Estimating the Parameters of the Bivariate Burr Type III Distribution by EM Algorithm

Afsaneh Azizi [1] **and Abdolreza Sayyareh** [2]

[1] Department of Statistics, Razi University, Kermanshah, Iran.
[2] Department of Computer Science and Statistics, Faculty of Mathematics, K.N. Toosi University of Technology, Tehran, Iran.

Abstract. In recent years, bivariate lifetime distributions are often used to model reliability and survival data. In this paper, we introduce a bivariate Burr III distribution, so that the marginals have Burr III distributions. It is observed that the joint probability density function, the joint cumulative distribution function and the joint survival distribution function can be expressed in compact forms. We suggest to use the ECM algorithm to compute the maximum likelihood estimators of the unknown parameters. We report some simulation results and perform one data analysis for illustrative purposes.

Keywords. Bivariate distribution, Burr III distribution, ECM algorithm , Pseudo likelihood.

MSC: 62F10; 62E15.

Afsaneh Azizi: (azizi.afsaneh@stu.razi.ac.ir)
Corresponding Author: Abdolreza Sayyareh (asayyareh@kntu.ac.ir)

1 Introduction

Burr (1942) has developed a system of twelve types of distribution functions based on generating the Pearson differential equations. The density function has a range of shapes that is applicable to a wide area of applications. From the system of Burr distributions, the Burr XII distribution is widely used. The inverse distribution of Burr XII is Burr III. It is more flexible and includes a variety of distributions with varying degrees of skewness and kurtosis. The Burr III distribution with two parameters c and k which is denoted by $BIII(c,k)$ has been used in a variety of settings for the purpose of statistical modeling. Some examples include applications in forestry by Gove *et al.* (2008) and Lindsay *et al.* (1996), in fracture roughness by Nadarajah and Kotz (2006, 2007), in life testing by Wingo (1983, 1993), in operational risk by Chernobai *et al.* (2007), in option market price distributions by Sherrick *et al.* (1996), in meteorology by Mielke (1973), in modeling crop prices by Tejeda and Goodwin (2008), and in reliability by Abdel-Ghaly *et al.* (1997).

The probability density function and the cumulative distribution function of $BIII(c,k)$ are given by

$$\begin{array}{rcl} f_{BIII}(x;c,k) & = & kcx^{-c-1}(1+x^{-c})^{-k-1}, \qquad x > 0, c > 0, k > 0, \\ F_{BIII}(x;c,k) & = & (1+x^{-c})^{-k}, \end{array}$$

respectively, where c and k are the shape parameters. Various fields of science have used the Burr III distribution. It is also called the Dagum distribution in studies of income, wage and wealth distribution (see Dempster *et al.* (1977)).

In many cases, time the life/failure data of interest is bivariate in nature. Any study on twins or on failure data recorded twice on the same system naturally leads to bivariate data. For example, Houggard *et al.* (1992) studied data on lifelength of Danish twins and Lin *et al.* (1999) considered a data on patients with colon cancer where the paired data consist of the time from treatment to recurrence of cancer and the time from treatment to death. Paired data could consist of blindness in the left/right eye, failure time of the left/right kidney or age at death of parent/child in a genetic study.

In recent years, bivariate lifetime data are often used to model reliability and survival data. For example, Sarhan and Balakrishnan (2007) studied Marshall and Olkin bivariate exponential distribution, Al-Khedhairi *et al.* (2008) presented a new class of bivariate Gompertz distributions, Kundu and Gupta (2009) proposed the bivariate

generalized exponential distribution, Kundu and Gupta (2009) studied an EM algorithm for computing maximum likelihood estimators of the parameters of the bivariate Weibull distribution in the case of complete data, Nandi and Dewan (2010) have considered the maximum likelihood estimators of parameters of bivariate Weibull distribution under random censoring and El-Sherpieny *et al.* (2013) have introduced a new bivariate generalized Gompertz distribution.

Dempster *et al.* (1977) introduced a general iterative approach commonly known as the EM algorithm as an excellent tool for finding MLEs in cases where observations are treated as incomplete data. The EM algorithm has two main applications. The first case occurs when the data has missing values due to limitations or problems with the observation process. The second case occurs when the likelihood function can be obtained and simplified by assuming that there is an additional but missing parameter. We assume that a complete data, $Z = (X; Y)$ exists with Y being the missing data and that a joint density function also exists as follows

$$p(z|\theta) = p(x; y|\theta) = p(y|x; \theta)p(x|\theta),$$

where θ is a set of unknown parameters from a distribution including a missing parameter. With this density function, we now define the complete-data likelihood as

$$L(\theta|Z) = L(\theta|X; Y) = p(X; Y|\theta).$$

The original likelihood $L(\theta|X)$ is called the incomplete-data likelihood function. Since the missing data Y is unknown under a certain distribution by assumption, we can think of $L(\theta|X; Y)$ as a function of a random variable, Y, with constant values, X and θ

$$L(\theta|X; Y) = f_{(X;\theta)}(Y).$$

Using the complete-data log-likelihood function with respect to the missing data Y given the observed data X, the EM algorithm finds its expected value as well as the current parameter estimates at the E-step and maximizes the expectation at the M-step. By repeating the E and M-Steps, the algorithm is guaranteed to converge to a local maximum of the likelihood function with each iteration increasing the log-likelihood.

In this paper, the usual maximum likelihood estimators can be obtained by solving nonlinear equations, which is not a trivial issue. To avoid difficult computation, we treat this problem as a missing value problem and use the EM algorithm, which can be implemented more conveniently than the direct maximization process. Another

advantage of the EM algorithm is that it can be used to obtain the observed Fisher information matrix, which is helpful for constructing the asymptotic confidence intervals for the parameters.

In this paper, we study the maximum likelihood estimators of the parameters of the bivariate Burr III distribution under random left censoring. This article is organized as follows. In Section 2, we define the bivariate Burr III distribution and discuss its diffrent properties. The EM algorithm to compute the MLEs of the unknown parameters is provided in Section 3. The findings of the numerical experiments are reported in Section 4. One real data set is analyzed in Section 5 and we conclude in Section 6. The proof of theorems are given in Appendix A and the observed Fisher information matrix is given in Appendix B.

2   Bivariate Burr III Distribution

Consider the Burr III distribution, $BIII(c, k)$, with shape parameters $c > 0$ and $k > 0$. Suppose U_1, U_2, U_3 are independent $BIII(c, k_1)$, $BIII(c, k_2)$ and $BIII(c, k_3)$, random variables, respectively. Let $X_1 = \max\{U_1, U_3\}$ and $X_2 = \max\{U_2, U_3\}$. Then we say that the bivariate vector (X_1, X_2) has bivariate Burr III distribution with parameters c, k_1, k_2 and k_3. We will denote it by $BBIII(c, k_1, k_2, k_3)$.

Note that if $\max\{U_1, U_2, U_3\} = U_3$, then the two random variables X_1 and X_2 are equal. For example, suppose a system has the two components. Each component is subject to individual independent stress say U_1 and U_2 respectively. The system has an overall stress U_3 which has been transmitted to both the components equally, independent of their individual stresses. Therefore, the observed stress at the two components are $X_1 = \max\{U_1, U_3\}$ and $X_2 = \max\{U_2, U_3\}$ respectively. Now suppose the overall stress, U_3, is greater than the individual stresses U_1 and U_2, then $X_1 = X_2 = U_3$.

Let $k_{13} = k_1 + k_3$, $k_{23} = k_2 + k_3$ and $k_{123} = k_1 + k_2 + k_3$. The following theorems will provide the joint cumulative distribution function, CDF, and joint probability density function, PDF.

Theorem 2.1. *If $(X_1, X_2) \sim BBIII(c, k_1, k_2, k_3)$, then the joint CDF of (X_1, X_2) for $x_1 > 0$ and $x_2 > 0$, is*

$$F_{X_1, X_2}(x_1, x_2) = (1 + x_1^{-c})^{-k_1}(1 + x_2^{-c})^{-k_2}(1 + z^{-c})^{-k_3},$$

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

Proof. Proof is given in the Appendix A. □

Theorem 2.2. *If* $(X_1, X_2) \sim BBIII(c, k_1, k_2, k_3)$*, then the joint PDF of* (X_1, X_2) *for* $x_1 > 0$ *and* $x_2 > 0$*, can be written as follows:*

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

where

$$\begin{aligned} f_1(x_1, x_2) &= f_{BIII}(x_1; c, k_{13}) f_{BIII}(x_2; c, k_2) \\ &= k_{13} k_2 c^2 x_1^{-c-1} x_2^{-c-1} (1 + x_1^{-c})^{-k_{13}-1} (1 + x_2^{-c})^{-k_2-1}, \end{aligned}$$

$$\begin{aligned} f_2(x_1, x_2) &= f_{BIII}(x_1; c, k_1) f_{BIII}(x_2; c, k_{23}) \\ &= k_1 k_{23} c^2 x_1^{-c-1} x_2^{-c-1} (1 + x_1^{-c})^{-k_1-1} (1 + x_2^{-c})^{-k_{23}-1}, \end{aligned}$$

$$\begin{aligned} f_0(x) &= \frac{k_3}{k_{123}} f_{BIII}(x; c, k_{123}) \\ &= k_3 c \, x^{-c-1} (1 + x^{-c})^{-k_{123}-1}. \end{aligned}$$

Proof. Proof is given in the Appendix A. □

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

Theorem 2.3. *If* $(X_1, X_2) \sim BBIII(c, k_1, k_2, k_3)$, *then*

$$F_{X_1, X_2}(x_1, x_2) = \frac{k_{12}}{k_{123}} F_a(x_1, x_2) + \frac{k_3}{k_{123}} F_s(z),$$

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

$$F_s(z) = (1 + z^{-c})^{-k_{123}}$$

and

$$
\begin{aligned}
F_a(x_1, x_2) &= \frac{k_{123}}{k_{12}}(1 + x_1^{-c})^{-k_1}(1 + x_2^{-c})^{-k_2}(1 + z^{-c})^{-k_3} \\
&\quad - \frac{k_3}{k_{12}}(1 + z^{-c})^{-k_{123}},
\end{aligned}
$$

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

Proof. Proof is given in the Appendix A. □

Corollary 2.1. *The joint PDF of* X_1 *and* X_2 *can be written as follows:*

$$f_{X_1, X_2}(x_1, x_2) = \frac{k_{12}}{k_{123}} f_a(x_1, x_2) + \frac{k_3}{k_{123}} f_s(z),$$

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

$$f_s(z) = k_{123} c\, z^{-c-1}(1 + z^{-c})^{-(k_{123} - 1)},$$

and

$$
f_a(x_1, x_2) = \frac{k_{123}}{k_{12}} \begin{cases} f_{BIII}(x_1, c, k_{13}) f_{BIII}(x_2, c, k_2) & \text{if } x_1 < x_2 \\ f_{BIII}(x_1, c, k_1) f_{BIII}(x_2, c, k_{23}) & \text{if } x_1 > x_2. \end{cases}
$$

here $f_s(.)$ *and* $f_a(., .)$ *are the singular and the absolute continuous parts, respectively.* □

The following theorem provides the marginal and the conditional distributions of the BBIII distribution.

Theorem 2.4. *If* $(X_1, X_2) \sim BBIII(c, k_1, k_2, k_3)$

a) $X_1 \sim BBIII(c, k_{13})$, $X_2 \sim BBIII(c, k_{23})$,

b) *The conditional distribution of* X_1 *given* $X_2 = x_2$, *say* $F_{X_1 | X_2 = x_2}$, *is a convex combination of an absolute continuous distribution function and discrete distribution function.*

$$F_{X_1|X_2=x_2}(x_1) = p\,G(x_1) + (1-p)\,H(x_1),$$

where

$$G(x_1) = \frac{1}{p}\begin{cases} \dfrac{k_2}{k_{23}}(1+x_2^{-c})^{k_3}\,(1+x_1^{-c})^{-k_{13}} & \text{if}\quad x_1 < x_2 \\[2mm] (1+x_1^{-c})^{-k_1} - \dfrac{k_3}{k_{23}}(1+x_2^{-c})^{-k_1} & \text{if}\quad x_1 > x_2, \end{cases}$$

$$H(x_1) = \begin{cases} 0 & \text{if}\quad x_1 < x_2 \\ 1 & \text{if}\quad x_1 > x_2 \end{cases},$$

and

$$p = 1 - \frac{k_3}{k_{23}}(1+x_2^{-c})^{-k_1}.$$

c) The conditional distribution of X_1 given $X_2 \leq x_2$, say $F_{X_1|X_2\leq x_2}$, is an absolute continuos distribution function.

$$P(X_1 \leq x_1 | X_2 \leq x_2) = F_{X_1|X_2\leq x_2}(x_1)$$
$$= \begin{cases} (1+x_1^{-c})^{-k_{13}}(1+x_2^{-c})^{k_3} & \text{if}\quad x_1 < x_2 \\ (1+x_1^{-c})^{-k_1} & \text{if}\quad x_1 > x_2. \end{cases}$$

Proof. Proof is given in the Appendix A. □

Corollary 2.2. *Since the joint survival function and the joint CDF have the following relation*

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

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

3   Maximum Likelihood Estimation

In this section, we address the problem of computing the maximum likelihood estimators of the unknown parameters of the bivariate Burr III distribution using the EM algorithm. It is assumed that $\{(X_{11}, X_{21}), ..., (X_{1n}, X_{2n})\}$ is a random sample from

BBIII(c, k_1, k_2, k_3) and our problem is to estimate c, k_1, k_2 and k_3 from the given sample. We use the following notation

$$I_0 = \{i | X_{1i} = X_{2i} = X_i\}, \qquad I_1 = \{i | X_{1i} < X_{2i}\}, \qquad I_2 = \{i | X_{1i} > X_{2i}\}.$$

$|I_0| = n_0, |I_1| = n_1, |I_2| = n_2, n = n_0 + n_1 + n_2$.

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

$$\ell(c, k_1, k_2, k_3) = \sum_{i \in I_0} \ln f_0(x_i) + \sum_{i \in I_1} \ln f_1(x_{1i}, x_{2i}) + \sum_{i \in I_2} \ln f_2(x_{1i}, x_{2i}). \qquad (3.1)$$

We need to maximize (3.1) with respect to the four unknown parameters. This is clearly a four-dimensional problem. However, no explicit expressions are available for the MLEs. We need to solve four non-linear equations simultaneously, which may not be very simple. The maximization can be performed using a command like the nlminb routine in the R software (R Development Core Team, 2014). But, it is related to initial guesses. Therefore, we present an expectation-maximization (EM) algorithm similar to Kundu and Gupta (2009) to find the MLEs of parameters.

We look at the log-likelihood with information on ordering of U_1, U_2 and U_3 are missing. Hence, we treat this problem as a missing value problem. Assume that for the bivariate random vector (X_1, X_2), there is an associated random vector (Δ_1, Δ_2) as follows

$$\Delta_1 = \begin{cases} 1 & \text{if} \quad U_1 > U_3 \\ 3 & \text{if} \quad U_1 < U_3 \end{cases}, \quad \text{and} \quad \Delta_2 = \begin{cases} 2 & \text{if} \quad U_2 > U_3 \\ 3 & \text{if} \quad U_2 < U_3. \end{cases}$$

Therefore, if $X_1 = X_2$, then $\Delta_1 = \Delta_2 = 3$, but if $X_1 < X_2$ or $X_1 > X_2$, then (Δ_1, Δ_2) is missing. If $(x_{1i}, x_{2i}) \in I_1$, then the possible values of (Δ_1, Δ_2) are $(1, 2)$ or $(3, 2)$, similarly, if $(x_{1i}, x_{2i}) \in I_2$, the possible values of (Δ_1, Δ_2) are $(1, 2)$ or $(1, 3)$ with non-zero probability.

Now we provide the E-step and M-step of the EM algorithm. In the E-step we treat the observations belonging to I_0 as complete observations and keep them intact. If the observations belong to I_1 or I_2, we treat it as a missing observation. If $(x_{1i}, x_{2i}) \in I_1$, we form the pseudo-observations by fractioning (x_1, x_2) to two partially complete pseudo-observations of the form $(x_1, x_2, \mu_1(\gamma))$ and $(x_1, x_2, \mu_2(\gamma))$, respectively. Here $\gamma = (c, k_1, k_2, k_3)$ and the fractional mass $\mu_1(\gamma), \mu_2(\gamma)$ assigned to the pseudo-observation is the conditional probability that the random vector (Δ_1, Δ_2) takes the values $(1, 2)$

and $(3, 2)$, respectively, given $X_1 < X_2$. Similarly, if $(x_{1i}, x_{2i}) \in I_2$, we form the pseudo-observation of the form $(x_1, x_2, v_1(\gamma))$ and $(x_1, x_2, v_2(\gamma))$. Here the fractional mass $v_1(\gamma)$ or $v_2(\gamma)$ assigned to the pseudo-observation is the conditional probability that the random vector (Δ_1, Δ_2) takes the values $(1, 2)$ and $(1, 3)$, respectively, given $X_2 < X_1$. Since

$$P(U_1 < U_3 < U_2) = \frac{k_2 \, k_3}{k_{123} \, k_{13}}, \quad P(U_3 < U_1 < U_2) = \frac{k_1 \, k_2}{k_{123} \, k_{13}},$$

$$P(U_2 < U_3 < U_1) = \frac{k_1 \, k_3}{k_{123} \, k_{23}}, \quad P(U_3 < U_2 < U_1) = \frac{k_1 \, k_2}{k_{123} \, k_{23}},$$

Therefore,

$$\mu_1(\gamma) = P(U_1 < U_3 < U_2 | X_1 < X_2) = \frac{k_3}{k_{13}},$$

$$\mu_2(\gamma) = P(U_3 < U_1 < U_2 | X_1 < X_2) = \frac{k_1}{k_{13}},$$

$$\nu_1(\gamma) = P(U_2 < U_3 < U_1 | X_1 > X_2) = \frac{k_3}{k_{23}},$$

$$\nu_2(\gamma) = P(U_3 < U_2 < U_1 | X_1 > X_2) = \frac{k_2}{k_{23}}.$$

From now on, we write $\mu_1(\gamma), \mu_2(\gamma), \nu_1(\gamma)$ and $\nu_2(\gamma)$ as μ_1, μ_2, ν_1 and ν_2, respectively. The log-likelihood function of the pseudo data can be written as

$$
\begin{aligned}
\ell_{pseudo}(c, k_1, k_2, k_3) \;=\; & n_0 \log k_3 + n_0 \log c - (c + 1) \sum_{i \in I_0} \log x_i - (k_{123} + 1) \sum_{i \in I_0} \log(1 + x_i^{-c}) \\
+ \; & \mu_1 \Big[n_1 \log k_1 - (c + 1) \sum_{i \in I_1} \log x_{1i} - (k_1 + k_3 + 1) \sum_{i \in I_1} \log(1 + x_{1i}^{-c}) \Big] \\
+ \; & \mu_2 \Big[n_1 \log k_3 - (c + 1) \sum_{i \in I_1} \log x_{1i} - (k_1 + k_3 + 1) \sum_{i \in I_1} \log(1 + x_{1i}^{-c}) \Big] \\
+ \; & 2n_1 \log c + n_1 \log k_2 - (c + 1) \sum_{i \in I_1} \log x_{2i} - (k_2 + 1) \sum_{i \in I_1} \log(1 + x_{2i}^{-c}) \\
+ \; & \nu_1 \Big[n_2 \log k_2 - (c + 1) \sum_{i \in I_2} \log x_{2i} - (k_2 + k_3 + 1) \sum_{i \in I_2} \log(1 + x_{2i}^{-c}) \Big] \\
+ \; & \nu_2 \Big[n_2 \log k_3 - (c + 1) \sum_{i \in I_2} \log x_{2i} - (k_2 + k_3 + 1) \sum_{i \in I_2} \log(1 + x_{2i}^{-c}) \Big] \\
+ \; & 2n_2 \log c + n_2 \log k_1 - (c + 1) \sum_{i \in I_2} \log x_{1i} - (k_1 + 1) \sum_{i \in I_2} \log(1 + x_{1i}^{-c}).
\end{aligned}
$$

It can be simplified as

$$
\begin{aligned}
\ell_{pseudo}(c, k_1, k_2, k_3) &= (n_0 + 2n_1 + 2n_2)\log c + (n_1\mu_1 + n_2)\log k_1 + (n_2\nu_1 + n_1)\log k_2 \\
&+ (n_0 + n_2\nu_2 + n_1\mu_2)\log k_3 - (c+1)\Big[\sum_{i\in I_0}\log x_i + \sum_{i\in I_1\cup I_2}\log x_{1i} \\
&+ \sum_{i\in I_1\cup I_2}\log x_{2i}\Big] - (k_{123}+1)\sum_{i\in I_0}\log(1 + x_i^{-c}) \\
&- (k_2+1)\sum_{i\in I_1}\log(1 + x_{2i}^{-c}) - (k_1+1)\sum_{i\in I_2}\log(1 + x_{1i}^{-c}) \\
&- (k_{23}+1)\sum_{i\in I_2}\log(1 + x_{2i}^{-c}) - (k_{13}+1)\sum_{i\in I_1}\log(1 + x_{1i}^{-c}).
\end{aligned}
$$

Now the M- step involves the maximization of the $\ell_{pseudo}(c, k_1, k_2, k_3)$ with respect to c, k_1, k_2 and k_3 at each step. For fixed c, the maximization of $\ell_{pseudo}(c, k_1, k_2, k_3)$ occurs at

$$
\hat{k}_1(c) = \frac{n_2 + n_1\mu_1}{\sum_{i\in I_0}\log(1 + x_i^{-c}) + \sum_{i\in I_1\cup I_2}\log(1 + x_{1i}^{-c})}. \tag{3.2}
$$

$$
\hat{k}_2(c) = \frac{n_1 + n_2\nu_1}{\sum_{i\in I_0}\log(1 + x_i^{-c}) + \sum_{i\in I_1\cup I_2}\log(1 + x_{2i}^{-c})}. \tag{3.3}
$$

$$
\hat{k}_3(c) = \frac{n_0 + n_1\mu_2 + n_2\nu_2}{\sum_{i\in I_0}\log(1 + x_i^{-c}) + \sum_{i\in I_1}\log(1 + x_{1i}^{-c}) + \sum_{i\in I_2}\log(1 + x_{2i}^{-c})}. \tag{3.4}
$$

and $\hat{c}$ can be obtained by solving the following nonlinear equation

$$
h(c) = c,
$$

where

$$
\begin{aligned}
h(c) &= \Big[\sum_{i\in I_0}\log x_i + \sum_{i\in I_1\cup I_2}\log x_{1i} + \sum_{i\in I_1\cup I_2}\log x_{2i} \\
&- (\hat{k}_1(c) + \hat{k}_2(c) + \hat{k}_3(c) + 1)\sum_{i\in I_0}\frac{\ln x_i x_i^{-c}}{1 + x_i^{-c}} - (\hat{k}_2(c) + 1)\sum_{i\in I_1}\frac{\ln x_{2i}x_{2i}^{-c}}{1 + x_{2i}^{-c}} \\
&- (\hat{k}_1(c) + 1)\sum_{i\in I_2}\frac{\ln x_{1i}x_{1i}^{-c}}{1 + x_{1i}^{-c}} - (\hat{k}_2(c) + \hat{k}_3(c) + 1)\sum_{i\in I_2}\frac{\ln x_{2i}x_{2i}^{-c}}{1 + x_{2i}^{-c}} \\
&- (\hat{k}_1(c) + \hat{k}_3(c) + 1)\sum_{i\in I_1}\frac{\ln x_{1i}x_{1i}^{-c}}{1 + x_{1i}^{-c}}\Big]^{-1}(n_0 + n_1\mu_2 + n_2\nu_2). \tag{3.5}
\end{aligned}
$$

Hence, in order to solve a fixed point equations. (3.2)-(3.5), we start with an initial value of the parameter vector $(c^{(0)}, k_1^{(0)}, k_2^{(0)}, k_3^{(0)})^T$. Suppose at the i-th step the estimates of the parameters c, k_1, k_2 and k_3 are $c^{(i)}, k_1^{(i)}, k_2^{(i)}$ and $k_3^{(i)}$, respectively. Then the $(i + 1)$-th step of the EM algorithm is obtained using the following algorithm.

Step 1: Compute μ_1, μ_2, ν_1 and ν_2 using $c^{(i)}, k_1^{(i)}, k_2^{(i)}$ and $k_3^{(i)}$.

Step 2: Find $c^{(i+1)}$ by solving (3.5).

Step 3: Once $c^{(i+1)}$ is obtained, compute $k_1^{(i+1)} = \hat{k}_1(c^{(i+1)})$, $k_2^{(i+1)} = \hat{k}_2(c^{(i+1)})$ and $k_3^{(i+1)} = \hat{k}_3(c^{(i+1)})$.

This version of the EM algorithm is called the ECM (expectation-conditional maximization) algorithm. Steps 1-3 describe one iteration of the algorithm.

4 Numerical Experiments

In this section, we present some simulation results to verify how the proposed EM algorithm works for different sample sizes and different parameter values. This simulation study data in R software was generated using the "actuar" package. We assume that the pair of random variables (X_1, X_2) is distributed as BBIII(c, k_1, k_2, k_3).

In Appendix B, we have provided the observed Fisher information matrix using the idea of Louis (1982). This matrix can be inverted to obtain the asymptotic covariance matrix.

The average estimate (AVEST), mean squared error (MSE) and average confidence lengths (ACL) are reported in Table 1 for $k_1 = k_2 = k_3 = 0.8$ and varying parameter $c = 0.6, 0.7, 0.8$ and the sample size $n = 50, 75, 100$ based on 1000 replications. In order to implement the proposed EM algorithm, we have used the $\epsilon = 10^{-5}$ and initial guesses as 0.5 for all the parameters. We have tried other initial guesses also, but the average estimates, the corresponding MSEs and confidence intervals are same.

Some of the salient features of the numerical experiments based on Table 1 are given below.

(i) We observe that the average estimators of all the four parameters c, k_1, k_2 and k_3 are very close to the true values of all choices of the parameter c.

(ii) The mean square error of the estimators decreases with increase in sample size.

Table 1: Average estimates, mean squared errors and average confidence interval.

Parameter	Iteration								
	50			75			100		
$c = 0.6$	AVEST	MSE	ACL	AVEST	MSE	ACL	AVEST	MSE	ACL
c	0.6485	$2.354e^{-3}$	(0.6464, 0.6506)	0.6481	$2.263\,e^{-4}$	(0.6479, 0.6524)	0.6475	$2.254\,e^{-3}$	(0.6459, 0.6489)
k_1	0.8162	$2.624e^{-4}$	(0.8060, 0.8263)	0.8135	$1.812\,e^{-4}$	(0.8052, 0.8217)	0.8109	$4.380\,e^{-4}$	(0.8137, 0.8281)
k_2	0.8260	$6.775\,e^{-4}$	(0.8158, 0.8363)	0.8253	$5.507\,e^{-4}$	(0.8190, 0.8357)	0.8195	$1.794\,e^{-4}$	(0.8123, 0.8266)
k_3	0.8478	$2.228e^{-4}$	(0.8389, 0.8567)	0.8473	$2.241\,e^{-3}$	(0.8401, 0.8546)	0.8468	$2.193\,e^{-4}$	(0.8405, 0.8531)
$c = 0.7$	AVEST	MSE	ACL	AVEST	MSE	ACL	AVEST	MSE	ACL
c	0.8239	$1.535e^{-2}$	(0.8214, 0.8264)	0.8225	$1.401e^{-2}$	(0.8205, 0.8286)	0.8225	$1.500e^{-2}$	(0.8207, 0.8243)
k_1	0.8257	$6.594e^{-4}$	(0.8154, 0.8359)	0.8234	$5.467e^{-4}$	(0.8150, 0.8317)	0.8204	$4.241e^{-4}$	(0.8131, 0.8377)
k_2	0.8350	$1.227e^{-3}$	(0.8216, 0.8454)	0.8345	$1.128\,e^{-3}$	(0.8280, 0.8449)	0.8291	$8.449\,e^{-4}$	(0.8218, 0.8364)
k_3	0.8806	$6.495\,e^{-3}$	(0.8709, 0.8902)	0.8799	$6.385\,e^{-3}$	(0.8721, 0.8877)	0.8797	$6.348e^{-3}$	(0.8729, 0.8865)
$c = 0.8$	AVEST	MSE	ACL	AVEST	MSE	ACL	AVEST	MSE	ACL
c	1.0238	$5.009\,e^{-2}$	(1.0208, 1.0268)	1.0220	$5.004\,e^{-2}$	(1.0246, 1.0295)	1.0219	$4.927\,e^{-2}$	(1.0198, 1.0241)
k_1	0.8288	$8.326\,e^{-4}$	(0.8185, 0.8392)	0.8269	$7.266\,e^{-4}$	(0.8185, 0.8354)	0.8253	$6.241\,e^{-3}$	(0.8262, 0.8409)
k_2	0.8378	$1.426\,e^{-3}$	(0.8273, 0.8482)	0.8353	$1.125\,e^{-3}$	(0.8308, 0.8479)	0.8323	$8.449\,e^{-4}$	(0.8249, 0.8396)
k_3	0.9037	$1.011\,e^{-3}$	(0.8935, 0.9139)	0.9029	$1.010\,e^{-2}$	(0.8945, 0.9112)	0.9028	$1.042\,e^{-3}$	(0.8956, 0.9104)

(iii) When n = 100, average length of confidence intervals for all the parameters is considerably lower compared to the case when n = 50.

(iv) The value of c makes no change in its numerical value.

The results of other numerical experiments are given in Figs. 1 and 2, for $c = k_1 = k_2 = 0.8$ and varying parameter $k_3 = 0.6, 0.7, 0.8$ and sample size $n = 50, 75, 100$. The results are similar to the ones discussed above for in Table 1.
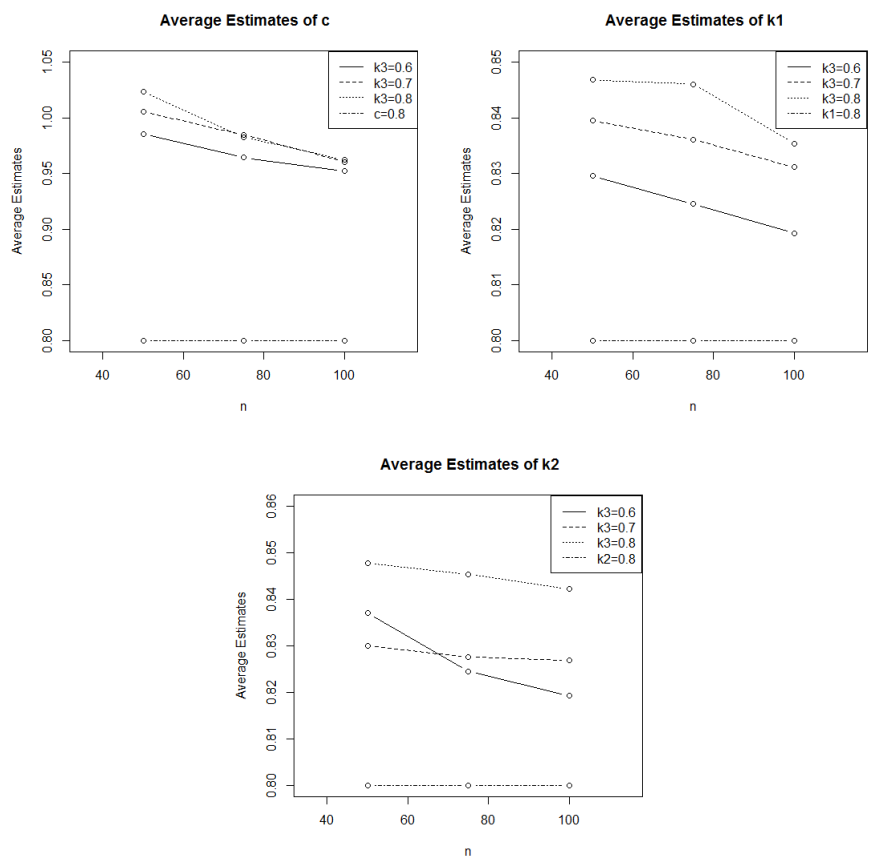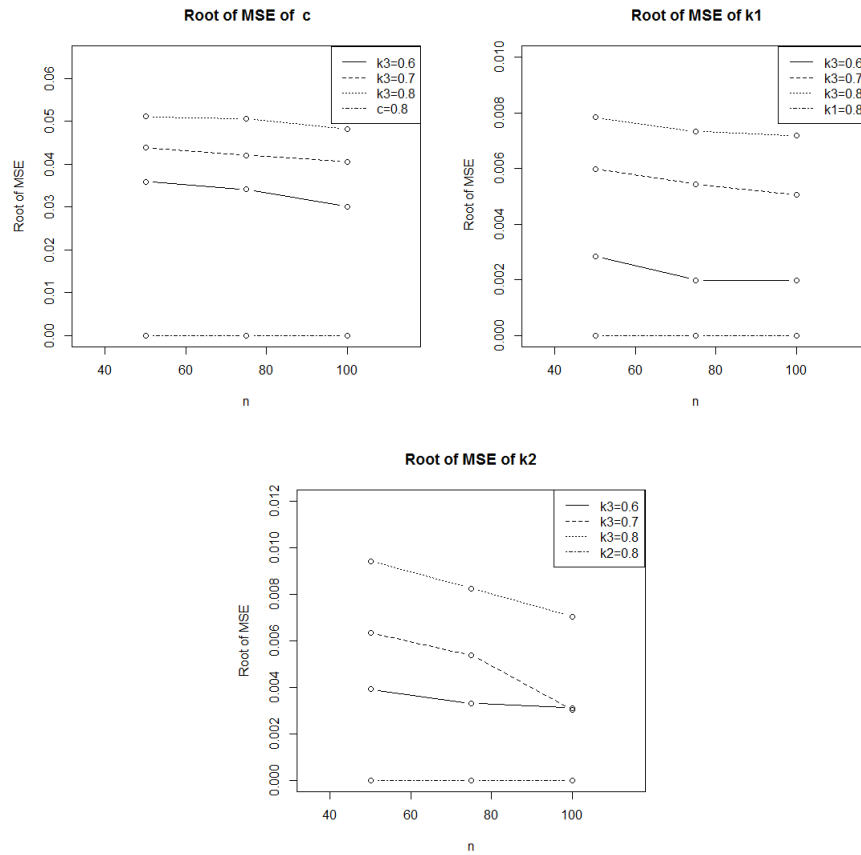

Figure 1: Average estimates of c, k_1 and k_2.

Root of MSE of c

Root of MSE of k1

Root of MSE of k2

Figure 2: Root of MSE of c, k_1 and k_2.

5 Data Analysis

The data set analyzed in this section, are from the American Football (National Football League) matches played on three consecutive weekends in 1986. It has been originally published in 'Washington Post' and it is also available in Csorgo and Welsh (1989). Kundu and Gupta (2010), Jamalizadeh and Kundu (2013), and Balakrishnan and Shiji (2014) analyzed this data. In this bivariate data set, X_1 represents the game time to the first points scored by kicking the ball between the goal posts, and X_2 represents the

game time to the first points scored by moving the ball into the end zone. The variables X_1 and X_2 have the following structure: (i)$X_1 < X_2$ means that the first score is a field goal, (ii) $X_1 = X_2$ means the first score is a converted touchdown, (iii) $X_1 > X_2$ means the first score is an unconverted touchdown or safety. In this case, the ties are exact because no 'game time' elapses between a touchdown and a point-after conversion attempt. It should be noted that the possible scoring times are restricted by the duration of the game, but it has been ignored similarly as in Csorgo and Welsh (1989).

Before analyzing the data using the proposed EM algorithm, we fit the Burr III distribution to X_1, X_2 and max(X_1, X_2), separately. The MLEs of the parameters of the Burr III distribution for X_1, X_2 and max(X_1, X_2) are (1.090, 5.152), (0.959, 5.244) and (0.952, 5.239), respectively. The Kolmogorov - Smirnov distances between the fitted distribution and the empirical distribution function and the corresponding p-values (in brackets) for X_1, X_2 and max(X_1, X_2) are 0.186 (0.11), 0.196 (0.09) and 0.192 (0.10), respectively. Based on the p-values Burr III distribution can be used for analyzing X_1, X_2 and max(X_1, X_2).

To start the EM algorithm we need some initial guesses of the unknown parameters, we used the idea of Kundu and Gupta (2009). For c, we suggest to take the average values of 1.090, 0.959 and 0.952, i.e. 0.996. Assuming the initial guess of c as 0.996, solving three equations in three unknowns for k's, we get the initial guess values of k_1, k_2 and k_3 as 1.079, 2.257 and 3.731, respectively. Using these initial values the EM algorithm converges to the same values. We have computed the MLEs using direct maximization and have obtained the same estimates. Therefore, the EM algorithm works quite well in this case. The average estimate, standard error and confidence intervals of the parameters are reported in Table 2.

Table 2: Average Estimate, Standard Error and Confidence Intervals for the real data set.

Parameter	Average Estimate	Standard Error	Confidence Interval
c	0.8087	0.1870	(0.6686, 0.9487)
k_1	1.6105	0.5314	(0.8905, 2.3304)
k_2	2.2410	0.4475	(1.1675, 3.3145)
k_3	3.2129	0.5170	(1.9637, 4.4622)

6   Conclusions

In this paper, we have proposed the bivariate Burr type III distribution function whose marginals are Burr type III distributions. It is observed that the BBIII distribution has an absolute continuous part and a singular part. This model has been obtained using a similar technique as of Kundu and Gupta (2009). Since the joint distribution function and the joint density function are in closed forms, this distribution can be used in practice for non-negative and positively correlated random variables. Since the maximum likelihood estimators of the unknown parameters cannot be obtained in closed form, we suggest the use of expectation-conditional maximization algorithm. We have looked at the pseudo-likelihood with information on the ordering of U_1, U_2 and U_3 missing and treated this problem as a missing value problem. The simulation results indicate that the ECM algorithm performs very well for different sample sizes and different parameter values that we have studied. We have also constructed the asymptotic confidence intervals using the idea of Louis (1982) and observed that the asymptotic confidence intervals give accurate results and hence can be used for testing purposes. Finally, one data set analyzed for illustrative purposes.

Acknowledgements

References

Abdel-Ghaly, A. A. Al-Dayian, G. R. and Al-Kashkari, F. H. (1997), The use of Burr Type XII distribution on software reliability growth modelling. *Microelectronics Reliability*, **37**(2), 305–313.

Al-Khedhairi, A. and El-Gohary, A. (2008), A new class of bivariate Gompertz distributions. *International Journal of Mathematics Analysis*, **2**(5), 235–253.

Balakrishna, N. and Shiji, K. (2014), On a class of bivariate exponential distributions. *Statistics and Probability Letters*, **85**, 153–160.

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*, **67**, 927–929.

Burr, I. W. (1942), Cumulative frequency distributions. *Annals of Mathematical Statistics*, *13*, 215-232.

Chernobai, A. S. Fabozzi, F. J. and Rachev, S. T. (2007), *Operational Risk: a guide to Basel II capital requirements, models and analysis*, New York. John Wiley and Sons.

Csorgo, S. and Welsh, A. (1989), Testing for exponential and Marshall-Olkin distributions. *Journal of Statistical Planning and Inference*, **23**(3), 287–300.

Dagum, C. A. (1977), New model of personal income distribution: specification and estimation. *Applied Economics*, **30**, 413–437.

Dempster, A. P. Laird, N. M. and Rubin, D. B. (1977), Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society*, **39**, 1–38.

El-Sherpieny, E. A. Ibrahim, S. A. and Bedar, R. E. (2013), A new bivariate generalized Gompertz distribution. *Asian Journal of Applied Sciences*, **1**(4), 2321–0893.

Gove, J. H. Ducey, M. J. Leak, W. B. and Zhang, L. (2008), Rotated sigmoid structures in managed uneven-aged northern hardwork stands: a look at the Burr Type III distribution. *Foresty* , **81**(2), 161–176.

Houggard, P. Harvald, B. and Holm, N. V. (1992), Measuring the similarities between the lifetimes of adult Danish twins born between 18811930. *Journal of the American Statititical Association*, **87**, 17–24.

Jamalizadeh, A. and Kundu, D. (2013), Weighted Marshall-Olkin bivariate exponential distribution. *Statistics*, **47**(5), 917–928.

Kundu, D. and Dey, A. K. (2009), Estimating the parameters of the Marshall-Olkin Bivariate Weibull distribution by EM algorithm. *Computational Statistics and Data Analysis*, **53** (4), 956–965.

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

Kundu, D. and Gupta, R. D. (2010), Modified Sarhan-Balakrishnan singular bivariate distribution. *Journal of Statistical Planning and Inference*, **140**(2), 526–538.

Lin, D. Y. Sun, W. and Ying, Z. (1999), Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika*, **86**, 59–70.

Lindsay, S. R. Wood, G. R. and Woollons, R. C. (1996), Modeling the diameter distribution of forest stands using the Burr distribution. *Journal of Applied Statistics*, **23**(6), 609–619.

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

Marshall, A. W. and Olkin, I. (1967), Multivariate exponential distribution. *Journal of the American Statistical Association*, **62**, 30–44.

Mielke, P. W. (1973), Another family of distributions for describing and analyzing precipitation data. *Journal of Applied Meteorology and Climatology*, **12**, 275–280.

Nadarajah, S. and Kotz, S. (2006), q exponential is a Burr distribution. *Physics Letters A*, **359**(6), 451–456.

Nadarajah, S. and Kotz, S. (2007), On the alternative to the Weibull function. *Engineering Fracture Mechanics*, **74**, 577–579.

Nandi, S. and Dewan, I. (2010), An EM algorithm for estimating the parameters of bivariate Weibull distribution under random censoring. *Computational Statistics and Data Analysis*, **54**, 1559–1569.

Sarhan, A. and Balakrishnan, N. (2007), A new class of bivariate distribution and its mixture. *Journal of the Multivariate Analysis*, **98**, 1508–1527.

Sherrick, B. J. Garcia, P. and Tirupattur, V. (1996), Recovering probabilistic information from option markets: tests of distributional assumptions. *Journal of Futures Markets*. **16**(5), 545–560.

Tejeda, H. A. and Goodwin, B. K., (2008), Modelling Crop Price through a Burr Distribution and analysis of correlation between crop prices and yields using copula method. *Agriculture and Applied Economics Association, Orlanndo, FL, USA*, **83**(3), 643–649.

Wingo, D. R. (1983), Maximum Likelihood Methods for fitting the Burr type XII distribution of life test data. *Journal of Biomedical Science*, *25*(1), 77–84.

Wingo, D. R. (1993), Maximum Likelihood Methods for fitting the Burr type XII distribution to multiply (progressively) censored life test data, *Metrika*, **40**(1), 203–210.

Apendix A

Proof of Theorem 2.1.

$$
\begin{aligned}
F_{X_1, X_2}(x_1, x_2) &= P(X_1 \leq x_1, X_2 \leq x_2) \\
&= P(\max\{U_1, U_3\} \leq x_1, \max\{U_2, U_3\} \leq x_2) \\
&= P(U_1 \leq x_1, U_3 \leq x_1, U_2 \leq x_2, U_3 \leq x_2) \\
&= P(U_1 \leq x_1)P(U_2 \leq x_2)P(U_3 \leq \min\{x_1, x_2\}) \\
&= F_{BIII}(x_1; c, k_1)F_{BIII}(x_2; c, k_2)F_{BIII}(z; c, k_3) \\
&= (1 + x_1^{-c})^{-k_1}(1 + x_2^{-c})^{-k_2}(1 + z^{-c})^{-k_3}.
\end{aligned}
$$

□

Proof of Theorem 2.2. The expressions for $f_1(.,.)$ and $f_2(.,.)$ can be obtained from $\frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1, X_2}(x_1, x_2)$ for $x_1 < x_2$ and for $x_1 > x_2$, respectively. But $f_0(.)$ can not be obtained in the same way. Now using the facts

$$
\int_0^\infty \int_0^{x_2} f_1(x_1, x_2) \, dx_1 \, dx_2 + \int_0^\infty \int_0^{x_1} f_2(x_1, x_2) \, dx_2 \, dx_1 + \int_0^\infty f_0(x) dx = 1
$$

$$
\begin{aligned}
\int_0^\infty \int_0^{x_2} f_1(x_1, x_2) \, dx_1 \, dx_2 &= \int_0^\infty f(x_2; c, k_2) \int_0^{x_2} f(x_1; c, k_{13}) \, dx_1 \, dx_2 \\
&= \int_0^\infty f(x_2; c, k_2)F(x_2; c, k_{13}) \, dx_2 \\
&= k_2 \int_0^\infty c \, x^{-c-1}(1 + x^{-c})^{-k_{123}-1} dx,
\end{aligned}
$$

and

$$
\begin{aligned}
\int_0^\infty \int_0^{x_1} f_2(x_1, x_2) \, dx_2 \, dx_1 &= \int_0^\infty f(x_1; c, k_1) \int_0^{x_1} f(x_2; c, k_{23}) \, dx_2 \, dx_1 \\
&= \int_0^\infty f(x_1; c, k_1)F(x_1; c, k_{23}) \, dx_1 \\
&= k_1 \int_0^\infty c \, x^{-c-1}(1 + x^{-c})^{-k_{123}-1} dx,
\end{aligned}
$$

then

$$\int_0^\infty f_0(x)\,dx = \int_0^\infty k_3\,c\,x^{-c-1}(1+x^{-c})^{-k_{123}-1}dx.$$

Therefore, the result follows. □

Proof of Theorem 2.3. To find $F_a(x_1, x_2)$ from $F_{X_1,X_2}(x_1, x_2) = pF_a(x_1, x_2) + (1-p)F_s(z)$, $0 \le p \le 1$, we compute

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

from which p may be obtained as

$$p = \int_0^\infty \int_0^{x_2} f_1(x_1, x_2)\,dx_1\,dx_2 + \int_0^\infty \int_0^{x_1} f_2(x_1, x_2)\,dx_2\,dx_1 = \frac{k_{12}}{k_{123}},$$

and

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

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

Alternatively, probabilistic arguments are also can be provided as follows. Suppose A is the following event: $A = \{U_1 < U_3\} \cap \{U_2 < U_3\}$, then $p(A) = \frac{k_3}{k_{123}}$. Therefore,

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2|A)P(A) + P(X_1 \le x_1, X_2 \le x_2|A^c)P(A^c).$$

Moreover, for z as defined before

$$P(X_1 \le x_1, X_2 \le x_2|A) = (1 + z^{-c})^{-k_{123}},$$

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

Clearly, $(1+z^{-c})^{-k_{123}}$ is the singular part as its mixed second partial derivative is zero when $x_1 \neq x_2$ and $P(X_1 \le x_1, X_2 \le x_2|A^c)$ is the absolute continuous part as its mixed partial derivative is a density function.

□

Proof of Theorem 2.4.   a)

$$
\begin{aligned}
P(X_1 \le x) &= P(\max\{U_1, U_3\} \le x) \\
&= P(U_1 \le x, U_3 \le x) \\
&= P(U_1 \le x)P(U_3 \le x) \\
&= F_{BIII}(x; c, k_1)F_{BIII}(x; c, k_3) \\
&= (1 + x^{-c})^{-k_{13}}.
\end{aligned}
$$

and

$$
\begin{aligned}
P(X_2 \le x) &= P(\max\{U_2, U_3\} \le x) \\
&= P(U_2 \le x, U_3 \le x) \\
&= P(U_2 \le x)P(U_3 \le x) \\
&= F_{BIII}(x; c, k_2)F_{BIII}(x; c, k_3) \\
&= (1 + x^{-c})^{-k_{23}}.
\end{aligned}
$$

b)

$$
\begin{aligned}
F_{X_1|X_2=x_2}(x_1) &= \int_0^{x_1} f_{X_1|X_2}(x|x_2)dx = \int_0^{x_1} \frac{f_{X_1,X_2}(x,x_2)}{f_{X_2}(x_2)}dx \\[2mm]
&= \begin{cases} \int_0^{x_1} \dfrac{f_1(x,x_2)}{f_{X_2}(x_2)}dx & \text{if } x_1 < x_2 \\[3mm] \int_0^{x_2} \dfrac{f_2(x,x_2)}{f_{X_2}(x_2)}dx & \text{if } x_1 > x_2 \end{cases} \\[4mm]
&= \begin{cases} \dfrac{k_2}{k_{23}}(1 + x_2^{-c})^{k_3} \int_0^{x_1} k_{13}cx^{-c-1}(1 + x^{-c})^{-k_{13}-1}\,dx & \text{if } x_1 < x_2 \\[3mm] \int_0^{x_1} k_1 cx^{-c-1}(1 + x^{-c})^{-k_1-1}\,dx & \text{if } x_1 > x_2 \end{cases} \\[4mm]
&= \begin{cases} \dfrac{k_2}{k_{23}}(1 + x_2^{-c})^{k_3}(1 + x_1^{-c})^{-k_{13}} & \text{if } x_1 < x_2 \\[3mm] (1 + x_1^{-c})^{-k_1} & \text{if } x_1 > x_2. \end{cases}
\end{aligned}
$$

c)

$$P(X_1 \le x_1 | X_2 \le x_2) = \frac{P(X_1 \le x_1, X_2 \le x_2)}{P(X_2 \le x_2)}$$

$$= \begin{cases} \dfrac{(1 + x_1^{-c})^{-k_{13}}(1 + x_2^{-c})^{-k_2}}{(1 + x_2^{-c})^{-k_{23}}} & \text{if } x_1 < x_2 \\[3ex] \dfrac{(1 + x_1^{-c})^{-k_1}(1 + x_2^{-c})^{-k_{23}}}{(1 + x_2^{-c})^{-k_{23}}} & \text{if } x_1 > x_2 \end{cases}$$

$$= \begin{cases} (1 + x_1^{-c})^{-k_{13}}(1 + x_2^{-c})^{k_3} & \text{if } x_1 < x_2 \\ (1 + x_1^{-c})^{-k_1} & \text{if } x_1 > x_2 \end{cases}.$$

$\square$

Apendix B: Observed Fisher Information Matrix

In this part, we present the observed Fisher information matrix obtained using the idea of Louis (1982), which is used when the EM algorithm is applied to obtain the MLEs in the case of incomplete data problem. The observed information matrix can then be inverted to obtain the asymptotic covariance matrix of the MLEs determined from the EM algorithm. Let S denote the derivative vector and H the Hessian matrix of the pseudo-log-likelihood function defined in (3.1). The observed Fisher information matrix is given by $H - SS^T$. The elements of vector $S = (S_1, S_2, S_3, S_4)^T$, are as follows

$$S_1 = \frac{n_0 + 2n_1 + 2n_2}{c} - \sum_{i \in I_0} \log x_i - \sum_{i \in I_1 \cup I_2} \log x_{1i} - \sum_{i \in I_1 \cup I_2} \log x_{2i}$$

$$+ (k_{123} + 1) \sum_{i \in I_0} \frac{\ln x_i \, x_i^{-c}}{1 + x_i^{-c}} + (k_2 + 1) \sum_{i \in I_1} \frac{\ln x_{2i} \, x_{2i}^{-c}}{1 + x_{2i}^{-c}} + (k_1 + 1) \sum_{i \in I_2} \frac{\ln x_{1i} \, x_{1i}^{-c}}{1 + x_{1i}^{-c}}$$

$$+ (k_{23} + 1) \sum_{i \in I_2} \frac{\ln x_{2i} \, x_{2i}^{-c}}{1 + x_{2i}^{-c}} + (k_{13} + 1) \sum_{i \in I_1} \frac{\ln x_{1i} \, x_{1i}^{-c}}{1 + x_{1i}^{-c}}.$$

$$S_2 = \frac{n_2 + n_1 \mu_1}{k_1} - \sum_{i \in I_0} \log(1 + x_i^{-c}) - \sum_{i \in I_1 \cup I_2} \log(1 + x_{1i}^{-c}).$$

$$S_3 = \frac{n_1 + n_2\nu_1}{k_2} - \sum_{i\in I_0} \log(1 + x_i^{-c}) - \sum_{i\in I_1\cup I_2} \log(1 + x_{2i}^{-c}).$$

$$S_4 = \frac{n_0 + n_1\mu_2 + n_2\nu_2}{k_3} - \sum_{i\in I_0} \log(1 + x_i^{-c}) - \sum_{i\in I_1} \log(1 + x_{1i}^{-c}) - \sum_{i\in I_2} \log(1 + x_{2i}^{-c}).$$

The Hessian matrix H is symmetric, so $H_{ij} = H_{ji}, i > j$ and in the following, the elements are given:

$$
\begin{aligned}
H_{11} =\ & -\frac{n_0 + 2n_1 + 2n_2}{c^2} \\
& - (k_{123} + 1)\Big[\sum_{i\in I_0} \frac{\ln^2 x_i\, x_i^{-c}}{1 + x_i^{-c}} - \Big[\sum_{i\in I_0} \frac{\ln x_i\, x_i^{-c}}{1 + x_i^{-c}}\Big]^2 \Big] - (k_2 + 1)\Big[\sum_{i\in I_1} \frac{\ln^2 x_{2i}\, x_{2i}^{-c}}{1 + x_{2i}^{-c}} - \Big[\sum_{i\in I_1} \frac{\ln x_{2i}\, x_{2i}^{-c}}{1 + x_{2i}^{-c}}\Big]^2 \Big] \\
& - (k_1 + 1)\Big[\sum_{i\in I_2} \frac{\ln^2 x_{1i}\, x_{1i}^{-c}}{1 + x_{1i}^{-c}} - \Big[\sum_{i\in I_2} \frac{\ln x_{1i}\, x_{1i}^{-c}}{1 + x_{1i}^{-c}}\Big]^2 \Big] - (k_{23} + 1)\Big[\sum_{i\in I_2} \frac{\ln^2 x_{2i}\, x_{2i}^{-c}}{1 + x_{2i}^{-c}} - \Big[\sum_{i\in I_2} \frac{\ln x_{2i}\, x_{2i}^{-c}}{1 + x_{2i}^{-c}}\Big]^2 \Big] \\
& - (k_{13} + 1)\Big[\sum_{i\in I_1} \frac{\ln^2 x_{1i}\, x_{1i}^{-c}}{1 + x_{1i}^{-c}} - \Big[\sum_{i\in I_1} \frac{\ln x_{1i}\, x_{1i}^{-c}}{1 + x_{1i}^{-c}}\Big]^2 \Big].
\end{aligned}
$$

$$H_{22} = -\frac{n_2 + n_1\mu_1}{k_1^2}.$$

$$H_{33} = -\frac{n_1 + n_2\nu_2}{k_2^2}.$$

$$H_{44} = -\frac{n_0 + n_1\mu_2 + n_2\nu_1}{k_3^2}.$$

$$H_{12} = \sum_{i\in I_0} \frac{\ln x_i\, x_i^{-c}}{1 + x_i^{-c}} + \sum_{i\in I_1\cup I_2} \frac{\ln x_{1i}\, x_{1i}^{-c}}{1 + x_{1i}^{-c}}.$$

$$H_{13} = \sum_{i\in I_0} \frac{\ln x_i\, x_i^{-c}}{1 + x_i^{-c}} + \sum_{i\in I_1\cup I_2} \frac{\ln x_{2i}\, x_{2i}^{-c}}{1 + x_{2i}^{-c}}.$$

$$H_{14} = \sum_{i\in I_0} \frac{\ln x_i\, x_i^{-c}}{1 + x_i^{-c}} + \sum_{i\in I_1} \frac{\ln x_{1i}\, x_{1i}^{-c}}{1 + x_{1i}^{-c}} + \sum_{i\in I_2} \frac{\ln x_{2i}\, x_{2i}^{-c}}{1 + x_{2i}^{-c}}.$$

$$H_{23} = 0, \qquad H_{23} = 0, \qquad H_{34} = 0.$$