STATISTICAL INFERENCE FOR THE BURR TYPE III DISTRIBUTION UNDER TYPE II CENSORED DATA

Ö. ALTINDAĞ, M. N. ÇANKAYA, A. YALÇINKAYA, AND H. AYDOĞDU

ABSTRACT. In this study, estimation and prediction problems for the Burr type III distribution under type II censored data are considered. Maximum likelihood and maximum product spacing estimation methods are used to estimate model parameters. EM algorithm is employed to obtain maximum likelihood estimates. Unobserved future order statistics are predicted with best unbiased prediction method. A simulation study is carried out to exhibit performance of the estimation methods. Further, a numerical example is presented to illustrate the usefulness of the Burr III distribution.

Keywords: Burr III distribution, type II censoring, maximum likelihood, EM algorithm, maximum product spacing, inference

1. INTRODUCTION

Irwing W. Burr [1] introduced a distribution family which includes twelve distribution types. Among these distributions, Burr type III (BIII) and type XII (BXII) have attracted considerable attention in the literature. It should be noted that they are reciprocal. Lindsay [2] used BIII distribution to model diameter distributions of forest stands as an alternative to Weibull distribution. Shao [3] has modelled the NOEC toxicity data with BIII distribution. Zimav et al. [4] showed that BXII distribution is a useful model for failure data in reliability studies. Some other usages of the BIII and BXII distributions can be found in [5-10]. However, BIII distribution is more flexible to fit data than BXII in the sense that it provides wider skewness and kurtosis region. One can refer to Rodriguez [11] and Tadikamalla [12] for the details.

When investigating lifetime characteristics of a unit, it is aimed to fit failure data with a flexible distribution. Since BIII distribution can properly approximates many classical lifetime models such as Weibull, gamma and lognormal [13], it is

worthwhile to model failure data with BIII distribution. So, BIII distribution plays an important role in lifetime analysis. In survival and reliability context, some censoring schemes are applied due to time and cost restrictions. A commonly used censoring scheme is type II censoring in which only the first $r \leq n$ failures among the n units are observed. Wingo [14] studied the maximum likelihood estimation for BXII distribution under type II censoring. Wang and Cheng [15] and Sayyareh [16] derived the expectation-maximization (EM) procedure for BXII distribution under multiple censoring and type II censoring, respectively. It should be pointed out that once the parameters of BXII distribution have been estimated from a random sample, the parameters of BIII distribution have already been obtained because of their reciprocal property, vice versa. However, this relation does not hold for the type II censored data though it holds for the complete sample. This fact can be explained as follows.

Let the random variable Y has BXII distribution, then $X = 1/Y$ has BIII distribution. So, when a complete sample from BIII distribution is observed, the estimates of parameters can be obtained from the estimation procedure of BXII model by replacement of each observation with its reciprocal. Let type II censored sample from BIII distribution has been observed. It is clear that, the inverses of these observations do not correspond to the type II censored sample from BXII distribution. Thus, estimation procedure for the parameters of BIII distribution under type II censoring has to be derived. Moradi et al. [17] have considered so called exponentiated-BIII distribution under type II censored data. However, extending the BIII distribution with exponentiation is meaningless since the distribution they have derived corresponds to reparametrized version of the BIII distribution. Azizi et al. [18] and Singh et al. [19] have studied parameter estimation for the BIII distribution under type II hybrid censoring and type II progressive censoring, respectively. It should be noted that these are generalizations of the classic type II censoring. In this study, estimation problem for the parameters of BIII distribution and prediction of the unobserved future order statistics under type II censored data are considered. Unlike the previous studies, the maximum product spacing (MPS) method is also taken into account as an alternative.

Remainder of the article is organized as the following. In Section 2, ML and MPS estimation methods for the parameters are considered. The EM algorithm is used to obtain ML estimates. Asymptotic variances and covariance of the ML estimates are calculated with missing information principle (MIP) introduced by Louis [20]. The Section 3 is divided to the prediction of unobserved future order statistics for the type II censored sample. A comprehensive simulation study is carried out to evaluate the performances of ML and MPS estimation methods with different parameter, sample size and censoring settings in Section 4. An illustrative real data example is studied in Section 5. Finally, the study is concluded in Section 6.

2. PARAMETER ESTIMATION

The cumulative distribution function (cdf) and the probability density function (pdf) of the BIII distribution are given by

$$F(x; c, k) = \left(1 + x^{-c}\right)^{-k}, \ x \geq 0; \ c, k > 0, \tag{2.1}$$

$$f(x; c, k) = kcx^{-(c+1)}\left(1 + x^{-c}\right)^{-(k+1)}, \ x \geq 0; \ c, k > 0, \tag{2.2}$$

respectively. The parameters c and k are the shape parameters of the distribution. Suppose that $X_1, X_2, ..., X_n$ be independent and identically BIII distributed random variables representing the lifetimes of n independent units. For type II censoring case, only the first prefixed $r \ (r \leq n)$ failures, say $X_{1:n}, X_{2:n}, ..., X_{r:n}$, are observed. These failures correspond to the first r order statistics of the random sample $X_1, X_2, ..., X_n$. To estimate the parameters of the BIII distribution we will consider ML and MPS methods, respectively.

2.1. **ML method.** The likelihood function of the BIII distribution under type II censored data is given by

$$
\begin{aligned}
L(c, k; \boldsymbol{x}) &= \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(x_{i:n}; c, k)\left[1 - F(x_{r:n}; c, k)\right]^{n-r} \\
&= \frac{n!}{(n-r)!} k^r c^r \left[1 - \left(1 + x_{r:n}^{-c}\right)^{-k}\right]^{n-r} \prod_{i=1}^{r} x_{i:n}^{-(c+1)}\left(1 + x_{i:n}^{-c}\right)^{-(k+1)}
\end{aligned}
$$

$$\tag{2.3}$$

where $\boldsymbol{x} = (x_{1:n}, x_{2:n}, ..., x_{r:n})$. It is generally tractable to work with log-likelihood function instead of the likelihood function. Therefore, the log-likelihood function is

$$
\begin{aligned}
\log L(c, k; \boldsymbol{x}) &= \text{constant} + r(\log c + \log k) + (n-r)\log\left(1 - \left(1 + x_{r:n}^{-c}\right)^{-k}\right) \\
&\quad - (c+1)\sum_{i=1}^{r}\log(x_{i:n}) - (k+1)\sum_{i=1}^{r}\log\left(1 + x_{i:n}^{-c}\right). \tag{2.4}
\end{aligned}
$$

The ML estimates of c and k are the arguments maximizing the $\log L$ function. For the Equation (2.4), the ML estimators cannot be obtained in an analytical form. One way to obtain ML estimates of c and k is that the first partial derivatives of the $\log L$ function with respect to c and k are equated to zero and these nonlinear equations are solved by a numerical procedure such as Newton-Raphson (NR). However, when employing the NR method some errors such as non-convergence or convergence to wrong value can occur. This may result in incorrect estimates. Dempster et al. [21] developed a method called EM to calculate the ML estimates in the presence of missing data. The EM algorithm is based on repeating the E and M steps until convergence occurs. Since the log-likelihood function increases in each repetition, the EM algorithm provides remarkable results for computation of the ML estimates. For the type II censoring case, the censored observations can

be thought of missing data and therefore the EM algorithm can be used to obtain the ML estimators. For the type II censored data with given prefixed constant r, let us denote the observed and censored data by $\boldsymbol{X} = (X_{1:n}, \; X_{2:n}, \; ..., \; X_{r:n})$ and $\boldsymbol{Z} = (Z_1, \; Z_2, \; ..., \; Z_{n-r})$, respectively. Then the complete data is $\boldsymbol{W} = (\boldsymbol{X}, \; \boldsymbol{Z})$. The log-likelihood function for the complete sample $\boldsymbol{W}$ is expressed as

$$
\begin{aligned}
\log L_c \left(c, k; \boldsymbol{W} \right) &= n \left(\log k + \log c \right) - (c+1) \sum_{i=1}^{n} \log w_i - (k+1) \sum_{i=1}^{n} \log(1 + w_i^{-c}) \\
&= n \left(\log k + \log c \right) - (c+1) \left(\sum_{i=1}^{r} \log x_{i:n} + \sum_{i=1}^{n-r} \log Z_i \right) \\
&\quad - (k+1) \left(\sum_{i=1}^{r} \log \left(1 + x_{i:n}^{-c} \right) + \sum_{i=1}^{n-r} \log \left(1 + Z_i^{-c} \right) \right) \quad (2.5)
\end{aligned}
$$

In the E step of the algorithm, the conditional expectation of the $\log L_c$ given $X_{1:n} = x_{1:n}, \; ..., X_{r:n} = x_{r:n}$, that is $E_{\boldsymbol{Z}|\boldsymbol{X}} \left[\log L_c \left(c, k; \boldsymbol{W} \right) \right]$ has to be calculated. By the theorem of Ng et al. [22], given $\boldsymbol{X} = (x_{1:n}, ..., x_{r:n})$, the unobserved variables $Z_1, Z_2, ..., Z_{n-r}$ are conditionally independent and their conditional distributions are identical truncated BIII distribution with left truncation at $x_{r:n}$. Therefore,

$E_{\boldsymbol{Z}|\boldsymbol{X}} \left[\log L_c \left(c, k; \boldsymbol{W} \right) \right]$

$$
\begin{aligned}
&= n \left(\log k + \log c \right) - (c+1) \left(\sum_{i=1}^{r} \log x_{i:n} + \sum_{i=1}^{n-r} E \left[\log Z_i | Z_i > x_{r:n} \right] \right) \\
&\quad - (k+1) \left(\sum_{i=1}^{r} \log \left(1 + x_{i:n}^{-c} \right) + \sum_{i=1}^{n-r} E \left[\log \left(1 + Z_i^{-c} \right) | Z_i > x_{r:n} \right] \right), \, (2.6)
\end{aligned}
$$

where

$$
\begin{aligned}
E \left[\log \left(1 + Z_i^{-c} \right) | Z_i > x_{r:n} \right] &= \frac{ck}{1 - F \left(x_{r:n} \right)} \int_{x_{r:n}}^{\infty} \log \left(1 + z^{-c} \right) z^{-(c+1)} \left(1 + z^{-c} \right)^{-(k+1)} dz \\
&= \frac{1}{k} - \frac{\left(1 + x_{r:n}^{-c} \right)^{-k}}{1 - \left(1 + x_{r:n}^{-c} \right)^{-k}} \log \left(1 + x_{r:n}^{-c} \right), \quad (2.7)
\end{aligned}
$$

$$
E \left[\log Z_i | Z_i > x_{r:n} \right] = \frac{ck}{1 - F \left(x_{r:n} \right)} \int_{x_{r:n}}^{\infty} \left(\log z \right) z^{-(c+1)} \left(1 + z^{-c} \right)^{-(k+1)} dz. \quad (2.8)
$$

The details for the integration are given in the appendix.

In the M step of the algorithm, the function $E_{\boldsymbol{Z}|\boldsymbol{X}} [\log L_c \left(c, k; \boldsymbol{W} \right)]$ is maximized with respect to the parameters c and k. Let us write $\lambda \left(c, k; x_{r:n} \right) = E \left[\log \left(1 + Z_i^{-c} \right) | Z_i > x_{r:n} \right]$ and $\tau \left(c, k; x_{r:n} \right) = E \left[\log Z_i | Z_i > x_{r:n} \right]$. Hereby, the

ML estimates $\hat{c}$ and $\hat{k}$ of the parameters c and k are obtained by repeating the iteration

$$
\begin{aligned}
\left(c_{(h+1)}, k_{(h+1)}\right) \;=\; & arg \max_{c,k>0} \Big\{ n \left(\log k + \log c\right) \\
& - (c+1) \left[\sum_{i=1}^{r} \log x_{i:n} + (n-r)\,\tau\left(c_{(h)}, k_{(h)}; x_{r:n}\right)\right] \\
& - (k+1) \left[\sum_{i=1}^{r} \log\left(1 + x_{i:n}^{-c}\right) + (n-r)\,\lambda\left(c_{(h)}, k_{(h)}; x_{r:n}\right)\right] \Big\}, \\
h \;=\; & 0, 1, 2, ...,
\end{aligned}
$$

until $\left\| \left(c_{(h+1)} - c_{(h)},\; k_{(h+1)} - k_{(h)}\right) \right\| \; < \; \epsilon$ is satisfied, where $\|.\|$ stands for the Euclidean norm in R^2, $c_{(0)}$ and $k_{(0)}$ are the initial guesses of $\hat{c}$ and $\hat{k}$ and $\epsilon > 0$ denotes pre-specified tolerance level. Thus, $\hat{c} = c_{(h+1)}$ and $\hat{k} = k_{(h+1)}$ for which h is the smallest integer such that the convergence criterion holds.

The maximization procedure can be done as given by Kundu and Pradhan [23]. First, $c_{(h+1)}$ is obtained by solving the fixed point type equation

$$
g\left(c\right) = c,
$$

where

$$
g\left(c\right) = n \left[\sum_{i=1}^{r} \log x_{i:n} + (n-r)\,\tau\left(c_{(h)}, k_{(h)}; x_{r:n}\right) - \left(k\left(c\right)+1\right) \sum_{i=1}^{r} \frac{x_{i:n}^{-c} \log x_{i:n}}{1 + x_{i:n}^{-c}}\right]^{-1}
\tag{2.9}
$$

and

$$
k\left(c\right) = n \left[\sum_{i=1}^{r} \log\left(1 + x_{i:n}^{-c}\right) + (n-r)\,\lambda\left(c_{(h)}, k_{(h)}; x_{r:n}\right)\right]^{-1}.
\tag{2.10}
$$

Then $k_{(h+1)}$ is obtained as $k_{(h+1)} = k\left(c_{(h+1)}\right)$.

One of the advantages using the EM algorithm when computing the ML estimates is that it enables us to obtain Fisher information matrix easily via MIP. The MIP is based on the idea

Observed information = Complete information - Missing information.

Let us write $\theta = (c, k)$. Then, the complete information and the missing information matrices are defined as

$$
I_{com}\left(\theta\right) = -nE \left[\begin{array}{cc} \frac{\partial^2 \log f(X;\theta)}{\partial c^2} & \frac{\partial^2 \log f(X;\theta)}{\partial c \partial k} \\ \frac{\partial^2 \log f(X;\theta)}{\partial c \partial k} & \frac{\partial^2 \log f(X;\theta)}{\partial k^2} \end{array} \right],
\tag{2.11}
$$

$$
I_{mis}\left(\theta\right) = -(n-r)E \left[\begin{array}{cc} \frac{\partial^2 \log f(X|X>x_{r:n};\theta)}{\partial c^2} & \frac{\partial^2 \log f(X|X>x_{r:n};\theta)}{\partial c \partial k} \\ \frac{\partial^2 \log f(X|X>x_{r:n};\theta)}{\partial c \partial k} & \frac{\partial^2 \log f(X|X>x_{r:n};\theta)}{\partial k^2} \end{array} \right],
\tag{2.12}
$$

respectively, where X is BIII random variate and f is the pdf of X. Thus, the observed information matrix is given as

$$I_{obs}(\theta) = I_{com}(\theta) - I_{mis}(\theta). \tag{2.13}$$

The elements of these matrices are given in the appendix. By inverting the matrix $I_{obs}(\theta)$, one can obtain the asymptotic variance-covariance matrix of the ML estimates $\hat{c}$ and $\hat{k}$.

In order to evaluate the performance of approximation for the variances and covariance of the ML estimates obtained from the information matrix, we have carried out a Monte Carlo simulation with 1000 replications. For the sake of simplicity, we take $c = 1$ and $k = 1$. The variances and covariance obtained from simulation and information matrix are given in Table 1. We also give coverage probability (CP) of %95 confidence interval of $\hat{c} + \hat{k}$. It is ratio satisfying $c + k \in (\hat{c} + \hat{k} - 1.96\sqrt{\text{Var}(\hat{c} + \hat{k})}, \hat{c} + \hat{k} + 1.96\sqrt{\text{Var}(\hat{c} + \hat{k})})$. Here, $\text{Var}(\hat{c} + \hat{k}) = \text{Var}(\hat{c}) + \text{Var}(\hat{k}) + 2\text{Cov}(\hat{c} + \hat{k})$ and it is computed from information.

TABLE 1. Variances, covariance and coverage probability of ML estimates using the EM algorithm

		Simulated			From Information			CP
n	r	$\text{Var}(\hat{c})$	$\text{Var}(\hat{k})$	$\text{Cov}(\hat{c},\hat{k})$	$\text{Var}(\hat{c})$	$\text{Var}(\hat{k})$	$\text{Cov}(\hat{c},\hat{k})$	$c+k$
30	21	0.0541	0.0452	-0.0251	0.0430	0.0555	-0.0221	0.956
	24	0.0412	0.0407	-0.0180	0.0392	0.0478	-0.0188	0.953
	27	0.0357	0.0445	-0.0170	0.0364	0.0450	-0.0162	0.954
50	35	0.0224	0.0270	-0.0117	0.0209	0.0329	-0.0123	0.951
	40	0.0222	0.0243	-0.0087	0.0211	0.0285	-0.0108	0.952
	45	0.0188	0.0274	-0.0089	0.0199	0.0262	-0.0095	0.954
100	70	0.0096	0.0126	-0.0051	0.0094	0.0163	-0.0059	0.946
	80	0.0101	0.0123	-0.0049	0.0097	0.0141	-0.0052	0.950
	90	0.0097	0.0120	-0.0046	0.0095	0.0127	-0.0047	0.955

It is seen from the Table 1 that the approximations for the variances and covariance determined from the information matrix are satisfactory in the case of moderate n and low censoring. The approximations become more accurate as n increases even if the censoring level is high. Also, the values of CP are near 0.95 in almost all cases.

2.2. **MPS method.** The MPS method was introduced as an alternative to ML method by Cheng and Amin [24] and independently by Ranneby [25]. Their aim is to overcome an unboundedness problem of the likelihood with a pdf having shifted origin for which the ML estimators may become inconsistent. For comprehensive content, one can refer to Ekström [26]. MPS and ML estimators are asymptotically equal and have same asymptotic sufficiency, consistency and efficiency properties as pointed out in Cheng and Amin [24]. The MPS method is based on maximizing the product of consecutive differences between values of the cdf at ordered observations.

Considering the probability integral transform, it is clear that MPS method chooses the values as estimates which make the spacings as uniform as possible. Since the cdf is being used as base function, there does not occur any unboundedness problem.

The MPS estimators of c and k are given for complete and type II censored data as follows. Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a complete sample from BIII distribution. Therefore, the MPS estimates are the arguments maximizing the function $G(c, k; \boldsymbol{x})$

$$G\left(c, k; \boldsymbol{x}\right) = \prod_{i=1}^{n+1} \left[F\left(x_{i:n}; c, k\right) - F\left(x_{i-1:n}; c, k\right)\right], \tag{2.14}$$

with respect to c and k where $(x_{1:n}, x_{2:n}, ..., x_{n:n})$ is the ordered sample and $F\left(x_{0:n}; c, k\right) := 0$, $F\left(x_{n+1:n}; c, k\right) := 1$. For the type II censored sample, the function $G\left(c, k; \boldsymbol{x}\right)$ can be obtained similarly as given in Ng et al. [27] by

$$G\left(c, k; \boldsymbol{x}\right) = \prod_{i=1}^{r+1} \left[F\left(x_{i:n}; c, k\right) - F\left(x_{i-1:n}; c, k\right)\right] \times \left[1 - F\left(x_{r:n}; c, k\right)\right]^{n-r}, \tag{2.15}$$

where $\boldsymbol{x} = (x_{1:n}, x_{2:n}, ..., x_{r:n})$ and $F\left(x_{r+1:n}; c, k\right)$ is taken to be 1 for convenience. Hence, the MPS estimators are

$$\left(\tilde{c}, \tilde{k}\right) = arg \max_{c, k > 0} G\left(c, k; \boldsymbol{X}\right). \tag{2.16}$$

Since there are no analytic expressions for $\tilde{c}$ and $\tilde{k}$, they can be obtained by numerical non-linear optimization techniques.

3. Prediction of future order statistics

In this section, we will consider predicting the unobserved future order statistics for type II censored data. As well as, there are some methods, we prefer to use the best unbiased prediction method. Let $\boldsymbol{X} = (X_{1:n}, X_{2:n}, ..., X_{r:n})$ be the sample will being observed and $Y_k = X_{r+k:n}$, $k = 1, ..., n-r$ be the unobservable variables for the type II censored data. Therefore, the best unbiased predictor (BUP) for Y_k, $k = 1, ..., n-r$ is defined as

$$\hat{Y}_k = E\left[Y_k | \boldsymbol{X}\right]. \tag{3.1}$$

Considering the Markovian property of order statistics, $E\left[Y_k | \boldsymbol{X}\right] = E\left[Y_k | X_{r:n}\right]$. It should be noted that, Y_k, $k = 1, ..., n-r$, corresponds to the kth order statistic of $n - r$ units from the truncated BIII distribution with left truncation at $x_{r:n}$, given $X_{r:n} = x_{r:n}$. Therefore, given $X_{r:n} = x_{r:n}$, the pdf of Y_k is

$$f_k\left(y | x_{r:n}; c, k\right) = k \binom{n-r}{k} \left[F\left(y; c, k\right) - F\left(x_{r:n}; c, k\right)\right]^{k-1} \left[1 - F\left(y; c, k\right)\right]^{n-r-k}$$

$$\times \left[1 - F\left(x_{r:n}; c, k\right)\right]^{r-n}, \quad y > x_{r:n}, \tag{3.2}$$

where f and F are the pdf and cdf of BIII distribution with parameters c and k, respectively. Thus, the BUP for Y_k, $k = 1, ..., n - r$ is

$$\hat{Y}_k = \int_{x_{r:n}}^{\infty} y f_k(y|x_{r:n}; \hat{c}, \hat{k}) dy, \qquad (3.3)$$

where $\hat{c}$ and $\hat{k}$ are the ML estimates based on $(x_{1:n}, \; x_{2:n}, \; ..., \; x_{r:n})$.

4. Simulation study

In this section, we will carry out a Monte Carlo simulation in order to investigate how the ML and MPS estimation methods perform for BIII distribution under different settings for sample sizes, censoring and parameter values. For this purpose, the number of replications is chosen as $N = 10^4$. Sample sizes are taken as 20, 40, 60 and 100. Two censoring levels, %10 and %30, are considered for each sample sizes. It should be noted that the pdf of the BIII distribution is L-shaped when $ck \leq 1$ and unimodal when $ck > 1$. So, we both consider L-shaped and unimodal cases by choosing the parameters as $(c = 1, \; k = 1)$, $(c = 0.6, \; k = 0.8)$, $(c = 2, \; k = 0.4)$ and $(c = 1, \; k = 2)$, $(c = 0.8, \; k = 1.5)$, respectively. For convergence criteria of the EM algorithm, we take the tolerance level as $\epsilon = 10^{-5}$. All the simulations are conducted in MATLAB^{TM} program. The nonlinear optimization tools are used for computation of the MPS estimators. We present the simulation mean, bias and mean square error (MSE) values for both ML and MPS estimators of c and k in Tables 2-6.

TABLE 2. ML and MPS estimations when $c = 1, k = 1$

		$\hat{c}_{ML}$			$\hat{k}_{ML}$			$\hat{c}_{MPS}$			$\hat{k}_{MPS}$		
n	r	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
20	14	0.9881	-0.0119	0.5874	1.0993	0.0993	0.0925	1.1196	0.1196	2.7714	0.9912	-0.0088	0.0744
	18	1.0682	0.0682	0.0799	1.0371	0.0371	0.0733	0.9690	-0.0310	0.0614	1.0038	0.0038	0.0594
40	28	0.8668	-0.1332	0.0554	1.1088	0.1088	0.0461	0.9993	-0.0007	0.1258	0.9928	-0.0072	0.0352
	36	1.0161	0.0161	0.0278	1.0255	0.0255	0.0336	0.9760	-0.0240	0.0263	0.9980	-0.0020	0.0301
60	42	0.8430	-0.1570	0.0423	1.1119	0.1119	0.0342	0.9935	-0.0065	0.0275	0.9946	-0.0054	0.0232
	54	1.0011	0.0011	0.0169	1.0227	0.0227	0.0218	0.9814	-0.0186	0.0168	0.9978	-0.0022	0.0201
100	70	0.8241	-0.1759	0.0404	1.1137	0.1137	0.0253	0.9907	-0.0093	0.0154	0.9963	-0.0037	0.0137
	90	0.9879	-0.0121	0.0096	1.0192	0.0192	0.0127	0.9856	-0.0144	0.0098	0.9971	-0.0029	0.0119

TABLE 3. ML and MPS estimations when $c = 0.6, k = 0.8$

		$\hat{c}_{ML}$			$\hat{k}_{ML}$			$\hat{c}_{MPS}$			$\hat{k}_{MPS}$		
n	r	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
20	14	0.6244	0.0244	0.3107	0.8938	0.0938	0.0834	0.7981	0.1981	2.3306	0.7962	-0.0038	0.0641
	18	0.6547	0.0547	0.0732	0.8263	0.0263	0.0540	0.5936	-0.0064	0.1635	0.8102	0.0102	0.0459
40	28	0.5239	-0.0761	0.0460	0.9148	0.1148	0.0440	0.6240	0.0240	0.2603	0.7983	-0.0017	0.0295
	36	0.6101	0.0101	0.0125	0.8243	0.0243	0.0244	0.5874	-0.0126	0.0119	0.8049	0.0049	0.0222
60	42	0.4989	-0.1011	0.0205	0.9210	0.1210	0.0336	0.6003	0.0003	0.0146	0.7988	-0.0012	0.0192
	54	0.5985	-0.0015	0.0071	0.8215	0.0215	0.0157	0.5890	-0.0110	0.0072	0.8017	0.0017	0.0147
100	70	0.4853	-0.1147	0.0174	0.9238	0.1238	0.0257	0.5976	-0.0024	0.0075	0.7978	-0.0022	0.0111
	90	0.5916	-0.0084	0.0041	0.8199	0.0199	0.0096	0.5930	-0.0070	0.0042	0.8002	0.0002	0.0090

TABLE 4. ML and MPS estimations when $c = 0.4, k = 2$

		$\hat{c}_{ML}$			$\hat{k}_{ML}$			$\hat{c}_{MPS}$			$\hat{k}_{MPS}$		
n	r	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
20	14	0.3720	-0.0280	0.0083	2.1095	0.1095	0.2441	0.3885	-0.0115	0.0084	1.9663	-0.0337	0.2231
	18	0.4202	0.0202	0.0071	2.1326	0.1326	0.3198	0.3824	-0.0176	0.0058	1.9618	-0.0382	0.2129
40	28	0.3543	-0.0457	0.0051	2.0514	0.0514	0.0923	0.3909	-0.0091	0.0037	1.9657	-0.0343	0.1002
	36	0.4061	0.0061	0.0030	2.0602	0.0602	0.1192	0.3880	-0.0120	0.0028	1.9639	-0.0361	0.0981
60	42	0.3477	-0.0523	0.0046	2.0360	0.0360	0.0560	0.3915	-0.0085	0.0024	1.9721	-0.0279	0.0647
	54	0.4016	0.0016	0.0018	2.0417	0.0417	0.0750	0.3908	-0.0092	0.0018	1.9718	-0.0282	0.0658
100	70	0.3440	-0.0560	0.0042	2.0244	0.0244	0.0328	0.3944	-0.0056	0.0014	1.9802	-0.0198	0.0399
	90	0.3979	-0.0021	0.0011	2.0260	0.0260	0.0425	0.3935	-0.0065	0.0011	1.9795	-0.0205	0.0395

TABLE 5. ML and MP Sestimations when $c = 0.8, k = 1.5$

		$\hat{c}_{ML}$			$\hat{k}_{ML}$			$\hat{c}_{MPS}$			$\hat{k}_{MPS}$		
n	r	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
20	14	0.7425	-0.0575	0.0493	1.6165	0.1165	0.1497	0.7830	-0.0170	0.0902	1.4883	-0.0117	0.1274
	18	0.8431	0.0431	0.0341	1.5833	0.0833	0.1629	0.7668	-0.0332	0.0277	1.4902	-0.0098	0.1180
40	28	0.7012	-0.0988	0.0236	1.5907	0.0907	0.0631	0.7834	-0.0166	0.0182	1.4865	-0.0135	0.0588
	36	0.8121	0.0121	0.0135	1.5399	0.0399	0.0663	0.7774	-0.0226	0.0128	1.4832	-0.0168	0.0565
60	42	0.6879	-0.1121	0.0211	1.5792	0.0792	0.0423	0.7854	-0.0146	0.0118	1.4844	-0.0156	0.0407
	54	0.8013	0.0013	0.0083	1.5290	0.0290	0.0420	0.7818	-0.0182	0.0083	1.4854	-0.0146	0.0378
100	70	0.6783	-0.1217	0.0196	1.5743	0.0743	0.0255	0.7888	-0.0112	0.0068	1.4884	-0.0116	0.0235
	90	0.7937	-0.0063	0.0049	1.5213	0.0213	0.0243	0.7872	-0.0128	0.0049	1.4893	-0.0107	0.0228

TABLE 6. ML and MPS estimations when $c = 1, k = 2$

		$\hat{c}_{ML}$			$\hat{k}_{ML}$			$\hat{c}_{MPS}$			$\hat{k}_{MPS}$		
n	r	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
20	14	0.9299	-0.0701	0.0519	2.1096	0.1096	0.2441	0.9711	-0.0289	0.0526	1.9663	-0.0337	0.2231
	18	1.0532	0.0532	0.0462	2.1314	0.1314	0.3115	0.9580	-0.0420	0.0377	1.9607	-0.0393	0.2079
40	28	0.8834	-0.1166	0.0321	2.0539	0.0539	0.0927	0.9746	-0.0254	0.0233	1.9684	-0.0316	0.1001
	36	1.0142	0.0142	0.0184	2.0611	0.0611	0.1235	0.9691	-0.0309	0.0176	1.9647	-0.0353	0.1014
60	42	0.8711	-0.1289	0.0284	2.0347	0.0347	0.0566	0.9810	-0.0190	0.0153	1.9708	-0.0292	0.0654
	54	1.0028	0.0028	0.0117	2.0398	0.0398	0.0730	0.9760	-0.0240	0.0116	1.9702	-0.0298	0.0644
100	70	0.8600	-0.1400	0.0261	2.0244	0.0244	0.0328	0.9859	-0.0141	0.0088	1.9802	-0.0198	0.0399
	90	0.9947	-0.0053	0.0067	2.0260	0.0260	0.0425	0.9837	-0.0163	0.0068	1.9795	-0.0205	0.0395

The simulation results have been summarized as follows. First of all, the behaviors of the ML and MPS estimators do not show a discrepancy depending on whether the pdf is L-shaped or unimodal. The biases of $\hat{c}$ and $\hat{k}$ are a bit greater, compared to the biases of $\tilde{c}$ and $\tilde{k}$. Although the biases of $\tilde{c}$ and $\tilde{k}$ do not change depending on the censoring proportion, the biases of $\hat{c}$ and $\hat{k}$ increase as the censoring proportion increases. All the biases get smaller as the sample size gets larger. $\tilde{k}$ mostly underestimate the parameter k for all cases whereas $\hat{k}$ overestimates absolutely. Similarly, $\tilde{c}$ usually underestimates the parameter c. However, there is no such a case for $\hat{c}$. For each sample size, the MSE values of all the estimators increase when the censoring proportion gets larger. Also, all the MSE values get smaller as the sample size increases. Since the ML and MPS estimators are consistent, this

is an expected situation. Further, it is observed that there is not any remarkable superiority between the ML and MPS estimators with respect to the MSE criterion, that is, all the MSE values are close to each other for both methods in much cases.

5. Illustrative example

In this section, we will consider a real data set given by Mann and Fertig [28]. This data set is also studied by Prakash and Singh [29] from a Bayesian perspective. The data set contains 10 failure times of airplane components of total 13 items. The censoring scheme is type II censoring with $r = 10$ and $n = 13$. The data set is

$$\{0.22,\ 0.50,\ 0.88,\ 1.00,\ 1.32,\ 1.33,\ 1.54,\ 1.76,\ 2.50,\ 3.00\}\,.$$

If we model this data set with BIII distribution, the ML estimates of parameters and the related standard errors (in parenthesis) are obtained as $\hat{c} = 1.4990(0.1458)$ and $\hat{k} = 1.7025(0.2279)$. This data set was also studied in Crowder et al. [30]. They considered Weibull and lognormal models. For these models, the rough estimates were found by graphical technique as $\hat{\alpha} = 1.4$, $\hat{\beta} = 2.2$, $\hat{\mu} = 0.5$, $\hat{\sigma} = 1$.

Here, α and β are shape and scale parameters of the Weibull distribution, respectively, μ and σ are the parameters of lognormal distribution which are mean and standard deviation of the corresponding normal distribution, respectively. We have computed the ML estimations of these parameters as given in Table 7. In order to compare BIII, Weibull and lognormal models for this data set, we consider loglikelihood, Akaike information criteria (AIC) values and Cramér-von Mises and Anderson-Darling test statistics. The results are presented in Table 7. It should be noted that the Cramér-von Mises and Anderson-Darling test statistics for the type II censored data are computed by using the $W_{r,n}^2$ and $A_{r,n}^2$ statistics given in D'Agostino and Stephens [31] at p. 114. Critical points of the $W_{r,n}^2$ and $A_{r,n}^2$ statistics at significance level 0.5 are nearly 0.085 and 0.503, respectively as given in Table 4.5 of [31]. So, the associated p-values of both statistics for BIII, Weibull and lognormal models are much bigger than 0.5.

Table 7. Comparison of BIII, Weibulland Lognormal models

	ML Estimation		Loglikelihood	AIC	Cramer-von Mises $(W_{r,n}^2)$	Anderson-Darling $(A_{r,n}^2)$
BIII	$\hat{c} = 1.4990$	$\hat{k} = 1.7025$	3.0228	-2.0456	0.0191 $(p > 0.5)$	0.1232 $(p > 0.5)$
Weibull	$\hat{\alpha} = 1.4174$	$\hat{\beta} = 2.2728$	3.1269	-2.2538	0.0215 $(p > 0.5)$	0.1119 $(p > 0.5)$
Lognormal	$\hat{\mu} = 0.4788$	$\hat{\sigma} = 0.9384$	3.1107	-2.2213	0.0188 $(p > 0.5)$	0.1234 $(p > 0.5)$

From Table 7, it can be considered that the BIII is successful to model this data set as much as Weibull and lognormal models. This indicates that the BIII distribution may be a good alternative for fitting lifetime data sets. Furthermore, the predictions for the unobserved variables are computed as $\hat{Y}_1 = 4.0601$, $\hat{Y}_2 = 6.3858$ and $\hat{Y}_3 = 20.0207$ via BUP approach. When a sample of size 13 from BIII distribution with $c = 1.4990$ and $k = 1.7025$ is generated 10000 times, the last

three observations of ordered sample are computed averagely as 4.2975, 6.8485 and 19.9901. These values agree with the predictions, as expected.

6. CONCLUSION

In this study, we have concerned with the statistical inference problem for the BIII distribution under type II censored data. For this purpose, we have taken into account the ML and MPS methods. We have implemented the EM algorithm for the computation of the ML estimators. Also, the information matrix is derived by means of MIP. The unobserved future order statistics have been predicted via BUP approach. Monte Carlo simulations indicate that in order to estimate the model parameters, both of the ML and MPS methods can be used. Especially, if the censoring proportion is large we suggest to use the MPS method since absolute biases of the MPS estimators are lower than that of ML estimators. Finally, the real life examination has illustrated modeling capacity of the BIII distribution.

Acknowledgement

The authors would like to express their gratitude to an anonymous referee for the valuable comments and suggestions.

7. APPENDIX

The integral given in Equation (2.7) is analytically obtained by first applying $1 + z^{-c} = y$ transformation and then applying partial integration. These steps are as follows:

$$E\left[\log\left(1 + Z_i^{-c}\right) | Z_i > x_{r:n}\right]$$

$$= \frac{ck}{1 - F(x_{r:n})} \int_{x_{r:n}}^{\infty} \log\left(1 + z^{-c}\right) z^{-(c+1)} \left(1 + z^{-c}\right)^{-(k+1)} dz$$

$$= \frac{k}{1 - F(x_{r:n})} \int_1^{1+x_{r:n}^{-c}} \log(y) y^{-(k+1)} dy$$

$$= \frac{k}{1 - F(x_{r:n})} \left[\log(y)\left(\frac{y^{-k}}{-k}\right) \Big|_1^{1+x_{r:n}^{-c}} + \int_1^{1+x_{r:n}^{-c}} y^{-(k+1)} dy\right]$$

$$= \frac{k}{1 - F(x_{r:n})} \left[\log\left(1 + x_{r:n}^{-c}\right)\left(\frac{(1 + x_{r:n}^{-c})^{-k}}{-k}\right) + \frac{1}{k^2}\left(1 - F(x_{r:n})\right)\right]$$

$$= \frac{1}{k} - \frac{(1 + x_{r:n}^{-c})^{-k}}{1 - (1 + x_{r:n}^{-c})^{-k}} \log\left(1 + x_{r:n}^{-c}\right).$$

Existence of the expectation given in Equation (2.8) can be easily shown by simple calculus.

To get the complete information $I_{com}(\theta)$ the second partial derivatives of $\log f$ with respect to c and k are

$$\frac{\partial^2 \log f}{\partial c^2} = \frac{-1}{c^2} - (k+1)\frac{\log^2(x)x^{-c}}{(1+x^{-c})^2},$$

$$\frac{\partial^2 \log f}{\partial k^2} = \frac{-1}{k^2},$$

$$\frac{\partial^2 \log f}{\partial c \partial k} = \frac{\log(x)x^{-c}}{1+x^{-c}}.$$

Therefore the elements of the complete information based on n units are

$$I_{com}(\theta)_{11} = n\left(\frac{1}{c^2} + (k+1)E\left[\frac{\log^2(X)X^{-c}}{(1+X^{-c})^2}\right]\right),$$

$$I_{com}(\theta)_{12} = -nE\left[\frac{\log(X)X^{-c}}{1+X^{-c}}\right],$$

$$I_{com}(\theta)_{21} = I_{com}(\theta)_{12},$$

$$I_{com}(\theta)_{22} = \frac{n}{k^2},$$

where X is BIII random variable. Since the expectations $E\left[\frac{\log^2(X)X^{-c}}{(1+X^{-c})^2}\right]$ and $E\left[\frac{\log(X)X^{-c}}{1+X^{-c}}\right]$ cannot be obtained analytically, they have to be computed numerically. The missing information is obtained in a similar manner. Skipping the derivation steps, we obtain the elements of the missing information as

$$I_{mis}(\theta)_{11} = (n-r)\left(\frac{1}{c^2} + (k+1)E\left[\frac{\log^2(X)X^{-c}}{(1+X^{-c})^2}|X > x_{r:n}\right] + A\right),$$

$$I_{mis}(\theta)_{12} = -(n-r)\left(E\left[\frac{\log(X)X^{-c}}{1+X^{-c}}|X > x_{r:n}\right] + B\right),$$

$$I_{mis}(\theta)_{21} = I_{mis}(\theta)_{12},$$

$$I_{mis}(\theta)_{22} = (n-r)\left(\frac{1}{k^2} - C\right),$$

where

$$A = \frac{\log^2(x_{r:n})k\left[(1+x_{r:n}^{-c})^k(x_{r:n}^c - k) - x_{r:n}^c\right]}{(1+x_{r:n}^c)^2\left(1 - (1+x_{r:n}^{-c})^{-k}\right)^2},$$

$$B = \frac{(1+x_{r:n}^{-c})^{-k-1}\log(x_{r:n})x_{r:n}^{-c}}{\left(1 - (1+x_{r:n}^{-c})^{-k}\right)}\left[1 - k\log(1+x_{r:n}^{-c})\left(1 - (1+x_{r:n}^{-c})^{-k}\right)^{-1}\right],$$

$$C = \frac{\log^2(1 + x_{r:n}^{-c})(1 + x_{r:n}^{-c})^{-k}}{\left(1 - \left(1 + x_{r:n}^{-c}\right)^{-k}\right)^2}.$$

Here, the expectations have to be computed numerically since they have no analytical solution. Also, it should be pointed out that all the expectations related with the information matrices exist. It is straightforward from the existence of $E[\log(X)]$.

References

[1] Burr, I. W. (1942). Cumulative frequency functions. *The Annals of mathematical statistics* 13(2): 215-232.

[2] Lindsay, S. R., et al. (1996). Modelling the diameter distribution of forest stands using the Burr distribution. *Journal of Applied Statistics* 23(6): 609-619.

[3] Shao, Q. (2000). Estimation for hazardous concentrations based on NOEC toxicity data: an alternative approach. *Environmetrics* 11(5): 583-595.

[4] Zimmer, W. J., et al. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology* 30(4): 386.

[5] AbdelGhaly, A. A., et al. (1997). The use of Burr type XII distribution on software reliability growth modelling. *Microelectronics and Reliability* 37(2): 305-313.

[6] Shankar, G. and Sahani, V. (1994). The Study of a Maintenance Float Model with Burr Failure Distribution. *Microelectronics and Reliability* 34(9): 1513-1517.

[7] Liu, P. H. and Chen, F. L. (2006). Process capability analysis of non-normal process data using the Burr XII distribution. International Journal of Advanced Manufacturing Technology 27(9-10): 975-984.

[8] Gove, J. H., et al. (2008). Rotated sigmoid structures in managed uneven-aged northern hardwood stands: a look at the Burr Type III distribution. Forestry 81(2): 161-176.

[9] Kar, R., et al. (2010). A closed form Delay Evaluation Approach using Burr's Distribution Function for High Speed On-Chip RC Interconnects. 2010 *Ieee 2nd International Advance Computing Conference*: 129-133.

[10] Ganora, D. and Laio, F. (2015). Hydrological Applications of the Burr Distribution: Practical Method for Parameter Estimation. *Journal of Hydrologic Engineering* 20(11).

[11] Rodriguez, R. N. (1977). A guide to the Burr type XII distributions. *Biometrika* 64(1): 129-134.

[12] Tadikamalla, P. R. (1980). A look at the Burr and related distributions. I*nternational Statistical Review/Revue Internationale de Statistique*: 337-344.

[13] Zoraghi, N., et al. (2012). Estimating the four parameters of the Burr III distribution using a hybrid method of variable neighborhood search and iterated local search algorithms. *Applied Mathematics and Computation* 218(19): 9664-9675.

[14] Wingo, D. R. (1993). Maximum likelihood estimation of Burr XII distribution parameters under type II censoring. *Microelectronics Reliability* 33(9): 1251-1257.

[15] Wang, F.K. and Cheng, Y. (2010). EM algorithm for estimating the Burr XII parameters with multiple censored data. *Quality and Reliability Engineering International* 26(6): 615-630.

[16] Panahi, H. and Sayyareh, A. (2014). Parameter estimation and prediction of order statistics for the Burr Type XII distribution with Type II censoring. *Journal of Applied Statistics* 41(1): 215-232.

[17] Moradi, N., et al. (2014). Estimation of the parameters of a exponentiated Burr type III distribution under type II censoring. *Journal of Statistical Sciences* 8(1): 93-109.

[18] Azizi, A. Z., et al. (2013). Inference about the Burr type III distribution under type-II hybrid censored data. *Journal of Statistical Research of Iran* 10(2): 209-233.

[19] Singh, D. P., et al. (2016). Estimation and prediction for a Burr III distribution with progressive censoring. *Communication in Statistics - Theory and Methods*. Accepted manuscript. DOI:10.1080/03610926.2016.1213290.

[20] Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. J*ournal of the Royal Statistical Society. Series B (Methodological)*: 226-233.

[21] Dempster, A. P., et al. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*: 1-38.

[22] Ng, H., et al. (2002). Estimation of parameters from progressively censored data using EM algorithm. *Computational Statistics & Data Analysis* 39(4): 371-386.

[23] Kundu, D. and Pradhan, B. (2009). Estimating the parameters of the generalized exponential distribution in presence of hybrid censoring. *Communications in Statistics - Theory and Methods* 38(12): 2030-2041.

[24] Cheng, R. and Amin, N. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)*: 394-403.

[25] Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*: 93-112.

[26] Ekström, M. (2008). Alternatives to maximum likelihood estimation based on spacings and the Kullback–Leibler divergence. *Journal of Statistical Planning and Inference* 138(6): 1778-1791.

[27] Ng, H., et al. (2012). Parameter estimation of three-parameter Weibull distribution based on progressively Type-II censored samples. *Journal of Statistical Computation and Simulation* 82(11): 1661-1678.

[28] Mann, N. R. and Fertig, K. W. (1973). Tables for obtaining Weibull confidence bounds and tolerance bounds based on best linear invariant estimates of parameters of the extreme-value distribution. *Technometrics* 15(1): 87-101.

[29] Prakash, G. and Singh, D. (2009). A Bayesian shrinkage approach in Weibull type-II censored data using prior point information. *REVSTAT–Statistical Journal* 7(2): 171-187.

[30] Crowder, M.J., et al. (1991). Statistical Analysis of Reliability Data. *Great Britain: Chapmann&Hall.*

[31] D'agostino, R.B. and Stephens, M.A. (1986). Goodness of Fit Techniques. *New York: Marcel Dekker.*

Current address: Ömer Altındağ: Bilecik Şeyh Edebali University, Faculty of Arts and Science, Department of Statistics, Bilecik, Turkey.

E-mail address: omer.altindag@bilecik.edu.tr, omeraltindag87@gmail.com

Current address: Mehmet Niyazi Çankaya: Uşak University, Faculty of Arts and Science, Department of Statistics, Uşak, Turkey.

E-mail address: mehmet.cankaya@usak.edu.tr, mehmetncankaya@gmail.com

Current address: Abdullah Yalçınkaya: Ankara University, Faculty of Science, Department of Statistics, Ankara, Turkey.

E-mail address: ayalcinkaya@ankara.edu.tr

Current address: Halil Aydoğdu: Ankara University, Faculty of Science, Department of Statistics, Ankara, Turkey.

E-mail address: aydogdu@ankara.edu.tr