. Here \( X \) represents the time in minutes of the first kick goal scored by any team and \( Y \) represents the time in minutes the first goal of any type scored by the home team. Clearly, in this case all the three possibilities are present namely \( X < Y \), \( X > Y \) and \( X = Y \).

Let us consider three random variables \( U_1, U_2 \) and \( U_3 \) as follows:

\[
U_1 = \text{time in minutes that a first kick goal has been scored by the opponent}
\]
\[
U_2 = \text{time in minutes that a first non-kick goal has been scored by the home team}
\]
\[
U_3 = \text{the time in minutes that a first kick goal has been scored by the home team}
\]

In this case \( X = \min\{U_1, U_3\} \) and \( Y = \min\{U_2, U_3\} \). If it is assumed that \( U_1, U_2 \) and \( U_3 \) are independently distributed, then \((X, Y)\) can be obtained as Model 1. In this case we have used three different \( S_B(t; \theta) \) and analyze the data set. The following \( S_B(t; \theta) \) has been used in this case:

\[
(a) \ S_B(t; \theta) = e^{-t}, \quad (b) \ S_B(t; \theta) = e^{-t^2}, \quad (c) \ S_B(t; \theta) = e^{-t^\theta}.
\]

Note that \( S_B(t; \theta) \) in three different cases are as follows; (a) exponential, (b) Rayleigh, (c) Weibull and the corresponding hazard functions are (a) constant, (b) increasing and (c) increasing or decreasing. We have divided all the data points by 100, and it is not going to make any difference in the inference procedure. We compute the MLEs of the unknown parameters based on the proposed EM algorithm. We start the EM algorithm with some initial guesses, and continue the process. We stop the EM algorithm when the ratio \( |l(k + 1) - l(k)|/l(k)| < 10^{-8} \), here \( l(k) \) denotes the value of the log-likelihood function at the \( k \)-th iterate. Based on the method proposed by Louis [40] it is possible to compute the observed Fisher information matrix at the final step of the EM algorithm. We have
provided the MLEs of the unknown parameters, the associate 95% confidence intervals and
the corresponding log-likelihood values for different cases in Table 4.

<table>
<thead>
<tr>
<th>$S_B(t; \theta)$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\theta$</th>
<th>ll</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e^{-t}$</td>
<td>0.7226</td>
<td>1.6352</td>
<td>1.7676</td>
<td>—</td>
<td>-22.757</td>
</tr>
<tr>
<td></td>
<td>(0.1844, 1.2608)</td>
<td>(0.8877, 2.3826)</td>
<td>(1.0378, 2.4975)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>$e^{-t^2}$</td>
<td>1.7523</td>
<td>3.2157</td>
<td>2.9556</td>
<td>—</td>
<td>-18.342</td>
</tr>
<tr>
<td></td>
<td>(0.5687, 2.3212)</td>
<td>(2.1033, 4.3281)</td>
<td>(1.9967, 3.9145)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>$e^{-t^\theta}$</td>
<td>1.2192</td>
<td>2.8052</td>
<td>2.6927</td>
<td>1.6954</td>
<td>-13.118</td>
</tr>
<tr>
<td></td>
<td>(0.2708, 2.1415)</td>
<td>(2.0921, 3.5184)</td>
<td>(1.5011, 3.8852)</td>
<td>(1.3248, 2.0623)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: MLEs of the different parameters, 95% confidence intervals and associated log-likelihood values

For illustrative purposes, we have used Model 2 also to analyze this data set. We have
used three different $F_B(t; \lambda)$ namely (i) $F_B(t; \lambda) = (1 - e^{-\lambda t})$, (ii) $F_B(t; \lambda) = (1 - e^{-\lambda t^2})$
and (iii) $F_B(t; \lambda_1, \lambda_2) = (1 - e^{-\lambda_1 t^\lambda_2})$. In this case (i) represents exponential, (ii) represents
Rayleigh and (iii) represents Weibull distributions. In this case also we have calculated the
MLEs of the unknown parameters using EM algorithm as mentioned before, and we have
used the same stopping rule as in the previous case here also. The MLEs of the unknown
parameters, the associated 95% confidence intervals and the corresponding log-likelihood values are reported in Table 5. In all the cases it is observed that the EM algorithm converges

<table>
<thead>
<tr>
<th>$F_B(t; \lambda)$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\lambda$</th>
<th>$\alpha$</th>
<th>ll</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 - e^{-\lambda t}$</td>
<td>1.4552</td>
<td>0.4686</td>
<td>1.1703</td>
<td>3.9991</td>
<td>—</td>
<td>-20.592</td>
</tr>
<tr>
<td></td>
<td>(0.6572, 2.2334)</td>
<td>(0.1675, 0.7694)</td>
<td>(0.6512, 1.6894)</td>
<td>(2.8009, 4.9991)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>$1 - e^{-\lambda t^2}$</td>
<td>0.4921</td>
<td>0.1659</td>
<td>0.4101</td>
<td>4.0263</td>
<td>—</td>
<td>-22.737</td>
</tr>
<tr>
<td></td>
<td>(0.1704, 0.7343)</td>
<td>(0.0648, 0.2087)</td>
<td>(0.2466, 0.5748)</td>
<td>(2.5141, 5.5382)</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>$1 - e^{-\lambda t^\alpha}$</td>
<td>1.2071</td>
<td>0.3387</td>
<td>0.4178</td>
<td>3.852</td>
<td>1.1328</td>
<td>-17.236</td>
</tr>
<tr>
<td></td>
<td>(0.5612, 1.8534)</td>
<td>(0.0888, 0.5893)</td>
<td>(0.4187, 1.2695)</td>
<td>(2.7163, 4.9897)</td>
<td>(0.7154, 1.5102)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: MLEs of the different parameters, 95% confidence intervals and associated log-likelihood values

within 25 iterations. It seems EM algorithm works quite well for both the models.
In this section we provide few other bivariate distributions with singular components which are available in the literature, which cannot be obtained as special cases of Model 1 or Model 2. It can be seen that in the first two examples, although the construction remains the same, their final forms are different. Reasons will be clear soon. Efficient estimation procedures using EM algorithm can be developed in both the cases. Marshall and Olkin [45] proposed a method to introduce an extra parameter to an univariate distribution and they have shown how it can be done in case of an exponential or Weibull distribution. They have indicated how it can be done for the bivariate cases also. Kundu and Gupta [33] developed bivariate distributions with singular components based on the idea of Marshall and Olkin [45]. Different properties and efficient estimation procedures have been developed.

7.1 Sarhan-Balakrishnan Bivariate Distribution

Sarhan and Balakrishnan [58] proposed this distribution based on minimization approach but they have not taken the base line distribution as the proportional hazard class. Sarhan and Balakrishnan [58] assumed the distribution functions of $U_1$, $U_2$ and $U_3$, for $t > 0$, as $F_{U_1}(t, \alpha_1) = (1-e^{-t})^{\alpha_1}$, $F_{U_2}(t, \alpha_2) = (1-e^{-t})^{\alpha_2}$ and $F_{U_3}(t; \lambda) = (1-e^{-\lambda t})$, respectively. Consider a bivariate random variable $(X, Y)$, such that $X = \min\{U_1, U_3\}$ and $Y = \min\{U_2, U_3\}$. It may be mentioned that although $U_1$, $U_2$ and $U_3$ belong to proportional reversed hazard class, the authors considered the minimization approach. Due to that reason the joint PDF of $X$ and $Y$ is not a very a compact form. The joint SF and the joint PDF of $X$ and $Y$ also can be derived along the same way, as we have mentioned in the previous section. The joint SF of $X$ and $Y$ in this case becomes

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

and the joint PDF can be obtained as

\[
f_{X,Y}(x, y) = \begin{cases}
  f_1(x, y) & \text{if } 0 < y < x < \infty, \\
  f_2(x, y) & \text{if } 0 < x < y < \infty, \\
  f_0(x) & \text{if } 0 < x = y < \infty.
\end{cases}
\]

Here

\[
\begin{align*}
f_1(x, y) &= \alpha_2 e^{-(\lambda x + y)}(1 - e^{-y})^{\alpha_2-1} (\lambda - \lambda (1 - e^{-x})^{\alpha_1} + \alpha_1 e^{-x} (1 - e^{-x})^{\alpha_1-1}) \\
f_2(x, y) &= \alpha_1 e^{-(\lambda y + x)}(1 - e^{-x})^{\alpha_1-1} (\lambda - \lambda (1 - e^{-y})^{\alpha_2} + \alpha_2 e^{-y} (1 - e^{-y})^{\alpha_2-1}) \\
f_0(x) &= \lambda e^{-\lambda x} (1 - (1 - e^{-x})^{\alpha_1})(1 - (1 - e^{-y})^{\alpha_2}).
\end{align*}
\]

Therefore, the SF and \(X\) and \(Y\) become

\[
S_X(x) = P(X > x) = e^{-\lambda x} (1 - (1 - e^{-x})^{\alpha_1}) \quad \text{and} \quad S_Y(y) = P(Y > y) = e^{-\lambda y} (1 - (1 - e^{-y})^{\alpha_2}),
\]

respectively. It should also be mentioned that the copula associated with the Sarhan-Balakrishnan bivariate distribution cannot be obtained in an explicit form.

### 7.2 Modified Sarhan-Balakrishnan Bivariate Distribution

Kundu and Gupta [37] proposed a modified version of the Sarhan-Balakrishnan bivariate distribution, and the later can be obtained as a special case of the former. The idea is as follows: Suppose \(U_1, U_2\) and \(U_3\) are three independent random variables, such that for \(t > 0\), the CDF of \(U_1, U_2\) and \(U_3\) are \(F_{U_1}(t; \alpha_1, \lambda) = (1 - e^{-\lambda t})^{\alpha_1}\), \(F_{U_2}(t; \alpha_2, \lambda) = (1 - e^{-\lambda t})^{\alpha_2}\) and \(F_{U_3}(t; \alpha_3, \lambda) = (1 - e^{-\lambda t})^{\alpha_3}\), respectively. In this case also consider a bivariate random variable \((X, Y)\), such that \(X = \min\{U_1, U_3\}\) and \(Y = \min\{U_2, U_3\}\). Clearly, the modified Sarhan-Balakrishnan bivariate distribution is more flexible than the Sarhan-Balakrishnan bivariate distribution.
The joint survival function in this case can be obtained as

\[ S_{X,Y}(x, y) = P(X > x, Y > y) = (1 - (1 - e^{-\lambda x})^{\alpha_1})(1 - (1 - e^{-\lambda y})^{\alpha_2})(1 - (1 - e^{-\lambda z})^{\alpha_3}), \]

where \( z = \max\{x, y\} \). Hence the marginal survival function can be obtained as

\[ P(X > x) = (1 - (1 - e^{-\lambda x})^{\alpha_1})(1 - (1 - e^{-\lambda z})^{\alpha_3}) \]
\[ P(Y > y) = (1 - (1 - e^{-\lambda y})^{\alpha_2})(1 - (1 - e^{-\lambda z})^{\alpha_3}). \]

Along the same line as before, the joint PDF of the modified Sarhan-Balakrishnan bivariate distribution can be obtained as

\[
f_{X,Y}(x, y) = \begin{cases} 
  f_1(x, y) & \text{if } 0 < y < x < \infty, \\
  f_2(x, y) & \text{if } 0 < x < y < \infty, \\
  f_0(x) & \text{if } 0 < x = y < \infty,
\end{cases}
\]

where

\[
f_1(x, y) = f(x; \alpha_1, \lambda)[f(y; \alpha_2, \lambda) + f(y; \alpha_3, \lambda)] - f(y; \alpha_2 + \alpha_3, \lambda),
\]
\[
f_2(x, y) = f(y; \alpha_2, \lambda)[f(x; \alpha_1, \lambda) + f(x; \alpha_3, \lambda)] - f(x; \alpha_1 + \alpha_3, \lambda),
\]
\[
f_0(x) = \frac{\alpha_1}{\alpha_1 + \alpha_3} f(x; \alpha_1 + \alpha_3, \lambda) + \frac{\alpha_2}{\alpha_2 + \alpha_3} f(x; \alpha_2 + \alpha_3, \lambda) - \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2 + \alpha_3} f(x; \alpha_1 + \alpha_2 + \alpha_3, \lambda).
\]

Here

\[
f(x; \alpha, \lambda) = \begin{cases} 
  \alpha \lambda e^{-\lambda x}(1 - e^{-\lambda x})^{\alpha - 1} & \text{if } x > 0, \\
  0 & \text{if } x \leq 0.
\end{cases}
\]

In this also the copula associated with the modified Sarhan-Balakrishnan bivariate distribution cannot be obtained in a compact form. Note that this idea can be extended for the proportional reversed hazard classes also. The necessary EM algorithm also can be developed for this model along the same line.
7.3 Bivariate Weibull-Geometric Distribution

As it has been mentioned before, Marshall and Olkin [45] proposed a method to introduce a parameter in a family of distributions. It introduces an extra parameter to a model, hence it brings more flexibility. The model has some interesting physical interpretations also. The basic idea is very simple and it can be defined as follows. Suppose \( \{X_n; n = 1, 2, \ldots\} \) is a sequence of i.i.d. random variables and \( N \) is a geometric random variable with parameter \( 0 < p \leq 1 \) having probability mass function

\[
P(N = n) = p(1 - p)^{n-1}; \quad n = 1, 2, \ldots.
\]

(31)

Here, \( X_n \)’s and \( N \) are independently distributed. Consider the random variables,

\[
Y = \min\{X_1, \ldots, X_N\}.
\]

We denote \( S_X(\cdot) \), \( S_Y(\cdot) \), \( f_X(\cdot) \) and \( f_Y(\cdot) \) as the survival function of \( X \), survival function of \( Y \), PDF of \( X_1 \) and PDF of \( Y \), respectively. The CDF of \( Y \) becomes

\[
S_Y(y) = P(Y > y) = P(X_1 > y, \ldots, X_N > y) = \sum_{n=1}^{\infty} (X_1 > y, \ldots, X_n > y | N = n) P(N = n) = p \sum_{n=1}^{\infty} S_X^n(y)(1-p)^{n-1} = \frac{pS_X(y)}{1 - (1-p)S_X(y)}.
\]

Therefore, the PDF of \( Y \) becomes;

\[
f_Y(y) = \frac{pf_X(y)}{(1 - (1-p)S_X(y))^2}.
\]

If \( X_1 \) is an exponential random variable with mean \( 1/\lambda \), then \( Y \) is called Marshall-Olkin exponential (MOE) random variable. The PDF of a MOE distribution with parameter \( \lambda \) and \( p \), \( \text{MOE}(p, \lambda) \), is

\[
f_Y(y; p, \lambda) = \frac{p\lambda e^{-\lambda y}}{(1 - (1-p)e^{-\lambda y})^2}; \quad y > 0.
\]

(32)
When \( p = 1 \), it becomes an exponential random variable. From the PDF (32), it is clear that MOE is a weighted exponential random variable, where the weight function is

\[
w(y) = \frac{p}{(1 - (1 - p)e^{-\lambda y})^2}; \quad y > 0.
\]

Hence, the two-parameter MOE distribution is more flexible than the one parameter exponential distribution. Similarly, if \( X_1 \) is a Weibull random variable with the shape parameter \( \alpha > 0 \) and scale parameter \( \lambda > 0 \), with the PDF

\[
f_{X}(x; \alpha, \lambda) = \alpha \lambda x^{\alpha-1}e^{-\lambda x^\alpha}; \quad x > 0
\]

then the associated \( Y \) is said to have Marshall-Olkin Weibull (MOWE) distribution. The MOWE has the PDF

\[
f_{Y}(y; p, \alpha, \lambda) = \frac{p \alpha \lambda y^{\alpha-1}e^{-\lambda y^\alpha}}{(1 - (1 - p)e^{-\lambda y^\alpha})^2}.
\]

It is immediate that the MOWE distribution is a weighted Weibull distribution and in this case the weight function is

\[
w(y) = \frac{p}{(1 - (1 - p)e^{-\lambda y^\alpha})^2}; \quad y > 0.
\]

The three-parameter MOWE distribution is more flexible than the two-parameter Weibull distribution. Marshall and Olkin [45] developed several properties of the MOE and MOWE distributions. Although Marshall and Olkin [45] did not develop any estimation procedure, very effective EM algorithm can be developed as it has been obtained in case of the corresponding bivariate model by Kundu and Gupta [33].

In the same paper Marshall and Olkin [45] mentioned how their method can be extended to the multivariate case also, although they did not provide much details. Moreover, the extension to the bivariate or multivariate case may not be unique. Kundu and Gupta [33] first used the method of Marshall and Olkin [45] to extend MOBW distribution and called it as the bivariate Weibull-Geometric (BWG) distribution. It has five parameters and MOBW can be obtained as a special case of BWG distribution. The details are given below.
Suppose \( \{(X_{1n}, X_{2n}); n = 1, 2, \ldots\} \) is a sequence of i.i.d. random variables with parameters \( \alpha_1, \alpha_2, \alpha_3, \theta \) as given in (17) and

\[
Y_1 = \min\{X_{11}, \ldots, X_{1N}\} \quad \text{and} \quad Y_2 = \min\{X_{21}, \ldots, X_{2N}\},
\]

then \((Y_1, Y_2)\) is said to have BWG distribution.

The joint survival function of \((Y_1, Y_2)\) becomes:

\[
P(Y_1 > y_1, Y_2 > y_2) = \sum_{n=1}^{\infty} P^n(X_{11} > y_1, X_{21} > y_2)p(1 - p)^{n-1}
\]

\[
= \frac{pP(X_{11} > y_1, X_{21} > y_2)}{1 - (1 - p)P(X_{11} > y_1, X_{21} > y_2)}
\]

\[
= \left\{ \begin{array}{ll}
pe^{-\alpha_1 y_1^\theta}e^{-(\alpha_2 + \alpha_3)y_2^\theta} & \text{if } y_1 \leq y_2 \\
pe^{-(\alpha_1 + \alpha_3)y_1^\theta}e^{-\alpha_2 y_2^\theta} & \text{if } y_1 > y_2
\end{array} \right.
\]

The joint PDF of \((Y_1, Y_2)\) can be obtained as

\[
g(y_1, y_2) = \left\{ \begin{array}{ll}
g_1(y_1, y_2) & \text{if } y_1 > y_2 \\
g_2(y_1, y_2) & \text{if } y_1 < y_2 \\
g_0(y) & \text{if } y_1 = y_2 = y,
\end{array} \right. \tag{33}
\]

where

\[
g_1(y_1, y_2) = \frac{\theta^2 y_1^{\theta-1}y_2^{\theta-1}(\alpha_3 + \alpha_1)\alpha_2e^{-(\alpha_1 + \alpha_3)y_1^\theta - \alpha_2 y_2^\theta}(1 + (1 - p)e^{-(\alpha_1 + \alpha_3)y_1^\theta - \alpha_2 y_2^\theta})}{(1 - (1 - p)e^{-(\alpha_1 + \alpha_3)y_1^\theta - \alpha_2 y_2^\theta})^3}
\]

\[
g_2(y_1, y_2) = \frac{\theta^2 y_1^{\theta-1}y_2^{\theta-1}(\alpha_3 + \alpha_2)\alpha_1e^{-(\alpha_2 + \alpha_3)y_1^\theta - \alpha_1 y_2^\theta}(1 + (1 - p)e^{-(\alpha_2 + \alpha_3)y_1^\theta - \alpha_1 y_2^\theta})}{(1 - (1 - p)e^{-(\alpha_1 + \alpha_3)y_1^\theta - \alpha_2 y_2^\theta})^3}
\]

\[
g_0(y) = \frac{\theta^2 y_1^{\theta-1}\alpha_3e^{-(\alpha_1 + \alpha_2 + \alpha_3)y^\theta}}{(\alpha_1 + \alpha_2 + \alpha_3)(1 - (1 - p)e^{-(\alpha_1 + \alpha_2 + \alpha_3)y^\theta})^2}
\]

The surface plot of the absolute continuous part of BWG for different parameter values are provided in Figure 6. It is clear from Figure 6 that BWG is more flexible than the MOBW distribution.

From the joint PDF of \((Y_1, Y_2)\) it can be seen that

\[
P(Y_1 = Y_2) = \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3} > 0,
\]

38
hence it has a singular component. The marginals of $Y_1$ and $Y_2$ are geometric-Weibull distribution as defined by Marshall and Olkin [45]. It may be mentioned that when $\theta = 1$, BWG distribution becomes the MOBW distribution. Therefore, MOBW distribution can be obtained as a special case of the BWG distribution. Estimation of the unknown five parameters for the BWG distribution based on a bivariate set of data set is a challenging problem. The MLEs of the unknown parameters can be obtained by solving a five dimensional optimization problem. To avoid that, Kundu and Gupta [33] proposed a very efficient EM algorithm, which involves solving a one dimensional optimization problem at each ‘E’-step of the EM algorithm. In the same paper Kundu and Gupta [33] developed the testing problem so that based of the observed data it should be possible to test whether it is coming from BWG or MOBW model. This model can be used for modeling dependent competing risks data.

7.4 Bivariate PHM-Geometric Distribution

Although Kundu and Gupta [33] proposed the Marshall-Olkin method for the MOBW distribution, it can be extended for a general bivariate PHM as provided in (11). Similar to the bivariate Weibull-geometric model, more general the bivariate PHM-geometric distribution can be obtained as follows: Suppose $\{(X_{1n}, X_{2n}) ; n = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables with parameters $\alpha_1, \alpha_2, \alpha_3$ as follows:

$$S_{X,Y}(x, y) = \begin{cases} [S_B(x)]^{\alpha_1}[S_B(y)]^{\alpha_2+\alpha_3} & \text{if } x < y \\ [S_B(x)]^{\alpha_1+\alpha_3}[S_B(y)]^{\alpha_2} & \text{if } x \geq y. \end{cases}$$

Here $S_B(\cdot)$ is the base line survival function. It may depend on some parameter, but we do not make it explicit. Further, $N$ is a geometric random variable with parameter $0 < p \leq 1$ having probability mass function as given in (31) and it is independent of $\{(X_{1n}, X_{2n}) ; n = 1, 2, \ldots\}$. Consider the random variables,

$$Y_1 = \min\{X_{11}, \ldots, X_{1N}\} \quad \text{and} \quad Y_2 = \min\{X_{21}, \ldots, X_{2N}\},$$
then \((Y_1, Y_2)\) is said to have bivariate PHM-Geometric (PHMG) distribution. The joint survival function of the bivariate PHMG becomes:

\[
P(Y_1 > y_1, Y_2 > y_2) = \sum_{n=1}^{\infty} P^n(X_{11} > y_1, X_{21} > y_2) p(1 - p)^{n-1}
= \frac{pP(X_{11} > y_1, X_{21} > y_2)}{1 - (1 - p)P(X_{11} > y_1, X_{21} > y_2)}
= \left\{ \begin{array}{ll}
p[S_B(y_1)]^{\alpha_1 + \alpha_2} & \text{if } y_1 \leq y_2 \\
p[S_B(y_1)]^{\alpha_1 + \alpha_3} & \text{if } y_1 > y_2.
\end{array} \right.
\]

If we denote \(f_B(\cdot)\) as the base line PDF, then the joint PDF of the bivariate PHMG can be written as

\[
g(y_1, y_2) = \left\{ \begin{array}{ll}
g_1(y_1, y_2) & \text{if } y_1 > y_2 \\
g_2(y_1, y_2) & \text{if } y_1 < y_2 \\
g_0(y) & \text{if } y_1 = y_2 = y,
\end{array} \right.
\]

where

\[
g_1(y_1, y_2) = \frac{c_1 f_B(y_1)[S_B(y_1)]^{\alpha_1 + \alpha_3 - 1} f_B(y_2)[S_B(y_2)]^{\alpha_2 - 1}(1 + (1 - p)[S_B(y_1)]^{\alpha_1 + \alpha_3}[S_B(y_2)]^{\alpha_2})}{(1 + (1 - p)[S_B(y_1)]^{\alpha_1 + \alpha_3}[S_B(y_2)]^{\alpha_2})^3}
\]

\[
g_2(y_1, y_2) = \frac{c_2 f_B(y_1)[S_B(y_1)]^{\alpha_1 - 1} f_B(y_2)[S_B(y_2)]^{\alpha_2 + \alpha_3 - 1}(1 + (1 - p)[S_B(y_1)]^{\alpha_1}[S_B(y_2)]^{\alpha_2 + \alpha_3})}{(1 + (1 - p)[S_B(y_1)]^{\alpha_1}[S_B(y_2)]^{\alpha_2 + \alpha_3})^3}
\]

\[
g_0(y) = \frac{p f_B(y)\alpha_3[S_B(y)]^{\alpha_1 + \alpha_2 + \alpha_3 - 1}}{(\alpha_1 + \alpha_2 + \alpha_3)(1 - (1 - p)[S_B(y)]^{\alpha_1 + \alpha_2 + \alpha_3})^2}
\]

\(c_1 = p(\alpha_3 + \alpha_1)\alpha_2, \ c_2 = p(\alpha_3 + \alpha_2)\alpha_1.\)

Clearly, bivariate PHMG is more flexible than the BWG distribution. When the base line distribution is Weibull with the scale parameter one, and the shape parameter \(\theta\), then the bivariate PHMG becomes the BWG distribution. This model can also be used for analyzing dependent competing risks data. It will be interesting to develop both classical and Bayesian inference procedures for this distribution. More work is needed in this direction.

### 7.5 Bivariate GE-Geometric Distribution

In this section first we introduce Marshall-Olkin GE (MOGE) distribution similar to the MOE and MOWE distribution. The idea is very similar and it can be defined as follows.
Suppose \( \{X_n; n = 1, 2, \ldots\} \) is a sequence of i.i.d. GE\((\alpha, \lambda)\) random variables and \(N\) is a geometric random variable with the PMF as given in (31). Moreover, \(N\) and \(\{X_n; n = 1, 2, \ldots\}\) are independently distributed. Let us define a new random variable \(Y\) as

\[
Y = \max\{X_1, \ldots, X_N\}.
\]

The random variable \(Y\) is said to have GE-Geometric (GEG) distribution with parameters \(p, \alpha, \lambda\), and it will be denoted by GEG\((p, \alpha, \lambda)\).

The CDF of \(Y\) for \(y > 0\) can be obtained as

\[
F_Y(y; p, \alpha, \lambda) = P(Y \leq y) = P(X_1 \leq y, \ldots, X_N \leq y) = p \sum_{n=1}^{\infty} P^n(X_1 \leq y)(1 - p)^{n-1}
\]

\[
= \frac{p(1 - e^{-\lambda x})^{\alpha}}{1 - (1 - p)(1 - e^{-\lambda x})^\alpha}.
\]

The corresponding PDF becomes

\[
f_Y(y; p, \alpha, \lambda) = \frac{p \alpha \lambda e^{-\lambda x}(1 - e^{-\lambda x})^{\alpha-1}}{(1 - (1 - p)(1 - e^{-\lambda x})^\alpha)^2} \quad \text{for} \quad y > 0,
\]

and zero, otherwise.

Now we can define bivariate GEG (BGEG) along the same line as before. Suppose \(\{(X_{1n}, X_{2n}); n = 1, 2, \ldots\}\) is a sequence of i.i.d. random variables with parameters \(\beta_1, \beta_2, \beta_3, \lambda\) having the joint PDF as given in (20) and

\[
Y_1 = \max\{X_{11}, \ldots, X_{1N}\} \quad \text{and} \quad Y_2 = \max\{X_{21}, \ldots, X_{2N}\},
\]

then \((Y_1, Y_2)\) is said to have BGEG distribution with parameters \(\beta_1, \beta_2, \beta_3, \lambda, p\).

The joint CDF of \((Y_1, Y_2)\) becomes:

\[
P(Y_1 \leq y_1, Y_2 \leq y_2) = \sum_{n=1}^{\infty} P^n(X_{11} \leq y_1, X_{21} \leq y_2)p(1 - p)^{n-1}
\]

\[
= \frac{pP(X_{11} \leq y_1, X_{21} \leq y_2)}{1 - (1 - p)P(X_{11} \leq y_1, X_{21} \leq y_2)}
\]

\[
= \begin{cases} 
\frac{p(1-e^{-\lambda y_1})^{\beta_1}(1-e^{-\lambda y_2})^{\beta_2+\beta_3}}{1-(1-p)(1-e^{-\lambda y_1})^{\beta_1}(1-e^{-\lambda y_2})^{\beta_2+\beta_3}} & \text{if } y_1 > y_2 \\
\frac{p(1-e^{-\lambda y_1})^{\beta_1+\beta_3}(1-e^{-\lambda y_2})^{\beta_2}}{1-(1-p)(1-e^{-\lambda y_1})^{\beta_1+\beta_3}(1-e^{-\lambda y_2})^{\beta_2}} & \text{if } y_1 \leq y_2.
\end{cases}
\]
It will be interesting to develop different properties of this distribution similar to the BWG distribution as it has been obtained by Kundu and Gupta [33]. Moreover, classical and Bayesian inference need to be developed for analyzing bivariate data sets with ties. This model also can be used for analyzing dependent complementary risks data. More work is needed along this direction.

8 Conclusions

In this paper we have considered the class of bivariate distributions with a singular component. Marshall and Olkin [44] first introduced the bivariate distribution with singular component based on exponential distributions. Since then an extensive amount of work has taken place in this direction. In this paper we have provided a comprehensive review of all the different models. It is observed that there are mainly two main approaches available in the literature to define bivariate distributions with singular components, and they produce two broad classes of bivariate distributions with singular components. We have shown that these two classes of distributions are related through their copulas under certain restriction. We have provided very general EM algorithms which can be used to compute the MLEs of the unknown parameters in these two cases. We have provided the analysis of one data set to show how the EM algorithms perform in real life.

There are several open problems associated with these models. Although, in this paper we have mainly discussed different methods for bivariate distributions, all the methods can be generalized for the multivariate distribution also. Franco and Vivo [20] first obtained the multivariate Sarhan-Balakrishnan distribution and developed several properties, although no inference procedures were developed. Sarhan et al. [59] provided the multivariate generalized failure rate distribution and Franco, Kundu and Vivo [19] developed multivariate modified Sarhan-Balakrishnan distribution. In both these papers, the authors developed
different properties and very efficient EM algorithms for computing the maximum likelihood estimators of the unknown parameters. More general multivariate distributions with proportional reversed hazard marginals can be found in Kundu, Franco and Vivo [31]. Note that all these models can be very useful to analyze dependent competing risks model with multiple causes. No work has been done along that like, more work is needed along these directions.

References


Figure 1: PDF plots of MOBE($\alpha_1, \alpha_2, \alpha_3$) distribution for different ($\alpha_1, \alpha_2, \alpha_3$) values: (a) (2.0,2.0,1.0) (b) (1.0,1.0,1.0) (c) (1.0,2.0,1.0) (d) (2.0,1.0,1.0).
Figure 2: PDF plots of $\text{MOBW}(\alpha_1, \alpha_2, \alpha_3, \theta)$ distribution for different $(\alpha_1, \alpha_2, \alpha_3, \theta)$ values: (a) $(1.0,1.0,1.0,0.5)$ (b) $(1.0,1.0,1.0,0.75)$ (c) $(2.0,2.0,1.0,1.5)$ (d) $(2.0,2.0,1.0,3.0)$ (e) $(3.0,3.0,3.0,1.5)$ (f) $(3.0,3.0,3.0,3.0)$. 
Figure 3: PDF plots of BWEE($\lambda_1, \lambda_2, \lambda_3, \alpha$) distribution for different ($\lambda_1, \lambda_2, \lambda_3, \alpha$) values: (a) (1.0,1.0,1.0,0.5) (b) (1.0,1.0,2.0,1.0) (c) (1.5,1.5,1.5,1.5) (d) (0.5,0.5,0.5,0.5).
Figure 4: PDF plots of $\text{BVK}(\alpha_1, \alpha_2, \alpha_3, \beta)$ distribution for different $(\alpha_1, \alpha_2, \alpha_3, \beta)$ values: (a) $(1.0,1.0,1.0,0.5)$ (b) $(2.0,2.0,2.0,2.0)$ (c) $(2.0,2.0,2.0,1.0)$ (d) $(2.0,2.0,2.0,3.0)$ (e) $(2.0,2.0,2.0,0.5)$ (f) $(0.75,0.75,0.75,1.1)$. 
Figure 5: PDF plots of MOBE($\alpha_1, \alpha_2, \alpha_3, \theta$) distribution for different ($\alpha_1, \alpha_2, \alpha_3, \theta$) values: (a) (1.0,1.0,1.0,0.5) (b) (1.0,1.0,1.0,0.75) (c) (2.0,2.0,1.0,1.5) (d) (2.0,2.0,1.0,3.0) (e) (3.0,3.0,3.0,1.5) (f) (3.0,3.0,3.0,3.0).
Figure 6: PDF plots of MOBE($\alpha_1, \alpha_2, \alpha_3, \theta, p$) distribution for different ($\alpha_1, \alpha_2, \alpha_3, \theta, p$) values: (a) (2.0,2.0,2.0,2.0,0.5) (b) (2.0,2.0,2.0,3.0,0.5) (c) (2.0,2.0,2.0,2.0,0.8) (d) (2.0,2.0,2.0,3.0,0.8) (e) (2.0,3.0,2.0,2.0,0.5) (f) (2.0,3.0,2.0,2.0,0.8).