# Bivariate Semi-parametric Singular Family of Distributions and its Applications

Debasis Kundu *

**Abstract**

In this paper we introduce a very general class of bivariate semiparametric distributions whose marginals belong to the proportional hazard class, and it has a singular component. This model can be used quite effectively to analyze a bivariate data set when there are ties. Note that the Marshall-Olkin bivariate exponential distribution is a special case of the proposed class. We derive several properties of the proposed distribution, and it is observed that it has a very convenient copula structure. Hence, several dependence properties and dependence measures can be obtained based on the copula. The main feature of the proposed distribution is that we do not use any specific parametric form of the base line distribution, instead we have assumed that the base line distribution has piecewise constant hazard function. It makes the proposed family a very flexible family. The maximum likelihood estimators cannot be obtained in explicit form, and we have used a very convenient EM algorithm to compute the maximum likelihood estimators. Simulation experiments have been performed to see the effectiveness of the proposed EM algorithm. Finally we have used this model to analyze a dependent competing risks data. Two data sets have been analyzed and the results are quite satisfactory.

Key Words and Phrases: Bivariate singular distribution; copula; maximum likelihood estimators; EM algorithm; asymptotic distribution; Lehmann family of distributions

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

We need bivariate distributions to analyze bivariate data sets and to find the association between two variables. An extensive amount of bivariate distributions both discrete and continuous have been proposed in the literature. Among the continuous distributions the most popular one is the bivariate normal distribution. Some of the other well known bivariate continuous distributions are bivariate $t$, bivariate log-normal, bivariate Birnbaum-Saunders, bivariate skew normal distribution, bivariate geometric skew normal distribution etc., see for example the books by Balakrishnan and Lai [3], Kotz et al. [16], the monograph by Azzalini and Capitanio [1] and see the references cited there in. In all these cases the authors introduced absolute continuous bivariate distributions and they cannot be used if it is known that there is a positive probability that the marginals can be equal. There exists several examples where it is known that the two marginals can be equal with a positive probability.

The Marshall-Olkin bivariate exponential (MOBE) distribution, introduced by Marshall and Olkin [18], is the most popular bivariate distribution which has been used in practice when there are ties in a bivariate data set. An extensive amount of work has been done since then related to Marshall-Olkin distribution and several extensions have been proposed in the literature. The main feature of the MOBE distribution is that it has a singular component unlike the absolute continuous bivariate distributions. In both these cases the marginals can be equal with a positive probability. An extensive amount of work has been done in developing different properties and providing efficient estimation of the unknown parameters of the MOBE distribution, see for example Pena and Gupta [20], Karlis [15], Meintanis [19] and the references cited therein.

Let us recall that a class of distribution functions $\{F(x;\theta); \theta > 0\}$, is said to be a proportional hazard class if its survival function $S(x;\theta)$ can be written as

$$S(x;\theta) = (S_0(x))^\theta. \tag{1}$$

Here $S_0(x)$ is known as the base line survival function. Throughout this article it is assumed that $S_0(x) = 0$ for $x \leq 0$, although all the results can be extended for general $S_0$ also. The results are valid even if $S_0(x)$ has a bounded support. From now it will be denoted by PHC($S_0, \theta$). Note that if $f_0(x)$ ($f(x;\theta)$) and $h_0(x)$ ($h(x;\theta)$) denote the probability density function (PDF) and the hazard function (HF) associated with $S_0(x)$ ($S(x;\theta)$), then

$$h(x;\theta) = \frac{f(x;\theta)}{S(x;\theta)} = \theta\frac{f_0(x)}{S_0(x)} = \theta h_0(x).$$

It can be easily seen that the exponential and Weibull families are members of the proportional hazard class. An extensive amount of work has been done in developing several properties and establishing the inference procedures of several univariate proportional hazard classes, see for example Cox and Oakes [9] in this respect.

The main aim of this paper is to develop a very general bivariate distribution with a singular component, and its marginals belong to a proportional hazard class. The MOBE distribution is a member of this family. We do not assume any specific parametric form of the base line distribution. The only assumption has been made that the base line distribution is an absolutely continuous distribution. It is observed that the proposed class of distribution has some physical interpretations also. Under certain restriction it has a very convenient copula structure also. Hence, several dependency measures and dependency properties can be easily established.

We have further developed the inference procedure based on the proposed bivariate distribution. In developing the inference procedure, any specific parametric form of the base line distribution has not been assumed, instead it has been assumed that the base line distribution has a piecewise constant hazard function with a given number of cut points. It makes the base line distribution to be a very flexible distribution and the joint PDF can take variety of shapes. The maximum likelihood estimators (MLEs) cannot be obtained in explicit form. We propose to use the expectation maximization (EM) algorithm to compute the MLEs of the unknown parameters. Simulation experiments have been performed to examine the behavior of the proposed EM algorithm. One data set has been analyzed to see how the proposed model and the inference procedure work in practice, and the results are quite satisfactory.

Analysis of competing risks data plays an important role in the lifetime data analysis. Recently, there is a growing interest in developing dependent competing risk model, see for example Feizjavdian and Hashemi [11], Cai et al. [7], Shen and Xu [24], Samanta and Kundu [21] and Kundu and Mondal [22]. The authors have mainly used the Marshall-Olkin bivariate Weibull and bivariate Weibull-Geometric distributions to analyze dependent competing risks data when an item can fail at a particular point due to more than two causes. In this paper we have explored analyzing dependent competing risks data based on the proposed model. We have developed the inference procedure of the unknown parameters, and analyze one data set to see how the proposed model works in practice.

The rest of the paper is organized as follows. In Section 2 we define the proposed bivariate proportional hazard class (BPHC) and provide several of its properties. The MLEs of the unknown parameters are provided in Section 3, and the simulation results have been presented in Section 4. In Section 5 we present the analysis of a bivariate data set, and its application on analyzing dependent competing risks data is provided in Section 6. Finally we conclude the paper in Section 7.

# 2 Bivariate Proportional Hazard Class

## 2.1 Definition and Interpretation

In this section first we will define the bivariate proportional hazard class (BPHC) of distributions. Suppose $U_1$ follows ($\sim$) PHC($S_0, \theta_1$), $U_2 \sim$ PHC($S_0, \theta_2$), $U_3 \sim$ PHC($S_0, \theta_3$) and they are independently distributed. Define

$$X_1 = \min\{U_1, U_3\} \quad \text{and} \quad X_2 = \min\{U_2, U_3\},$$

then $(X_1, X_2)$ is said to have BPHC with the base line CDF $F_0$ or base line survival function $S_0 = 1 - F_0$, with parameters $\theta_1$, $\theta_2$ and $\theta_3$. From now on it will be denoted by BPHC($S_0, \theta_1, \theta_2, \theta_3$). Before deriving several properties let us first provide some physical interpretation of the BPHC.

Shock Model: It is the classical interpretation as it was originally given by Marshall and Olkin [18] in case MOBE distribution. It is valid in this general case also. Suppose there are two components of a system, and there are three shocks which can affect the components. Suppose the Shock 1 affects Component 1, similarly the Shock 2 affects Component 2, and the Shock 3 affects both the components. As soon as the shock appears the component fails. In this case the lifetime of both the components can be modeled using BPHC.

Soccer Model: In a soccer game a 'kick' goal means a goal which has been scored directly either by penalty kick, free kick, corner kick etc. Suppose the first component of a bivariate data represents the time of the first kick goal scored by any team and the second component provides the time of the first goal scored by the home team. In that case the bivariate data can be modelled using BPHC.

## 2.2 Joint and Marginal Distribution Functions

**Theorem 1:** If $(X_1, X_2) \sim \text{BPHC}(S_0, \theta_1, \theta_2, \theta_3)$ then the joint survival function of $(X_1, X_2)$ is

$$S_{X_1,X_2}(x_1, x_2) = P(X_1 \geq x_1, X_2 \geq x_2) = \begin{cases} (S_0(x_1))^{\theta_1}(S_0(x_2))^{\theta_2+\theta_3} & \text{if} \quad x_1 \leq x_2 \\ (S_0(x_1))^{\theta_1+\theta_3}(S_0(x_2))^{\theta_2} & \text{if} \quad x_1 > x_2. \end{cases}$$

**Proof:** It is trivial. The details are avoided. ∎

The joint PDF of $(X_1, X_2)$ can be obtained as follows.

**Theorem 2:** If $(X_1, X_2) \sim \text{BPHC}(S_0, \theta_1, \theta_2, \theta_3)$ then the joint PDF of $(X_1, X_2)$ is

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} (\theta_2 + \theta_3)\theta_1 f_0(x_1) f_0(x_2)(S_0(x_1))^{\theta_1-1}(S_0(x_2))^{\theta_2+\theta_3-1} & \text{if} \quad x_1 < x_2 \\ (\theta_1 + \theta_3)\theta_2 f_0(x_1) f_0(x_2)(S_0(x_1))^{\theta_1+\theta_3-1}(S_0(x_2))^{\theta_2-1} & \text{if} \quad x_1 > x_2 \\ \frac{\theta_3}{\theta_1+\theta_2+\theta_3} f_0(x_3)(S_0(x_3))^{\theta_1+\theta_2+\theta_3-1} & \text{if} \quad x_1 = x_2 = x_3. \end{cases}$$

**Proof:** The proof can be obtained along the same line as the proof of Theorem 2.2 of Kundu and Gupta [17]. The details are avoided. ∎

It should be mentioned here that the joint PDF does not exist with respect to two dimensional Lebesgue measure as it is not an absolutely continuous distribution function. Here, the joint PDF is with respect to two-dimensional Lebesgue measure on the set $\{(x, y); x \neq y\}$ and with respect to one dimensional Lebesgue measure on the set $\{(x, y); x = y\}$, similar to the MOBE distribution, see for example Bemis et al. [6]. The BPHC has an absolute continuous part and a singular part. The absolute continuous part has the joint PDF

$$f_{ac}(x_1, x_2) = \frac{\theta_1 + \theta_2 + \theta_3}{\theta_1 + \theta_2} \times \begin{cases} (\theta_1 + \theta_3)\theta_2 f_0(x_1) f_0(x_2)(S_0(x_1))^{\theta_1-1}(S_0(x_2))^{\theta_2+\theta_3-1} & \text{if} \quad x_1 \leq x_2 \\ (\theta_2 + \theta_3)\theta_1 f_0(x_1) f_0(x_2)(S_0(x_1))^{\theta_1+\theta_3-1}(S_0(x_2))^{\theta_2-1} & \text{if} \quad x_1 > x_2. \end{cases}$$

The joint PDF of the absolute continuous part is a continuous function on $x_1 \neq x_2$. It is continuous every where if and only if $\theta_1 = \theta_2$. Now we will provide the shape of the joint PDF of $f_{ac}(x_1, x_2)$ under different cases.

**Theorem 3:** Suppose $f_0(x)$, the base line PDF, is a 'smooth' (at least twice differentiable) concave function, and $\lim_{x \to 0^+} f_0(x) = \lim_{x \to \infty} f_0(x) = 0$. Let us define the following sets $C_1 = \{(x_1, x_2); x_1 \geq 0, x_2 \geq 0, x_1 < x_2\}$, $C_2 = \{(x_1, x_2); x_1 \geq 0, x_2 \geq 0, x_1 > x_2\}$, $C_0 = \{(x_1, x_2); x_1 \geq 0, x_2 \geq 0, x_1 = x_2\}$

(a) For $\theta_1 = \theta_2 = \theta \geq 1$, $f_{ac}(x_1, x_2)$ is a continuous unimodal function, and the mode occurs on $C_0$.

(b) If $\theta_1 > \theta_2 > 1$, then $f_{ac}(x_1, x_2)$ is an unimodal function, and the mode occurs on $C_1$.

(c) If $\theta_2 > \theta_1 > 1$, then $f_{ac}(x_1, x_2)$ is an unimodal function, and the mode occurs on $C_2$.

**Proof:**

(a) It is immediate that $f_{ac}(x_1, x_2)$ is continuous on $C_1 \cup C_2$, and since, $f_{ac}(x, x) = \lim_{x_1, x_2 \to x} f_{ac}(x_1, x_2)$, it follows that $f_{ac}(x_1, x_2)$ is continuous on $C_1 \cup C_2 \cup C_0$. Since for all $0 < x_1, x_2 < \infty$,

$$f_{ac}(0,0) = f_{ac}(\infty, \infty) = f_{ac}(x_1, 0) = f_{ac}(x_1, \infty) = f_{ac}(0, x_2) = f_{ac}(\infty, x_2) = 0,$$

$f_{ac}(x_1, x_2)$ has a local maximum in $C_1 \cup C_2 \cup C_0$. Because of the assumption on $f_0(x)$, it follows that $f_0(x)$ has a unique maximum, and suppose that is $x_0$. Now let us consider the two partial derivatives of $\ln f_{ac}(x_1, x_2)$, for $(x_1, x_2) \in C_1$, and they are

$$\frac{\partial}{\partial x_1} \ln f_{ac}(x_1, x_2) = \frac{f_0'(x_1)}{f_0(x_1)} - (\theta - 1) \frac{f_0(x_1)}{S_0(x_1)} \tag{2}$$

$$\frac{\partial}{\partial x_2} \ln f_{ac}(x_1, x_2) = \frac{f_0'(x_2)}{f_0(x_2)} - (\theta + \theta_3 - 1) \frac{f_0(x_2)}{S_0(x_2)}. \tag{3}$$

Since, $\frac{f_0'(x_1)}{f_0(x_1)}$ is a decreasing function and $\frac{f_0(x_1)}{S_0(x_1)}$ is an increasing of $x_1 \in (0, x_0)$, then if there exists a solution, say, $x_1'$, of $\frac{\partial}{\partial x_1} \ln f_{ac}(x_1, x_2) = 0$, then $x_1' < x_0$. Following the same argument, and since $\theta_3 > 0$, it follows that if there exists a solution, say $x_2'$, of $\frac{\partial}{\partial x_2} \ln f_{ac}(x_1, x_2) = 0$, then $x_2' < x_1$. Therefore, $(x_1', x_2') \notin C_1$. Hence, there is no critical point in $C_1$. Along the same line it follows that there is no critical point in $C_2$. Hence, the critical point will be in $C_0$, and it follows that it is unique.

(b) Following the same argument as above it follows that under this assumption there exists unique solutions of (2) and (3). (c) It follows similarly. ∎

**Theorem 4:** If the base line hazard function $h_0(t)$ is a decreasing function in $t$, then $f_{ac}(x_1, x_2)$ is a decreasing function of $x_1$ and $x_2$.

**Proof:** Since $f_{ac}(x_1, x_2)$ can be written in terms of $h_0(x_1)$, $h_0(x_2)$, $S_0(x_1)$ and $S_0(x_2)$, the result follows immediately. ∎

**Theorem 5:** If $(X_1, X_2) \sim \text{BPHC}(F_0, \theta_1, \theta_2, \theta_3)$ then

1. $X_1 \sim \text{PHC}(S_0, \theta_1 + \theta_3)$.
2. $X_2 \sim \text{PHC}(S_0, \theta_2 + \theta_3)$.
3. $\min\{X_1, X_2\} \sim \text{PHC}(S_0, \theta_1 + \theta_2 + \theta_3)$.
4. $X_1|\{X_1 < X_2\} \sim \text{PHC}(S_0, \theta_1 + \theta_2 + \theta_3)$.
5. $X_2|\{X_2 < X_1\} \sim \text{PHC}(S_0, \theta_1 + \theta_2 + \theta_3)$.

**Proof:** The proof is not difficult to obtain. Hence, the details are avoided. ∎

The above results will be used in finding initial estimates of our proposed EM algorithm to compute the maximum likelihood estimators. It can be easily seen that

$$P(X_1 < X_2) = \frac{\theta_1}{\theta_1 + \theta_2 + \theta_3}, \quad P(X_2 < X_1) = \frac{\theta_2}{\theta_1 + \theta_2 + \theta_3}, \quad P(X_1 = X_2) = \frac{\theta_3}{\theta_1 + \theta_2 + \theta_3}.$$

The correlation coefficient of $X_1$ and $X_2$ varies from zero to one. $X_1$ and $X_2$ are independent if $\theta_3 = 0$, and the correlation tends to one as $\theta_3 \to \infty$.

## 2.3 Bivariate Hazard Rates and Some Related Properties

Now we will be discussing about the bivariate failure rate of the proposed BPHC. It may be mentioned that there are several ways of defining the bivariate failure rate. Basu [5] first defined the bivariate failure rate for an absolute continuous bivariate distribution. But his definition cannot be used here as the BPHC is not an absolutely continuous distribution function. Moreover, the definition of Basu may not uniquely define the joint distribution function, which is not a very desirable feature. Hence, we have used the definition of Johnson and Kotz [14] as it can be used when the marginals are absolutely continuous, and moreover the bivariate hazard gradients uniquely define the bivariate joint distribution function.

The hazard gradients of the BPHC are defined as

$$h_1(x_1,x_2) = -\frac{\partial}{\partial x_1}\ln S_{X_1,X_2}(x_1,x_2) \quad \text{and} \quad h_2(x_1,x_2) = -\frac{\partial}{\partial x_2}\ln S_{X_1,X_2}(x_1,x_2).$$

Hence, the hazard gradients of the BPHC are as follows:

$$\begin{aligned} h_1(x_1,x_2) = -\frac{\partial}{\partial x_1}\ln S_{X_1,X_2}(x_1,x_2) &= \begin{cases} \theta_1 h_0(x_1) & \text{if } \; x_1 < x_2 \\ (\theta_1+\theta_3)h_0(x_1) & \text{if } \; x_1 > x_2 \end{cases} \\ h_2(x_1,x_2) = -\frac{\partial}{\partial x_2}\ln S_{X_1,X_2}(x_1,x_2) &= \begin{cases} (\theta_2+\theta_3)h_0(x_2) & \text{if } \; x_1 < x_2 \\ \theta_2 h_0(x_2) & \text{if } \; x_1 > x_2. \end{cases} \end{aligned}$$

Now we will show that the survival function of a BPHC satisfies the total positivity of order two ($TP_2$) property, and it also satisfies some hazard rate ordering properties. Let us recall the following definitions. A function $g : \mathbb{R}^2 \to \mathbb{R}$ is said to have a $TP_2$ property, if for all $\mathbf{x} = (x_1,x_2), \;\; \mathbf{y} = (y_1,y_2) \in \mathbb{R}^2$, $g(\mathbf{x})g(\mathbf{y}) \le g(\mathbf{x}\wedge\mathbf{y})g(\mathbf{x}\vee\mathbf{y})$. Here $\mathbf{x}\wedge\mathbf{y} = (\min\{x_1,y_1\},\min\{x_2,y_2\})$ and $\mathbf{x}\vee\mathbf{y} = (\max\{x_1,y_1\},\max\{x_2,y_2\})$. Let $\mathbf{X}$ and $\mathbf{Y}$ be two bivariate random vectors, with survival functions $S_X$ and $S_Y$, respectively. We say that $\mathbf{X}$ is smaller than $\mathbf{Y}$ in the bivariate hazard rate order (denoted by $\mathbf{X} \le_{\text{hr}} \mathbf{Y}$), if

$$S_X(\mathbf{x})S_Y(\mathbf{y}) \le S_X(\mathbf{x}\wedge\mathbf{y})S_Y(\mathbf{x}\vee\mathbf{y}); \quad \mathbf{x},\mathbf{y} \in \mathbb{R}^2.$$

We say that $\mathbf{X}$ is smaller than $\mathbf{Y}$ in the weak bivariate hazard rate order (denoted by $\mathbf{X} \leq_{\text{whr}} \mathbf{Y}$), if

$$S_X(\mathbf{y})S_Y(\mathbf{x}) \leq S_X(\mathbf{x})S_Y(\mathbf{y}); \quad \text{whenever} \quad \mathbf{x} \leq \mathbf{y}, \quad i.e.\ x_1 \leq y_1, x_2 \leq y_2. \tag{4}$$

We have the following results.

**Theorem 6:**

(a) If $(X_1, X_2) \sim \text{BPHC}(S_0, \theta_1, \theta_2, \theta_3)$, then the survival function of $(X_1, X_2)$ has the $\text{TP}_2$ property.

(b) Suppose $\mathbf{X} = (X_1, X_2) \sim \text{BPHC}(S_0, \theta_1, \theta_2, \theta_3)$ and $\mathbf{Y} = (Y_1, Y_2) \sim \text{BPHC}(S_0, \theta_1, \theta_2, \theta_3')$. If $\theta_3 > \theta_3'$, then $\mathbf{X} \leq_{\text{whr}} \mathbf{Y}$.

(c) Suppose $\mathbf{X} = (X_1, X_2) \sim \text{BPHC}(S_0, \theta_1, \theta_2, \theta_3)$ and $\mathbf{Y} = (Y_1, Y_2) \sim \text{BPHC}(S_0, \theta_1, \theta_2, \theta_3')$. If $\theta_3 > \theta_3'$, then $\mathbf{X} \leq_{\text{hr}} \mathbf{Y}$.

**Proof:**

(a) To prove this we need to show for $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$,

$$S(x_1, x_2)S(y_1, y_2) \leq S(x_1 \vee y_1, x_2 \vee y_2)S(x_1 \wedge y_1, x_2 \wedge y_2). \tag{5}$$

The above inequality (5) can be shown by considering all possible twenty four cases namely $x_1 < x_2 < y_1 < y_2$, $x_1 < x_2 < y_2 < y_1$, and so on.

(b) To prove this we need to show (4). This also can be shown by considering all the twenty four cases as above.

(c) Using (a), (b), and Theorem 2.1 of Hu, Khaledi and Shaked [12], the result follows. ∎

### 2.4 Associated Copula

Now we would like to provide the survival copula associated with the BPHC. Based on the marginal distributions, namely PHC($S_0, \theta_1 + \theta_3$) and PHC($S_0, \theta_2 + \theta_3$), respectively, the BPHC has the following survival copula for $0 \le u, v \le 1$;

$$\overline{C}(u,v) = S\left[S_1^{-1}(u), S_2^{-1}(v)\right] = \begin{cases} u^{\frac{\theta_1}{\theta_1+\theta_3}} v & \text{if} \quad u < v^{\frac{\theta_1+\theta_3}{\theta_2+\theta_3}} \\ u v^{\frac{\theta_2}{\theta_1+\theta_3}} & \text{if} \quad u \ge v^{\frac{\theta_1+\theta_3}{\theta_2+\theta_3}}. \end{cases}$$

If we denote $\beta = \dfrac{\theta_3}{\theta_1 + \theta_3}$ and $\delta = \dfrac{\theta_3}{\theta_2 + \theta_3}$, then

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

Further, in the symmetric case, i.e. when $\theta_2 = \theta_3$, if we denote $\gamma = \dfrac{\theta_3}{\theta_1 + \theta_3} = \dfrac{\theta_3}{\theta_2 + \theta_3}$,

$$\overline{C}(u,v) = \begin{cases} u^{1-\gamma} v & \text{if} \quad u < v \\ u v^{1-\gamma} & \text{if} \quad u \ge v. \end{cases} \tag{6}$$

Note that (6) is the famous Marshall-Olkin copula. Based on this copula several dependence properties and dependence measures can be easily established. The above copula is the famous Marshall-Olkin copula, see for example Nelsen. Therefore, it follows that the Kendall's $\tau$ and Spearman's $\rho$ become as follows:

$$\tau = \frac{\theta_1 \theta_2}{\theta_1 - \theta_1 \theta_2 + \theta_2} \quad \text{and} \quad \rho = \frac{3\theta_1 \theta_2}{2\theta_1 - \theta_1 \theta_2 + 2\theta_2}.$$

# 3 Maximum Likelihood Estimators

### 3.1 Log-Likelihood Function

In this section we will be discussing about the maximum likelihood estimators of the unknown parameters in a fairly general set up. So far many authors have discussed about the maximum likelihood estimation of the unknown parameters under specific assumption of the base line distributions. For example Karlis [15], Barreto-Souza and Lemonte [4] and the references cited therein. In this case we would also like to discuss the maximum likelihood estimation of the unknown parameters without any specific assumption on the base line distribution. Instead, we have assumed a fairly general class of distribution. It is assumed that the base line hazard function $h_0(t)$ is piecewise constant, i.e. it is based on the following: the number of constants, $M$ and cut points $\tau_0 < \tau_1 < \ldots < \tau_{M-1} < \tau_M$ to be used. They are assumed to be known. Further, $h_0(t)$ has the form

$$h_0(t) = \sum_{j=1}^{M} c_j I_{[\tau_{j-1},\tau_j]}(t), \tag{7}$$

here $c_j$'s are unknown constants, and

$$I_{[\tau_{j-1},\tau_j]}(t) = \begin{cases} 1 & t \in [\tau_{j-1}, \tau_j] \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that under this set up, $c_1, \ldots, c_M$ are identifiable only up to a multiplicative constant. Hence, without loss of generality it is assumed that $c_M = 1$. Note that when $M = 1$, it becomes an exponential model. Hence, the proposed BPHC coincides with the MOBE distribution. The assumption of piecewise constant hazard function is not very uncommon in the statistical literature. It is has been attempted before by different authors, see for example Ibrahim et al. [13], Balakrishnan et al. [2] and see the references cited therein. Now let us compute $S_0(t)$ and $f_0(t)$ which will be needed later. Note that

$$H_0(t) = \sum_{j=1}^{M} c_j(\min\{t, \tau_j\} - \tau_{j-1}) I_{[\tau_{j-1},\tau_M]}(t).$$

Once $H_0(t)$ is obtained, $S_0(t)$ and $f_0(t)$ can be obtained from the following formulas

$$S_0(t) = e^{-H_0(t)} \quad \text{and} \quad f_0(t) = h_0(t)e^{-H_0(t)}.$$

It is assumed that we have a random sample of size $n$ namely

$$\mathcal{D}_1 = \{(x_1, y_1), \ldots, (x_n, y_n)\},$$

from a BPHC$(S_0, \theta_1, \theta_2, \theta_3)$. The problem is to estimate $\theta_1, \theta_2, \theta_3$ and $c_1, \ldots, c_M$, assuming $\tau_0, \ldots, \tau_M$ to be known. Let us use the following notation:

$I_1 = \{i : 1 \leq i \leq n, x_i < y_i\}$, $I_2 = \{i : 1 \leq i \leq n, x_i > y_i\}$, $I_3 = \{i : 1 \leq i \leq n, x_i = y_i = z_i\}$.

Further $|I_j| = n_j$, where $|I_j|$ denotes the number of elements in the set $I_j$, for $j = 1, 2, 3$, and $I = I_1 \cup I_2 \cup I_3$. We would like to compute the maximum likelihood estimates of the unknown parameters. First we provide the log-likelihood contribution of $(x, y)$ in different cases.

**Case 1:** $x < y$

Suppose $\tau_{m-1} \leq x < \tau_m$ and $\tau_{k-1} \leq y < \tau_k$.

$$\ln(\theta_2 + \theta_3) + \ln\theta_1 + \ln f_0(x) + \ln f_0(y) + (\theta_1 - 1)\ln S_0(x) + (\theta_2 + \theta_3 - 1)\ln S_0(y) =$$

$$\ln(\theta_2 + \theta_3) + \ln\theta_1 + \ln h_0(x) + \ln h_0(y) - \theta_1 H_0(x) - (\theta_2 + \theta_3)H_0(y) =$$

$$\ln(\theta_2 + \theta_3) + \ln\theta_1 + \ln c_m + \ln c_k - \theta_1\{\sum_{j=1}^{m-1} c_j(\tau_j - \tau_{j-1}) + c_m(x - \tau_{m-1})\} -$$

$$(\theta_2 + \theta_3)\{\sum_{j=1}^{k-1} c_j(\tau_j - \tau_{j-1}) + c_k(y - \tau_{k-1})\}$$

**Case 2:** $y < x$

Suppose $\tau_{m-1} \leq x < \tau_m$ and $\tau_{k-1} \leq y < \tau_k$.

$$\ln\theta_2 + \ln(\theta_1 + \theta_3) + \ln f_0(x) + \ln f_0(y) + (\theta_1 + \theta_3 - 1)\ln S_0(x) + \theta_2 - 1)\ln S_0(y) =$$

$$\ln\theta_2 + \ln(\theta_1 + \theta_3) + \ln h_0(x) + \ln h_0(y) - (\theta_1 + \theta_3)H_0(x) - \theta_2 H_0(y) =$$

$$\ln\theta_2 + \ln(\theta_1 + \theta_3) + \ln c_m + \ln c_k - (\theta_1 + \theta_3)\{\sum_{j=1}^{m-1} c_j(\tau_j - \tau_{j-1}) + c_m(x - \tau_{m-1})\} -$$

$$\theta_2\{\sum_{j=1}^{k-1} c_j(\tau_j - \tau_{j-1}) + c_k(y - \tau_{k-1})\}$$

**Case 3:** $x = y = z$

Suppose $\tau_{m-1} \leq z < \tau_m$.

$$\ln\theta_3 - \ln(\theta_1+\theta_2+\theta_3) + \ln f_0(z) + (\theta_1+\theta_2+\theta_3-1)\ln S_0(z) =$$

$$\ln\theta_3 - \ln(\theta_1+\theta_2+\theta_3) + \ln h_0(z) - (\theta_1+\theta_2+\theta_3)H_0(z) =$$

$$\ln\theta_3 - \ln(\theta_1+\theta_2+\theta_3) + \ln c_m - (\theta_1+\theta_2+\theta_3)\{\sum_{j=1}^{m-1} c_j(\tau_j-\tau_{j-1}) + c_m(x-\tau_{m-1})\}.$$

Based on the above three cases we are in a position to write the log-likelihood function of the observed data $\mathcal{D}_1$. For convenience we express $x_i$'s, $y_i$'s and $z_i$'s in the following manner. Any $x_i$, for $i \in I_1 \cup I_2$ will be denoted $x_{uvw}$, here $u = j$ if $i \in I_j$, for $j = 1, 2$. The index $v$ indicates $\tau_{v-1} < x_i \leq \tau_v$ , for $v = 1, \ldots, M$, and $w \in \{1, \ldots, n_{1uv}\}$, where $n_{1uv}$ is the number of $x_i$'s in $(\tau_{v-1}, \tau_v]$ for $i \in I_u$. Similarly, $y_i$ for $i \in I_1 \cup I_2$ will be denoted by $y_{uvw}$. As in $x_i$'s the index $v$ indicates $\tau_{v-1} < y_i \leq \tau_v$ , for $v = 1, \ldots, M$, and $w \in \{1, \ldots, n_{2uv}\}$, where $n_{2uv}$ is the number of $y_i$'s in $(\tau_{v-1}, \tau_v]$ for $i \in I_u$. Finally, any $z_i$, for $i \in I_3$ will be denoted by $z_{vw}$. Here also, the index $v$ indicates $\tau_{v-1} < z_i \leq \tau_v$ , for $v = 1, \ldots, M$, and $w \in \{1, \ldots, n_{3v}\}$, where $n_{3v}$ is the number of $z_i$'s in $(\tau_{v-1}, \tau_v]$ for $i \in I_3$. If we denote $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ and $\boldsymbol{c} = (c_1, \ldots, c_{M-1})$, $n_{\cdot v} = n_{11v} + n_{12v} + n_{21v} + n_{22v} + n_{3v}$, $n_{1\cdot v} = n_{11v} + n_{12v}$, $n_{2\cdot v} = n_{21v} + n_{22v}$ the log-likelihood function of the observed data becomes

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{c}|\mathcal{D}_1) &= n_1\ln(\theta_2+\theta_3) + n_1\ln\theta_1 + n_2\ln(\theta_1+\theta_3) + n_2\ln\theta_2 + n_3\ln\theta_3 - n_3\ln(\theta_1+\theta_2+\theta_3) \\ &\quad + \sum_{v=1}^{M} n_{\cdot v}\ln c_v - \theta_1 A_1(\boldsymbol{c}|\mathcal{D}_1) - \theta_2 A_2(\boldsymbol{c}|\mathcal{D}_1) - \theta_3 A_3(\boldsymbol{c}|\mathcal{D}_1), \end{aligned} \tag{8}$$

where

$$\begin{aligned} A_1(\boldsymbol{c}|\mathcal{D}_1) &= \left\{\sum_{j=1}^{M-1}\sum_{v=1+j}^{M} c_j(n_{1\cdot v}+n_{3v})(\tau_j-\tau_{j-1}) + \sum_{j=1}^{M}\sum_{u=1}^{2}\sum_{w=1}^{n_{1uj}} c_j(x_{ujw}-\tau_{j-1}) + \sum_{j=1}^{M}\sum_{w=1}^{n_j} c_j(z_{jw}-\tau_{j-1})\right\} \\ A_2(\boldsymbol{c}|\mathcal{D}_1) &= \left\{\sum_{j=1}^{M-1}\sum_{v=1+j}^{M} c_j(n_{2\cdot v}+n_{3v})(\tau_j-\tau_{j-1}) + \sum_{u=1}^{2}\sum_{j=1}^{M}\sum_{w=1}^{n_{2uj}} c_j(y_{ujw}-\tau_{j-1}) + \sum_{j=1}^{M}\sum_{w=1}^{n_j} c_j(z_{jw}-\tau_{j-1})\right\} \\ A_3(\boldsymbol{c}|\mathcal{D}_1) &= \left\{\sum_{j=1}^{M-1}\sum_{v=1+j}^{M} c_j(n_{12v}+n_{21v})(\tau_j-\tau_{j-1}) + \sum_{j=1}^{M}\sum_{w=1}^{n_{12j}} c_j(x_{2jw}-\tau_{j-1}) + \sum_{j=1}^{M}\sum_{w=1}^{n_{21j}} c_j(y_{1jw}-\tau_{j-1})\right. \\ &\qquad \left.\sum_{j=1}^{M-1}\sum_{v=1+j}^{M} c_j n_{3v}(\tau_j-\tau_{j-1}) + \sum_{j=1}^{M}\sum_{w=1}^{n_j} c_j(z_{jw}-\tau_{j-1})\right\} \end{aligned}$$

Therefore, it is clear that the MLEs of the unknown parameters can be obtained by solving $M+2$ dimensional optimization problem.

## 3.2 EM Algorithm

In practice it has been seen that with $M = 1$, 2 and 3, the method works quite well, see for example Balakrishnan et al. [2]. Hence, one needs to solve a 3, 4 or 5 dimensional optimization problem to compute the MLEs. To avoid that we treat this problem as a missing value problem, and we propose to use a very efficient EM algorithm. In this case one needs to solve only a $M$ dimensional optimization problem at each 'E'-step of the EM algorithm.

For each observation $(X_1, X_2)$, it is assumed that there exists an associated random vector $(\Delta_1, \Delta_2)$, which takes the following values:

$$(\Delta_1, \Delta_2) = \begin{cases} (1,2) & \text{if} \quad X_1 = U_1, X_2 = U_2 \\ (1,3) & \text{if} \quad X_1 = U_1, X_2 = U_3 \\ (3,2) & \text{if} \quad X_1 = U_3, X_2 = U_2 \\ (3,3) & \text{if} \quad X_1 = U_3, X_2 = U_3. \end{cases}$$

If $X_1 < X_2$, then $(\Delta_1, \Delta_2)$ is either (1,2) or (1,3). Similarly, if $X_2 < X_1$, then $(\Delta_1, \Delta_2)$ is either (1,2) or (3,2). Further, if $X_1 = X_2$, then $(\Delta_1, \Delta_2) =$ (3,3). Hence, if $X_1 < X_2$, in the observed data $\Delta_2$ is missing. Similarly, if $X_2 < X_1$, $\Delta_1$ is missing. Also observe that

$$\begin{aligned} P(\Delta_1 = 1, \Delta_2 = 2 | X_1 < X_2) &= \frac{\theta_2}{\theta_2 + \theta_3} \\ P(\Delta_1 = 1, \Delta_2 = 3 | X_1 < X_2) &= \frac{\theta_3}{\theta_2 + \theta_3} \\ P(\Delta_1 = 1, \Delta_2 = 2 | X_2 < X_1) &= \frac{\theta_1}{\theta_1 + \theta_3} \\ P(\Delta_1 = 3, \Delta_2 = 2 | X_1 < X_2) &= \frac{\theta_3}{\theta_1 + \theta_3}. \end{aligned}$$

Now let us denote at the $k$-th stage of the EM algorithm, the estimates of $\theta_1, \theta_2, \theta_3$ and $c_1, \ldots, c_M$ as $\theta_1^{(k)}, \theta_2^{(k)}, \theta_3^{(k)}$ and $c_1^{(k)}, \ldots, c_M^{(k)}$, respectively. Let us further denote $a^{(k)} =$

$\dfrac{\theta_1^{(k)}}{\theta_1^{(k)}+\theta_3^{(k)}}$ and $b^{(k)} = \dfrac{\theta_2^{(k)}}{\theta_2^{(k)}+\theta_3^{(k)}}$. Now following the same procedure as in Dinse [10] we obtain the 'pseudo' log-likelihood function at the $(k+1)$-th stage as

$$\begin{aligned}
l_p(\boldsymbol{\theta}, \boldsymbol{c}|\boldsymbol{\theta}^{(k)}, \boldsymbol{c}^{(k)}) &= (n_1 + n_2 b^{(k)})\ln\theta_1 + (n_1 a^{(k)} + n_2)\ln\theta_2 + (n_1(1-a^{(k)}) + n_2(1-b^{(k)}) + n_3)\ln\theta_3 \\
&\quad + \sum_{i\in I_1\cup I_2}\{\ln h_0(x_i) + \ln h_0(y_i)\} + \sum_{i\in I_3}\{\ln h_0(x_i)\} - \theta_1\sum_{i\in I} H_0(x_i) - \theta_2\sum_{i\in I} H_0(y_i) \\
&\quad -\theta_3(\sum_{i\in I_1} H_0(y_i) + \sum_{i\in I_2} H_0(x_i) + \sum_{i\in I_3} H_0(x_i)) \\
&= (n_1 + n_2 b^{(k)})\ln\theta_1 + (n_1 a^{(k)} + n_2)\ln\theta_2 + (n_1(1-a^{(k)}) + n_2(1-b^{(k)}) + n_3)\ln\theta_3 \\
&\quad + \sum_{v=1}^{M} n_{\cdot v}\ln c_v - \theta_1 A_1(\boldsymbol{c}|\mathcal{D}_1) - \theta_2 A_2(\boldsymbol{c}|\mathcal{D}_1) - \theta_3 A_3(\boldsymbol{c}|\mathcal{D}_1). \qquad (9)
\end{aligned}$$

Therefore, at the $(k+1)$-th step of the EM algorithm, for a given $\boldsymbol{c}$, $\theta_1^{(k+1)}(\boldsymbol{c})$, $\theta_2^{(k+1)}(\boldsymbol{c})$ and $\theta_3^{(k+1)}(\boldsymbol{c})$ can be obtained as

$$\theta_1^{(k+1)}(\boldsymbol{c}) = \frac{n_1 + n_2 b^{(k)}}{A_1(\boldsymbol{c}|\mathcal{D}_1)}, \quad \theta_2^{(k+1)}(\boldsymbol{c}) = \frac{n_1 a^{(k)} + n_2}{A_2(\boldsymbol{c}|\mathcal{D}_1)}, \quad \theta_1^{(k+1)}(\boldsymbol{c}) = \frac{n_1(1-a^{(k)}) + n_2(1-b^{(k)}) + n_3}{A_3(\boldsymbol{c}|\mathcal{D}_1)}.$$

Hence, $\boldsymbol{c}^{(k+1)}$ can be obtained by maximizing

$$\begin{aligned}
g(\boldsymbol{c}|\boldsymbol{\theta}^{(k)}) &= \sum_{v=1}^{M} n_{\cdot v}\ln c_v - (n_1 + n_2 b^{(k)})\ln A_1(\boldsymbol{c}|\mathcal{D}_1) - (n_1 a^{(k)} + n_2)\ln A_2(\boldsymbol{c}|\mathcal{D}_1) - \\
&\qquad (n_1(1-a^{(k)}) + n_2(1-b^{(k)}) + n_3)\ln A_3(\boldsymbol{c}|\mathcal{D}_1), \qquad (10)
\end{aligned}$$

with respect to $\boldsymbol{c}$. Once $\boldsymbol{c}^{(k+1)}$ is obtained, $\theta_1^{(k+1)}$, $\theta_2^{(k+1)}$ and $\theta_3^{(k+1)}$ can be obtained as $\theta_1^{(k+1)}(\boldsymbol{c}^{(k+1)})$, $\theta_2^{(k+1)}(\boldsymbol{c}^{(k+1)})$ and $\theta_3^{(k+1)}(\boldsymbol{c}^{(k+1)})$.

### 3.3 Some Special Cases

In this section we present explicit expressions of $A_1(\mathbf{c})$, $A_2(\mathbf{c})$, $A_3(\mathbf{c})$, and $g(\mathbf{c}|\theta^{(k)})$, for some special cases namely $M$ = 1,2 and 3, which are useful in practice.

**Case 1:** $M = 1$.

In this case clearly, $A_1(\mathbf{c})$, $A_2(\mathbf{c})$ and $A_3(\mathbf{c})$ do not depend on $\mathbf{c}$, and let us denote them as $A_1$, $A_2$ and $A_3$, respectively. Here, they are as follows:

$$A_1 = \sum_{i \in I_1 \cup I_2} x_i + \sum_{i \in I_3} z_i, A_2 = \sum_{i \in I_1 \cup I_2} y_i + \sum_{i \in I_3} z_i, A_3 = \sum_{i \in I_2} x_i + \sum_{i \in I_1} y_i + \sum_{i \in I_3} z_i.$$

Therefore, in this case at each 'E'-step of the EM algorithm, the corresponding 'M'-step can be obtained explicitly. It may be mentioned that Karlis [15] also has proposed an EM algorithm in case of MOBE distribution, but this proposed EM algorithm is slightly different than that. This proposed EM algorithm does not need any expectation calculation explicitly.

**Case 2:** $M = 2$.

In this case let us denote $A_1(\mathbf{c})$, $A_2(\mathbf{c})$,$A_3(\mathbf{c})$ as $A_1(c)$, $A_2(c)$, $A_3(c)$, respectively, and $\tau_1 = \tau$. Then

$$\begin{aligned}
A_1(c) &= c\tau(\#\{i : x_i > \tau\} + \#\{i : z_i > \tau\}) + c(\sum_{i:x_i<\tau} x_i + \sum_{i:z_i<\tau} z_i) + \sum_{i:x_i>\tau} (x_i - \tau) + \sum_{i:z_i>\tau} (z_i - \tau) \\
A_2(c) &= c\tau(\#\{i : y_i > \tau\} + \#\{i : z_i > \tau\}) + c(\sum_{i:y_i<\tau} y_i + \sum_{i:z_i<\tau} z_i) + \sum_{i:y_i>\tau} (y_i - \tau) + \sum_{i:z_i>\tau} (z_i - \tau) \\
A_3(c) &= c\tau(\#\{i : i \in I_2, x_i > \tau\} + \#\{i : i \in I_1, y_i > \tau\} + \#\{i : z_i > \tau\}) + \\
& \quad c\left(\sum_{i:i\in I_2, x_i<\tau} x_i + \sum_{i:i\in I_1, y_i<\tau} y_i + \sum_{i:z_i<\tau} z_i\right) + \sum_{i:i\in I_2, x_i>\tau} (x_i - \tau) + \sum_{i:i\in I_1, y_i>\tau} (y_i - \tau) + \sum_{i:z_i>\tau} (z_i - \tau),
\end{aligned}$$

$$\begin{aligned}
g(c|\boldsymbol{\theta}^{(k)}) &= (\#\{i : x_i < \tau\} + \#\{i : y_i < \tau\} + \#\{i : z_i < \tau\}) \ln c - (n_1 + n_2 b^{(k)}) \ln A_1(c) \\
& \quad -(n_1 a^{(k)} + n_2) \ln A_2(c) - (n_1(1 - a^{(k)}) + n_2(1 - b^{(k)}) + n_3) \ln A_3(c)
\end{aligned}$$

**Case 3:** $M = 3$.

In this case let us denote $A_1(\mathbf{c})$, $A_2(\mathbf{c})$,$A_3(\mathbf{c})$ as $A_1(c_1, c_2)$, $A_2(c_1, c_2)$, $A_3(c_1, c_2)$, respectively,

$$A_1(c_1, c_2) = c_1\tau_1[\#\{i : \tau_1 < x_i \leq \tau_3\} + \#\{i : \tau_1 < z_i \leq \tau_3\}] +$$

$$
\begin{aligned}
& c_2(\tau_2-\tau_1)[\#\{i:\tau_2<x_i\leq\tau_3\}+\#\{i:\tau_2<z_i\leq\tau_3\}]+ \\
& c_1\sum_{i:x_i\leq\tau_1}x_i+c_2\sum_{i:\tau_1<x_i\leq\tau_2}(x_i-\tau_1)+\sum_{i:\tau_2<x_i\leq\tau_3}(x_i-\tau_2)+ \\
& c_1\sum_{i:z_i\leq\tau_1}z_i+c_2\sum_{i:\tau_1<z_i\leq\tau_2}(z_i-\tau_1)+\sum_{i:\tau_2<z_i\leq\tau_3}(z_i-\tau_2) \\
A_2(c_1,c_2) \;=\; & c_1\tau_1[\#\{i:\tau_1<y_i\leq\tau_3\}+\#\{i:\tau_1<z_i\leq\tau_3\}]+ \\
& c_2(\tau_2-\tau_1)[\#\{i:\tau_2<y_i\leq\tau_3\}+\#\{i:\tau_2<z_i\leq\tau_3\}]+ \\
& c_1\sum_{i:y_i\leq\tau_1}y_i+c_2\sum_{i:\tau_1<y_i\leq\tau_2}(y_i-\tau_1)+\sum_{i:\tau_2<y_i\leq\tau_3}(y_i-\tau_2)+ \\
& c_1\sum_{i:z_i\leq\tau_1}z_i+c_2\sum_{i:\tau_1<z_i\leq\tau_2}(z_i-\tau_1)+\sum_{i:\tau_2<z_i\leq\tau_3}(z_i-\tau_2) \\
A_3(c_1,c_2) \;=\; & c_1\tau_1(\#\{i:i\in I_2,x_i>\tau_1\}+\#\{i:i\in I_1,y_i>\tau_1\}+\#\{i:z_i>\tau_1\})+ \\
& c_2(\tau_2-\tau_1)(\#\{i:i\in I_2,\tau_2<x_i\leq\tau_3\}+\#\{i:i\in I_1,\tau_2<y_i\leq\tau_3\}+\#\{i:\tau_2<z_i\leq\tau_3\})+ \\
& c_1\left(\sum_{i:i\in I_2,x_i<\tau_1}x_i+\sum_{i:i\in I_1,y_i<\tau_1}y_i+\sum_{i:z_i<\tau_1}z_i\right)+ \\
& c_2\left(\sum_{i:i\in I_2,\tau_1<x_i\leq\tau_2}(x_i-\tau_1)+\sum_{i:i\in I_1,\tau_1<y_i\leq\tau_2}(y_i-\tau_1)+\sum_{i:\tau_1<z_i\leq\tau_2}(z_i-\tau_1)\right)+ \\
& \left(\sum_{i:i\in I_2,\tau_2<x_i}(x_i-\tau_2)+\sum_{i:i\in I_1,\tau_2<y_i}(y_i-\tau_2)+\sum_{i:\tau_2<z_i}(z_i-\tau_2)\right),
\end{aligned}
$$

$$
\begin{aligned}
g(c_1,c_2|\boldsymbol{\theta}^{(k)}) \;=\; & (\#\{i:x_i<\tau_1\}+\#\{i:y_i<\tau_1\}+\#\{i:z_i<\tau_1\})\ln c_1+ \\
& (\#\{i:\tau_1<x_i\leq\tau_2\}+\#\{i:\tau_1<y_i\leq\tau_2\}+\#\{i:\tau_1<z_i\leq\tau_2\})\ln c_2- \\
& (n_1+n_2b^{(k)})\ln A_1(c_1,c_2)-(n_1a^{(k)}+n_2)\ln A_2(c_1,c_2)- \\
& (n_1(1-a^{(k)})+n_2(1-b^{(k)})+n_3)\ln A_3(c)
\end{aligned}
$$

# 4 Simulation Results

In this section we have performed some simulation experiments to see the performance of the proposed EM algorithm. We have considered different $M$, different set of parameters and different sample sizes. In each case we have calculated the maximum likelihood estimators based on EM algorithm. About the initial estimates of the EM algorithm, we have used Theorem 5. We have stopped the EM algorithm when the difference between the two consecutive log-likelihood values become less than $10^{-5}$. In each case we have reported the average estimates (AEs), the associated mean squared errors (MSEs), and the average number of iterations (ANI). All the results are based on 1000 replications. We have considered the following four models: Model 1: $M = 1$, $\theta_1 = \theta_2 = \theta_3 = 1.0$; Model 2: $M = 1$, $\theta_1 = \theta_2 = 1.0$, $\theta_3 = 2.0$; Model 3: $M = 2$, $\theta_1 = \theta_2 = \theta_3 = 1.0$, $c = 2.0$ and $\tau = 0.1$; Model 4: $\theta_1 = \theta_2 = 2$, $\theta_3 = 1.0$, $c = 2.0$ and $\tau = 0.1$. The results are reported in Tables 1 to 4.

Table 1: The average estimates, the associated mean squared, the average number of iterations for Model 1.

| $n$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | ANI |
|---|---|---|---|---|
| 25 | 1.0512 | 1.0497 | 1.0115 | 12.4 |
| | (0.0944) | (0.0884) | (0.0958) | |
| 50 | 1.0299 | 1.0288 | 0.9962 | 12.6 |
| | (0.0435) | (0.0407) | (0.0475) | |
| 75 | 1.0234 | 1.0206 | 0.9988 | 12.7 |
| | (0.0285) | (0.0261) | (0.0287) | |
| 100 | 1.0180 | 1.0182 | 0.9981 | 13.2 |
| | (0.0204) | (0.0188) | (0.0217) | |

Some of the points are quite clear from these table values. It is observed that as the sample size increases the average number of iterations required for the EM algorithm to converge also increases. In all the cases as the sample size increases the biases and the MSEs of all the estimators decrease, as expected. It verifies the consistency properties of the MLEs. From Tables 1 and 3 it is apparent that when $\theta_1 = \theta_2 = \theta_3$, the biases and the MSEs of the corresponding MLEs are almost equal, when the sample sizes are also equal. Comparing Tables 1 and 2 and similarly Tables 3 and 4 it is observed that if $\theta_1 = \theta_2$ and $\theta_3$

Table 2: The average estimates, the associated mean squared, the average number of iterations for Model 2.

| $n$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | ANI |
|---|---|---|---|---|
| 25 | 1.0610 | 1.0587 | 2.0286 | 9.5 |
| | (0.1482) | (0.1353) | (0.2658) | |
| 50 | 1.0313 | 1.0279 | 2.0136 | 9.8 |
| | (0.0628) | (0.0605) | (0.1286) | |
| 75 | 1.0230 | 1.0191 | 2.0049 | 9.9 |
| | (0.0423) | (0.0384) | (0.0888) | |
| 100 | 1.0187 | 1.0187 | 2.0000 | 10.3 |
| | (0.0302) | (0.0291) | (0.0597) | |

is different, then the biases and the MSEs of all the estimators increase for the same sample sizes. Comparing Tables 1 and 3 we can infer that if $\theta_1$, $\theta_2$ and $\theta_3$ are fixed and number of cut points increases, then the biases and MSEs of the corresponding MLEs of $\theta_1$, $\theta_2$, $\theta_3$ increase. It is quite clear that the proposed EM algorithm is working quite well in this case and it can be used in practice quite conveniently.

# 5 Data Analysis: UEFA Champion League

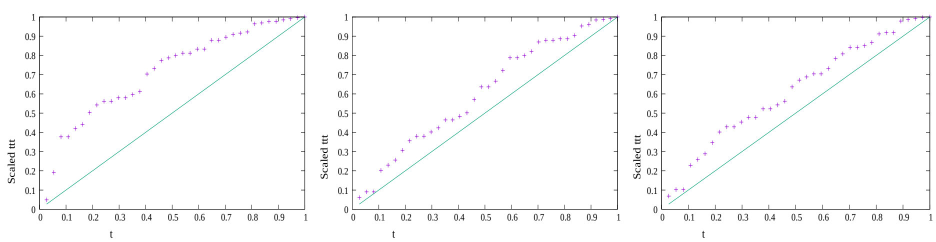
In this section we have analyzed one data set to see how the proposed model and the EM algorithm work in practice. This data set is a bivariate data set with some ties and it has been obtained from Meintanis [19] It is a the UEFA Champion's League soccer data set during the year 2004-2005 and 2005-2006. Here the variable $X_1$ represents the time of the first *kick* goal scored by any team, and $X_2$ represents the time of the first goal (any) scored by the home team. It may be noted that in this case all possibilities are open, namely $X_1 < X_2$, $X_2 > X_1$ or $X_1 = X_2$. The data set has been presented in Table 5.

Before progressing any further, we have scaled all the values of the data set, so that all

Table 3: The average estimates, the associated mean squared, the average number of iterations for Model 3.

| $n$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $c$ | ANI |
|---|---|---|---|---|---|
| 25 | 1.2520 | 1.2481 | 1.1202 | 1.2287 | 20.8 |
| | (0.2217) | (0.2023) | (0.2058) | (0.8227) | |
| 50 | 1.1953 | 1.1991 | 1.1106 | 1.2461 | 26.7 |
| | (0.1017) | (0.0999) | (0.0892) | (0.6937) | |
| 75 | 1.1830 | 1.1789 | 1.1000 | 1.2498 | 28.9 |
| | (0.0740) | (0.0700) | (0.0588) | (0.6479) | |
| 100 | 1.1738 | 1.1737 | 1.0941 | 1.2559 | 32.4 |
| | (0.0585) | (0.0579) | (0.0446) | (0.6203) | |

values are between 0 and 1, it is not going to make any difference in the inference procedure. In Figure 1, we have plotted the scaled TTT transforms of $X_1$, $X_2$ and $\min\{X_1, X_2\}$, respectively. It is clear from the plots that the empirical hazard functions of $X_1$, $X_2$ and $\min\{X_1, X_2\}$ are all increasing functions. We have fitted the proposed model with $M$ = 1,2




(a) (b) (c)

Figure 1: The PDF plot of the marginal distribution of $X$, when (a) Scaled TTT plot for $X_1$ , (b) Scaled TTT plot for $X_2$, (c) Scaled TTT plot for $\min\{X_1, X_2\}$.

and 3. In all the cases we have used the EM algorithm as proposed in Section 3.2. We have used the same initial estimates and the stopping criterion as it has been used in the previous section. We have considered different $\tau_1$ and $\tau_2$ in case of $M$ = 2 and 3. We have reported the MLEs of the unknown parameters, the associated log-likelihood and BIC values and the number of iterations needed to converge in each case in Table 6.

Table 4: The average estimates, the associated mean squared, the average number of iterations for Model 4.

| $n$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $c$ | ANI |
|---|---|---|---|---|---|
| 25 | 2.4990 | 2.4893 | 1.1172 | 1.4546 | 23.9 |
| | (0.8462) | (0.7788) | (0.3518) | (0.5126) | |
| 50 | 2.3690 | 2.3738 | 1.0931 | 1.4870 | 29.1 |
| | (0.3581) | (0.3605) | (0.1286) | (0.3774) | |
| 75 | 2.3395 | 2.3296 | 1.0861 | 1.4901 | 33.1 |
| | (0.2529) | (0.2406) | (0.0899) | (0.3370) | |
| 100 | 2.3165 | 2.3169 | 1.0800 | 1.4994 | 33.6 |
| | (0.1996) | (0.2006) | (0.0659) | (0.3127) | |

Based on the above results, it is observed that the best fitted model is $M = 2$ and $\tau_1 = 0.5$. The MLEs of the unknown parameters and the associated 95% asymptotic confidence intervals become $\widehat{\theta}_1 = 1.5521(\pm 0.3145)$, $\widehat{\theta}_2 = 3.4913(\pm 0.7812)$, $\widehat{\theta}_3 = 3.5191(\pm 0.7542)$ and $\widehat{c}_1 = 0.3223(\pm 0.0112)$. The Kolmogorov-Smirnov distances and the associated $p$ values, reported within brackets, between the empirical distribution functions of $X$, $Y$, $\min\{X, Y\}$ and the fitted distribution functions based on the MLEs of the above model are 0.198 (0.109), 0.089 (0.929) and 0.139 (0.467), respectively. Therefore, we cannot reject the null hypothesis at the 5% level of significance that $X$, $Y$ and $\min\{X, Y\}$ follow PHC. Hence, it seems reasonable to fit BPHC to this data set.

# 6 Applications

In lifetime experiment in many cases an item can fail due to more than one causes. In such a situation the experimenter wants to analyze the effect of one cause in presence of other causes. In the statistical literature it is known as the competing risks problem. Analyzing competing risks data has played an important role in the lifetime data analysis. There are mainly two approaches to analyze competing risks data. One is known as the latent failure

Table 5: UEFA Champion's League data

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

time model, proposed by Cox [8] and the other is known as the cause specific hazard function model, introduced by Prentice et al. [23]. An extensive amount of work has been done in analyzing different competing risks data.

## 6.1 Data Description

The data set has been obtained from the Diabetic Retinopathy study (DRS) conducted by the National Eye Institute. The study was conducted to examine the effect of laser treatment in delaying the onset of blindness in patients with diabetic Retinopathy. The experiment can be briefly described as follows. At the beginning of the experiment, one eye for each patient was selected at random for laser treatment, and the other eye was not given the laser

| $M$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\tau_1$ | $\tau_2$ | $c_1$ | $c_2$ | # iter | log-ll | BIC |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.6503 | 1.4716 | 1.5909 | | | | | 15 | -20.4720 | 51.777 |
| 2 | 0.9454 | 2.1648 | 2.1694 | 0.25 | | 0.4198 | | 25 | -15.4752 | 45.390 |
| 2 | 1.5521 | 3.4913 | 3.5191 | 0.50 | | 0.3223 | | 28 | -12.7311 | 39.906 |
| 2 | 2.7920 | 6.0049 | 6.2587 | 0.75 | | 0.2247 | | 26 | -15.9677 | 46.379 |
| 3 | 1.5630 | 3.5385 | 3.4801 | 0.25 | 0.50 | 0.2579 | 0.4210 | 25 | -11.6868 | 41.428 |
| 3 | 2.8324 | 6.1623 | 6.1582 | 0.50 | 0.75 | 0.1820 | 0.4760 | 24 | -11.6552 | 41.365 |
| 3 | 2.8414 | 6.2023 | 6.1045 | 0.25 | 0.75 | 0.1458 | 0.3089 | 23 | -12.5964 | 43.247 |

Table 6: Maximum likelihood estimates of the unknown parameters for different $M$ and for different $\tau_1$ and $\tau_2$

treatment. The minimum time to blindness ($T$) and the indicator specifying whether the treated eye ($\Delta = 1$), the untreated eye ($\Delta = 2$) or both the eyes have failed simultaneously ($\Delta = 3$) have been recorded. The data set has been presented in Table 7.

| $T$ | $\Delta$ | $T$ | $\Delta$ | $T$ | $\Delta$ | $T$ | $\Delta$ | $T$ | $\Delta$ | $T$ | $\Delta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 266 | 1 | 272 | 3 | 203 | 3 | 91 | 2 | 1137 | 3 | 84 | 1 |
| 154 | 2 | 1484 | 1 | 392 | 1 | 285 | 3 | 315 | 1 | 1140 | 2 |
| 583 | 1 | 287 | 2 | 901 | 1 | 547 | 2 | 1252 | 1 | 1247 | 3 |
| 79 | 1 | 717 | 2 | 448 | 2 | 622 | 0 | 642 | 1 | 904 | 2 |
| 707 | 2 | 141 | 2 | 276 | 1 | 469 | 2 | 407 | 1 | 520 | 1 |
| 93 | 1 | 356 | 1 | 485 | 2 | 1313 | 2 | 1653 | 3 | 248 | 2 |
| 805 | 1 | 427 | 2 | 503 | 1 | 344 | 1 | 699 | 1 | 423 | 2 |
| 790 | 2 | 36 | 2 | 285 | 2 | 125 | 2 | 667 | 1 | 315 | 2 |
| 777 | 2 | 588 | 2 | 727 | 2 | 306 | 1 | 471 | 3 | 210 | 2 |
| 415 | 1 | 126 | 1 | 409 | 2 | 307 | 2 | 350 | 2 | 584 | 1 |
| 637 | 2 | 350 | 1 | 355 | 1 | 577 | 2 | 663 | 3 | 1302 | 1 |
| 178 | 1 | 567 | 2 | 227 | 2 | 517 | 2 | 966 | 3 | | |

Table 7: Minimum time to blindness in days and its causes

## 6.2 Model Formulation & Maximum Likelihood Estimators

Suppose $X_1$ and $X_2$ denote the time to blindness of a patient with diabetic Retinopathy due to laser treatment and without laser treatment, respectively. Let $T$ denote the time to blindness of an eye. Therefore, $T = \min\{X_1, X_2\}$. Clearly, $X_1$ and $X_2$ are not independent.

Feizjavadian and Hashemi [11], Cai et al. [7] and Samanta and Kundu [21] analyzed this data set based on the assumption that $(X_1, X_2)$ follows Marshall-Olkin bivariate Weibull distribution. Here, it is assumed that $(X_1, X_2)$ follows BPHC$(S_0, \theta_1, \theta_2, \theta_3)$. Based on the above assumption, the log-likelihood contributions for different cases can be written as follows:

**Case 1** $(T = t, \Delta = 1)$: The likelihood contribution is:

$$\theta_1 f_0(t)(S_0(t))^{\theta_1-1}(S_0(t))^{\theta_2+\theta_3} = \theta_1 h_0(t)(S_0(t))^{\theta_1+\theta_2+\theta_3}.$$

**Case 2** $(T = t, \Delta = 2)$: The likelihood contribution is:

$$\theta_2 f_0(t)(S_0(t))^{\theta_2-1}(S_0(t))^{\theta_1+\theta_3} = \theta_2 h_0(t)(S_0(t))^{\theta_1+\theta_2+\theta_3}.$$

**Case 3** $(T = t, \Delta = 3)$: The likelihood contribution is:

$$\theta_3 f_0(t)(S_0(t))^{\theta_1-1}(S_0(t))^{\theta_2+\theta_3} = \theta_3 h_0(t)(S_0(t))^{\theta_1+\theta_2+\theta_3}.$$

Let us make the following assumption. We have the following competing risks data:

$$\mathcal{D}_2 = \{(t_1, \delta_1), \ldots, (t_m, \delta_m)\},$$

and let $m_j$ denote the $\#\{i : \delta_i = m_j\}$, for $j =$ 1,2 and 3. It is further assumed that $h_0(t)$ has the same assumption as (7). Moreover, we introduce the following notations for further development.

$$J_k = \{i : \tau_{k-1} < t_i \leq \tau_k\}; \qquad k = 1, \ldots, M,$$

and the number of elements in $J_k$ is $u_k$, for $k = 1, \ldots, M$. Therefore, the log-likelihood function of the observed data $\mathcal{D}_2$ becomes

$$\begin{aligned} l_c(\boldsymbol{\theta}, \mathbf{c}|\mathcal{D}_2) &= m_1 \ln\theta_1 + m_2 \ln\theta_2 + m_3 \ln\theta_3 + \sum_{i=1}^{m} \ln h_0(t_i) - (\theta_1 + \theta_2 + \theta_3) \sum_{i=1}^{m} H_0(t_i) \\ &= m_1 \ln\theta_1 + m_2 \ln\theta_2 + m_3 \ln\theta_3 + \sum_{j=1}^{M-1} u_j \ln c_j + (\theta_1 + \theta_2 + \theta_3) \left\{ \sum_{j=1}^{M-1} c_j B_j + \sum_{i \in J_M} (t_i - \tau_{M-1}) \right\} \end{aligned}$$

where

$$B_j = \left\{ (\tau_j - \tau_{j-1}) \sum_{k=1+j}^{M} u_k + \sum_{i \in J_j} (t_i - \tau_{j-1}) \right\}.$$

Hence, for a given $\mathbf{c}$, the MLEs of $\theta_1$, $\theta_2$ and $\theta_3$ can be obtained as

$$\widehat{\theta}_1(\mathbf{c}) = \frac{m_1}{\sum_{j=1}^{M-1} c_j B_j + \sum_{i \in J_M} (t_i - \tau_{M-1})}, \quad \widehat{\theta}_2(\mathbf{c}) = \frac{m_2}{\sum_{j=1}^{M-1} c_j B_j + \sum_{i \in J_M} (t_i - \tau_{M-1})},$$

and

$$\widehat{\theta}_3(\mathbf{c}) = \frac{m_3}{\sum_{j=1}^{M-1} c_j B_j + \sum_{i \in J_M} (t_i - \tau_{M-1})}.$$

The MLE of $\mathbf{c}$ can be obtained by maximizing $w(\mathbf{c})$, where

$$w(\mathbf{c}) = m_1 \ln \widehat{\theta}_1(\mathbf{c}) + m_2 \ln \widehat{\theta}_2(\mathbf{c}) + m_3 \ln \widehat{\theta}_3(\mathbf{c}) + \sum_{j=1}^{M-1} u_j \ln c_j.$$

## 6.3 Some Special Cases

In this section we present explicit expressions of $\widehat{\theta}_1(\mathbf{c})$, $\widehat{\theta}_2(\mathbf{c})$, and $\widehat{\theta}_3(\mathbf{c})$ for some special cases namely $M$ = 1,2 and 3, which are useful in practice.

**Case 1:** $M = 1$. In this case the unknown parameters are $\theta_1$, $\theta_2$ and $\theta_3$, and they can be obtained in explicit forms as follows:

$$\widehat{\theta}_1 = \frac{m_1}{\sum_{i=1}^{m} t_i}, \quad \widehat{\theta}_2 = \frac{m_2}{\sum_{i=1}^{m} t_i}, \widehat{\theta}_3 = \frac{m_3}{\sum_{i=1}^{m} t_i}.$$

**Case 2:** $M = 2$. In this case the unknown parameters are $\theta_1$, $\theta_2$, $\theta_3$, and $c_1$. The explicit expressions of $\widehat{\theta}_1(c_1)$, $\widehat{\theta}_2(c_1)$, and $\widehat{\theta}_3(c_1)$ are as follows:

$$\begin{aligned}
\widehat{\theta}_1(c_1) &= \frac{m_1}{c_1 \tau_1 u_2 + c_1 \sum_{i \in J_1} t_i + \sum_{i \in J_2} (t_i - \tau_1)}, \\
\widehat{\theta}_2(c_1) &= \frac{m_2}{c_1 \tau_1 u_2 + c_1 \sum_{i \in J_1} t_i + \sum_{i \in J_2} (t_i - \tau_1)}, \\
\widehat{\theta}_3(c_1) &= \frac{m_3}{c_1 \tau_1 u_2 + c_1 \sum_{i \in J_1} t_i + \sum_{i \in J_2} (t_i - \tau_1)},
\end{aligned}$$

**Case 3:** $M = 3$. In this case the unknown parameters are $\theta_1$, $\theta_2$, $\theta_3$, $c_1$ and $c_2$. The explicit expressions of $\widehat{\theta}_1(c_1, c_2)$, $\widehat{\theta}_2(c_1, c_2)$, and $\widehat{\theta}_3(c_1, c_2)$ are as follows:

$$\begin{aligned}
\widehat{\theta}_1(c_1, c_2) &= \frac{m_1}{c_1\{\tau_1(u_2+u_3) + \sum_{i\in J_1} t_i\} + c_2\{(\tau_2-\tau_1)u_3 + \sum_{i\in J_2}(t_i-\tau_1)\} + \sum_{i\in J_3}(t_i-\tau_2)}, \\
\widehat{\theta}_2(c_1, c_2) &= \frac{m_2}{c_1\{\tau_1(u_2+u_3) + \sum_{i\in J_1} t_i\} + c_2\{(\tau_2-\tau_1)u_3 + \sum_{i\in J_2}(t_i-\tau_1)\} + \sum_{i\in J_3}(t_i-\tau_2)}, \\
\widehat{\theta}_3(c_1, c_2) &= \frac{m_3}{c_1\{\tau_1(u_2+u_3) + \sum_{i\in J_1} t_i\} + c_2\{(\tau_2-\tau_1)u_3 + \sum_{i\in J_2}(t_i-\tau_1)\} + \sum_{i\in J_3}(t_i-\tau_2)}.
\end{aligned}$$

## 6.4 Data Analysis: Diabetic Retinopathy

In this section we present the analysis of the diabetic Retinopathy data set as presented in Table 7. In this case we have fitted the model as described in Section 6.2 with $M = 1$, 2 and 3. We have used different $\tau_1$ and $\tau_2$ values for $M = 2$ and 3. We have presented the MLEs of the unknown parameters, the associated log-likelihood and BIC values in Table 8. According to the smallest BIC values, the best fitted model is when $M = 2$, and $\tau_1 = 0.25$.

| $M$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\tau_1$ | $\tau_2$ | $c_1$ | $c_2$ | log-ll | BIC |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.3125 | 1.5469 | 0.4687 | | | | | -56.5649 | 125.918 |
| 2 | 1.8290 | 2.1557 | 0.6532 | 0.25 | | 0.5630 | | -53.6840 | 124.419 |
| 2 | 1.8763 | 2.2114 | 0.6701 | 0.50 | | 0.6629 | | -55.8521 | 128.755 |
| 2 | 3.9935 | 4.7066 | 1.4262 | 0.75 | | 0.3160 | | -54.7472 | 126.545 |
| 3 | 1.8758 | 2.2108 | 0.6699 | 0.25 | 0.50 | 0.5491 | 0.9639 | -53.6789 | 128.671 |
| 3 | 3.9877 | 4.6998 | 1.4142 | 0.50 | 0.75 | 0.3120 | 0.3610 | -54.6828 | 130.679 |
| 3 | 3.9927 | 4.7057 | 1.4259 | 0.25 | 0.75 | 0.2580 | 0.4280 | -52.6523 | 126.618 |

Table 8: Maximum likelihood estimates of the unknown parameters for different $M$ and for different $\tau_1$ and $\tau_2$

The corresponding MLEs of the unknown parameters, and the associated 95% asymptotic confidence intervals are $\widehat{\theta}_1 = 1.8290(\pm 0.3476)$, $\widehat{\theta}_2 = 2.1557(\pm 0.6487)$, $\widehat{\theta}_3 = 0.6532(\pm 0.0868)$ and $\widehat{c}_1 = 0.5630(\pm 0.0468)$. The Kolmogorov-Smirnov distance and the associated $p$ value, between the empirical CDF and the fitted CDF are 0.1085 and 0.3728, respectively. Based on the $p$ value, we conclude that the proposed distribution fits the data well.

Now we would like to answer the question whether the laser treatment has any effect in delaying the blindness or not. It means we need to test the following hypothesis

$$H_0 : \theta_1 = \theta_2 \qquad vs. \qquad H_1 : \theta_1 \neq \theta_2.$$

Under the null hypothesis the MLEs of the unknown parameters are $\widehat{\theta}_1 = \widehat{\theta}_2 = 1.9924$, $\widehat{\theta}_3 = 0.6532$, and the associated log-likelihood value becomes -53.8894. The likelihood ratio test has the $p$-value 0.65. Hence, the data indicate that the laser treatment does not have any significant effect in delaying the blindness of a diabetic patient.

# 7 Conclusions

In this paper we have introduced a bivariate distribution with a singular component and it has marginals belong to the PHC. We did not assume any specific form of the base line distribution, instead we have assumed that it has piecewise constants hazard function. It makes the proposed distribution quite flexible. We have proposed a very convenient EM algorithm to estimate the unknown parameters. Further, we have used this model to analyze competing risks data with dependent cause of failures. The results are quite satisfactory. Note that in this paper we have mainly discussed the classical inference of the unknown parameters. It will be interesting to consider the corresponding Bayesian inference also. More work is needed along that direction.

# Acknowledgements:

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

# Conflict of Interest and Funding Statements

The author does not have any conflict of interest. This manuscript has not been funded by any funding agent.

## References

[1] Azzalini, A.A. and Capitanio, A. (2014), *The skew-normal and related families*, Cambridge University Press, Cambridge, UK

[2] Balakrishnan, N., Koutras, M.V., Milenos, F.S., Pal, S. (2016), "Piecewise linear approximation for cure rate models and associated inferential issues", *Methodology and Computing in Applied Probability*, vol. 18, 937 – 966.

[3] Balakrishnan, N. and Lai, C. (2009), *Continuous bivariate distributions*, 2nd edition, Springer, New York.

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

[5] Basu, A.P. (1971), "Bivariate failure rate", *Journal of the American Statistical Association*, vol. 66, 103–104.

[6] Bemis, B., Bain, L. J. and Higgins, J. J. (1972), "Estimation and hypothesis testing for the parameters of a bivariate exponential distribution", *Journal of the American Statistical Association*, vol. 67, 927-929.

[7] Cai, J., Shi, Y., Liu, B. (2017), "Analysis of incomplete data in the presence of dependent competing risks from Marshall-Olkin bivariate Weibull distribution

under progressive hybrid censoring", *Communications in Statistics - Theory and Methods*, vol. 46, 6497 – 6511.

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

[9] Cox, D. R. and Oakes, D. (1984), *Analysis of Survival Data*, Chapman & Hall, New York.

[10] Dinse, G. E. (1982), "Non-parametric estimation of partially incomplete time and types of failure data", *Biometrics*, vol. 38, 417–431.

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

[12] Hu. T., Khaledi, B-E and Shaked, M. (2003), "Multivariate hazard rate", *Journal of Multivariate Analysis*, vol. 84, 173 - 189.

[13] Ibrahim, J.G., Chen, M.H. and Sinha, D. (2001), *Bayesian Survival Analysis*, Wiley, New York.

[14] Johnson, N.L. and Kotz, S. (1975), "A vector multivariate hazard rate", *Journal of Multivariate Analysis*, vol. 5, 53–66.

[15] Karlis, D. (2003), "ML estimation for multivariate shock models via an EM algorithm", *Annals of the Institute of Statistical Mathematics*, vol. 55, 817 – 830.

[16] Kotz, S., Balakrishnan, N. and Johnson, N. (2000), *Continuous multivariate distributions*, 2nd. edition, John Wiley and Sons, New York.

[17] Kundu, D. and Gupta, R.D. (2009), "Bivariate generalized exponential distribution", *Journal of Multivariate Analysis*, vol. 100, no. 4, 581 – 593, 2009.

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

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

[20] Pena, E.A. and Gupta, A.K. (1990), "Bayes estimation for the Marshall- Olkin exponential distribution", *Journal of the Royal Statistical Society*, Ser. B, vol. 52,379 – 389.

[21] Samanta, D. and Kundu, D. (2021), "Bayesian Inference of a Dependent Competing Risk Data", *Journal of Statistical Computation and Simulation*, DOI:10.1080/00949655.2021.1917575.

[22] Kundu, D. and Mondal, S. (2021), "Analyzing Competing Risks Data Using Bivariate Weibull-Geometric Distribution", *Statistics*, vol. 55, 276 – 295.

[23] Prentice, R. L., Kalbfleish, J. D., Peterson, Jr. A. V., Flurnoy, N., Farewell, V. T.,Breslow, N. E. (1978), "The analysis of failure times in presence of competing risks", *Biometrics*, vol. 34, 541 – 554.

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