REVSTAT – Statistical Journal
Volume 21, Number 4, October 2023, 447–468
https://doi.org/10.57805/revstat.v21i4.425

A Note on the Stochastic EM Algorithm Based on Left Truncated Right Censored Data from Burr XII Distribution

Author: Debanjan Mitra
– Operations Management, Quantitative Methods, and Information Systems,
Indian Institute of Management Udaipur, India
debanjan.mitra@iimu.ac.in

Received: June 2021 Revised: January 2022 Accepted: January 2022

Abstract:

- The Burr XII distribution is a flexible model for failure-time data. A very general and commonly observed structure for failure-time data involves left truncation and right censoring. In this article, modelling of left truncated right censored failure-time data by the Burr XII distribution is discussed. The steps of the stochastic expectation maximization algorithm, which is a useful technique of estimation for incomplete data structures, are developed to estimate the model parameters of the Burr XII distribution. The Newton-Raphson method, which is a direct method of obtaining maximum likelihood estimates by optimizing the observed likelihood is also used. The two methods of inference are assessed and compared through a Monte Carlo simulation study. Discussions of the inferential methods are extended to the cases of a three-parameter Burr XII model, and a covariate-included model. An illustrative example based on real data is provided. Finally, an application of the inferential results to a prediction issue is discussed with an illustration.

Keywords:

- *lifetime distribution; censoring; truncation; stochastic expectation maximization algorithm; prediction.*

AMS Subject Classification:

- 62F10, 62H10.

1. INTRODUCTION

The Burr system of distributions was proposed by Burr, for modelling a wide variety of data observed in real life [8]. Subsequently, among this system of distributions, the Burr XII distribution received special attention by researchers, see Tadikamalla [22] and the references therein. The Burr XII distribution is given by the cumulative distribution function (CDF)

$$F_Y(y;\alpha,\beta) = 1 - (1 + y^\beta)^{-\alpha}, \quad y > 0, \tag{1.1}$$

with corresponding probability density function (PDF)

$$f_Y(y;\alpha,\beta) = \alpha\beta y^{\beta-1}(1 + y^\beta)^{-(\alpha+1)}, \quad y > 0. \tag{1.2}$$

with $\alpha > 0$ and $\beta > 0$, both of which are shape parameters. The Burr XII distribution gets its flexibility through the shape parameters β and α. Figures 1 and 2 display the PDF of the Burr XII distribution for different values of β (keeping α fixed), and α (keeping β fixed), respectively. Note that for $\beta \le 1$, the distribution is L-shaped, while it is unimodal for $\beta > 1$, as observed by Beirlant *et al.* [6].

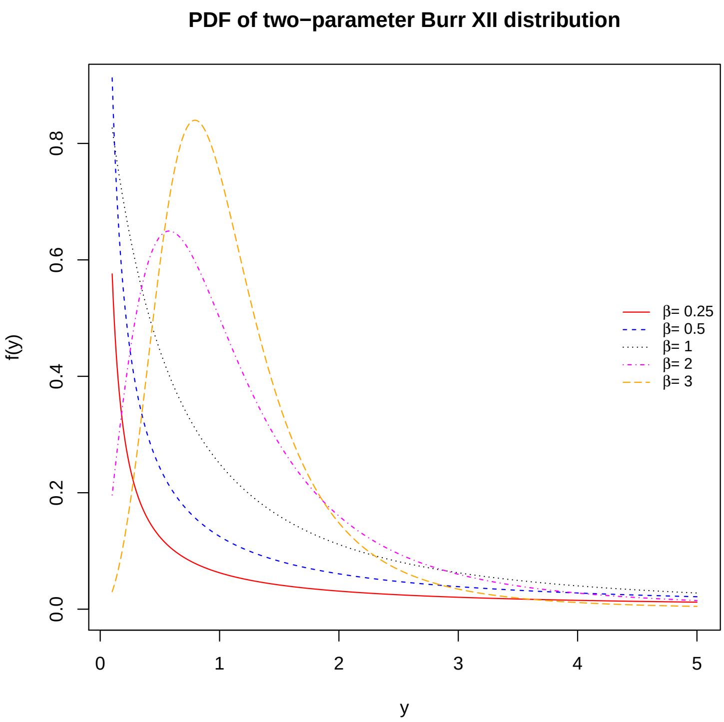

Figure 1: Density function of the Burr XII model for different values of β when $\alpha = 1$.

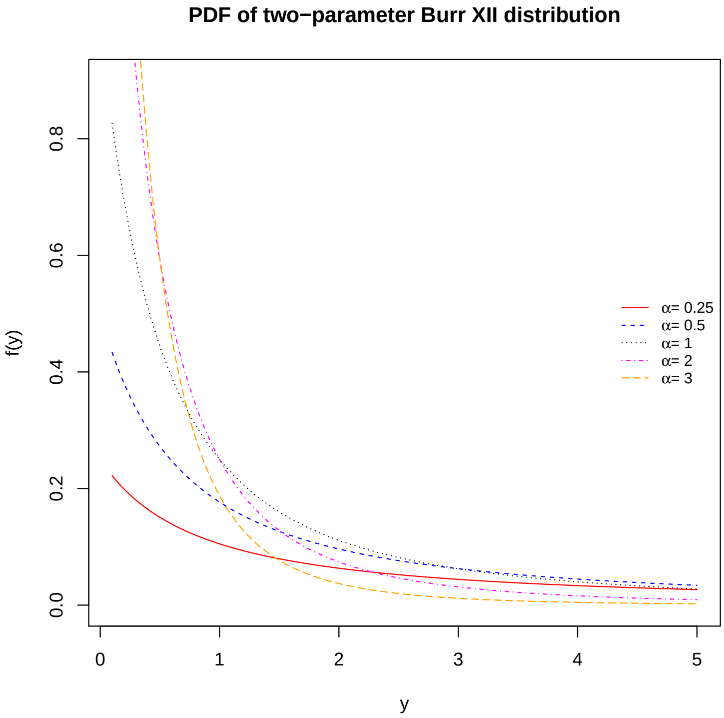

Figure 2: Density function of the Burr XII model for different values of α when $\beta = 1$.

Zimmer *et al.* [24] advocated for the use of the Burr XII distribution as an alternative to lognormal and Weibull distributions, and mentioned the many advantages this model has. The log-logistic distribution, which is another important lifetime model, is a special case of the Burr XII distribution; this also is a motivation to use the Burr XII distribution to model failure-time data [24].

Despite its great flexibility, however, the Burr XII distribution was relatively less used in survival and reliability studies, especially compared to the well-known models like Weibull, gamma etc. Recently, some researchers have used the Burr XII model in the context of failure-time data. For example, Soliman [21] modelled progressively type-II censored data by the Burr XII distribution. Silva *et al.* [19] and Silva *et al.* [20] discussed regression models for the Burr XII distribution based on censored data.

Left truncated right censored (LTRC) data are commonly observed in studies involving lifetimes of experimental units [14]. For example, in many reliability and survival experiments, the main event of interest is the failure of experimental units. Due to practical time constraints on sample collection in such experiments, the observed samples are often either left truncated, or right censored, or both. In medical studies, for example, groups of subjects are often followed over time for observing the occurrence of certain disease or event such as death. LTRC data arise naturally in situations of this type. Another example of LTRC data may be found in oraganisational or social science studies where start-up businesses are observed over a time-window during which they may fail.

As left truncation and right censoring are quite commonly observed features among data arising out of survival and reliability studies, it is of natural importance to develop inferential methods for the Burr XII distribution based on LTRC data, especially as due to its flexible nature the Burr XII model has been posed as a general purpose model for failure-time data by Zimmer *et al.* [24]. To use the Burr XII distribution as a general purpose failure-time model, it is important to develop inferential methods for the model based on LTRC data which is one of the most common and general structures among incomplete data formats in lifetime studies. However, so far, no researcher has attempted modelling LTRC data by using the Burr XII distribution.

In this article, we discuss modelling LTRC failure-time data by the Burr XII distribution in detail. First, we consider the two-parameter version of the Burr XII distribution, as it is the more frequently used version. The stochastic expectation maximization (St-EM) algorithm has emerged as a stable, efficient, and convenient method for parameter estimation for incomplete data problems. For estimating the parameters of the Burr XII model, we develop the steps of the St-EM algorithm based on LTRC data. We discuss two approaches for constructing confidence intervals, one of them being based on an adaptation of the missing information principle of Louis [15], and the other being based on parametric bootstrap approach. For comparison purposes, we also use the Newton-Raphson (NR) method which is a direct approach to obtain maximum likelihood estimates by optimizing the observed likelihood function. Through detailed Monte Carlo simulations, we study the performance of the proposed methods of inferences. Further, we extend our discussion of inferential methods to the cases for a covariate-included model, and the three-parameter Burr XII distribution. These are the main contributions of these paper.

The article is organized as follows. A brief introduction to LTRC data is provided in Section 2. The St-EM algorithm for the two-parameter Burr XII model based on LTRC data is discussed in detail in Section 3; both point and interval estimation procedures are presented. The direct method of obtaining MLEs is presented in this section too. This section also contains a discussion of inferential methods for a covariate-included model. The detailed results of the numerical experiments are presented in Section 4. Then, in Section 5, discussion of the St-EM algorithm is extended to a three-parameter Burr XII model with a scale parameter in addition to the two shape parameters α and β. This three-parameter model is also used in failure-time data modelling [24]. An application of the inferential methods in predicting the expected number of failures in a future time interval is presented in Section 6. Along with an estimate of the expected number of failures in a future time interval, we provide asymptotic confidence intervals for this expected number of failures. This is of direct practical relevance, as in many situations like maintenance, the researcher may want to have an estimate of the expected number of future failures during a certain time period. In Section 7, a numerical illustration based on a real data is provided. Finally, mentioning some future directions of research in this area, the paper is concluded with some remarks in Section 8.

2. LEFT TRUNCATED RIGHT CENSORED DATA

Hong *et al.* [13] analyzed LTRC data obtained from an electrical industry in the US. Following the setup used by Hong *et al.* [13], Balakrishnan and Mitra [1, 2, 3, 4, 5] discussed the EM algorithm based on LTRC data for some commonly used failure-time models such as lognormal, Weibull, gamma, and generalized gamma; see also Mitra *et al.* [16].

Consider a life-test involving n industrial units. Let T denote the underlying failure-time variable. Let L and R denote points of left truncation and right censoring, respectively; that is, we suppose that the study starts at time point L and continues till time point R. Some units start operating before L, while some start after L. No information are available for units that fail before L, making the data left truncated. Some units may not have failed when the study ends at R, and those units become right censored at R. A unit that starts operating before L, has to live through a threshold time, say κ_L, before its failure become an observable event. We call κ_L the left truncation time. An indicator variable ν indicates whether a unit is left truncated or not; for a left truncated unit ν is 0, otherwise it is 1. Note that failures are observable only in the window from L to R. As a result, for each operating unit, there is a time κ_R depending on the starting point of the unit, such that the unit is right censored if $T > \kappa_R$. As different units may have different starting points, values of κ_L and κ_R may be differ from unit to unit. Thus in effect, for each unit we can define the observed lifetime as $Y = \text{Min}(T, \kappa_R)$, provided $Y > \kappa_L$. Let δ denote an indicator variable for censoring; δ is 0 for a right censored unit, and 1 otherwise.

For subsequent formulation of the problem, let S_1 and S_2 denote index sets for untruncated and truncated units, respectively, that is,

$$S_1 = \{i : \nu_i = 1\}, \quad \text{and} \quad S_2 = \{i : \nu_i = 0\},$$

where ν_i is the truncation indicator for the $i-$th unit, $i = 1, ..., n$. Incorporating the censoring indicator δ, we define the index sets

$$S_{11} = \{i : i \in S_1, \delta_i = 1\}, \quad S_{10} = \{i : i \in S_1, \delta_i = 0\},$$

$$S_{21} = \{i : i \in S_2, \delta_i = 1\}, \quad S_{20} = \{i : i \in S_2, \delta_i = 0\}.$$

We further define S_{cen}:

$$S_{cen} = S_{10} \cup S_{20}.$$

We assume that the underlying lifetime T follows the Burr XII distribution with parameters α and β, i.e., $T \sim Burr(\alpha, \beta)$.

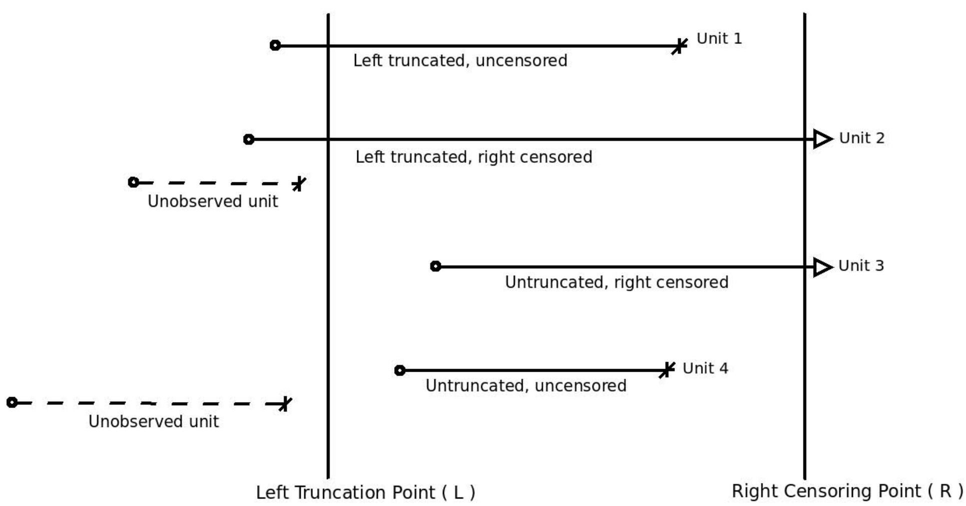

Figure 3: Illustration of LTRC Data.

In Figure 3, we present an illustration of the structure of LTRC data we consider here. We would also like to point out that this is a very general structure that can accommodate units with different combinations of truncation and censoring: left truncated and right censored, left truncated and uncensored, untruncated and right censored, and untruncated and uncensored. This enhances the scope of this model greatly, to be applied to a wide array of observational studies involving failure-times.

3. INFERENCE VIA THE STOCHASTIC EM ALGORITHM

The St-EM algorithm has emerged as a strong tool for analyzing incomplete data. Compared to the traditional EM algorithm, the St-EM algorithm has some distinct advantages. For example, in the EM algorithm, one needs to analytically calculate the conditional expectation of the complete data log-likelihood given the observed data and the current parameter values. Analytical calculation of this conditional expectation may be very difficult, or even intractable, for complex problems. However, in St-EM algorithm, one does not require the analytic calculation of the conditional expectations unlike the EM algorithm. Moreover, in the EM algorithm, the sequence of estimated parameters may get trapped in saddle points depending on the nature of the likelihood surface. But in the St-EM algorithm, due to its stochastic nature, one does not encounter such a problem [23].

The St-EM algorithm has been used for various incomplete data problems in statistical literature; see [9], [23], for example. The asymptotic properties of the St-EM algorithm have been explored by Nielsen [17], among others. Bordes and Chauveau [7] and Ng and Ye [18] recommended the use of the St-EM algorithm for LTRC data.

In St-EM algorithm, for each censored failure-time, a randomly drawn observation from an appropriate conditional distribution is obtained given the observed data and the current value of the parameter. By replacing all censored failure times by such randomly drawn observations, a pseudo-complete dataset is obtained and the pseudo-complete likelihood function is constructed. Then, the pseudo-complete likelihood is optimized to obtain updated parameter estimates. The whole process is then iterated large number of times, to get a sequence of estimates corresponding to each stage of the algorithm. Finally, after discarding some initial values of the estimates for burn-in, the remaining values are averaged to obtain the final estimates.

Note that corresponding to the underlying failure-time variable T, the observed data can be written as

$$\boldsymbol{t} = \boldsymbol{\Upsilon} \cup \boldsymbol{\Gamma},$$

where $\boldsymbol{\Upsilon} = \{t_i : \delta_i = 1\}$ and $\boldsymbol{\Gamma} = \{t_i : \delta_i = 0\}$ contain the observed and right censored failure-times, respectively. For each unit in $\boldsymbol{\Gamma}$, we generate a random observation from the conditional distribution

$$\begin{aligned} f_{T_i|T_i>y_i}(t_i|t_i > y_i; \boldsymbol{\theta}) &= \frac{f_T(t_i; \boldsymbol{\theta})}{1 - F_T(y_i; \boldsymbol{\theta})} \\ &= \alpha\beta t_i^{\beta-1}(1 + t_i^{\beta})^{-(\alpha+1)}(1 + y_i^{\beta})^{\alpha}, \quad t_i > y_i, \end{aligned} \tag{3.1}$$

where $\boldsymbol{\theta} = (\alpha, \beta)$, and y_i is the censored failure-time. By replacing the censored failure-times by these randomly drawn observations, we obtain the pseudo-complete data

$$\boldsymbol{t}_{PC} = \boldsymbol{\Upsilon} \cup \boldsymbol{\Gamma}_{PC}.$$

In the EM algorithm, the complete data likelihood is constructed considering the situation where there would be no incompleteness in the data. In case of the St-EM algorithm, having imputed the censored lifetimes by randomly generated observations from the above conditional distributions, the pseudo-complete data $\boldsymbol{t}_{PC}$ will now be used in a similar fashion, i.e., as if there was no censoring in the data. For a unit that belongs to the untruncated group, the contribution to the likelihood would be $f_T(t_i;\boldsymbol{\theta})$; and for a unit that belongs to the left truncated group, the contribution would be $\frac{f_T(t_i;\boldsymbol{\theta})}{1-F_T(\kappa_{Li};\boldsymbol{\theta})}$. Therefore, by using the pseudo-complete data, the pseudo-complete likelihood is constructed as

$$\begin{aligned} L_{PC}(\boldsymbol{\theta}) &= \prod_{i\in S_1}\{f_T(t_i;\boldsymbol{\theta})\} \times \prod_{i\in S_2}\left\{\frac{f_T(t_i;\boldsymbol{\theta})}{1-F_T(\kappa_{Li};\boldsymbol{\theta})}\right\} \\ &= \prod_{i=1}^{n}\{\alpha\beta t_i^{\beta-1}(1+t_i^{\beta})^{-(\alpha+1)}\} \times \prod_{i\in S_2}\{(1+\kappa_{Li}^{\beta})^{\alpha}\}. \end{aligned} \tag{3.2}$$

The pseudo-complete log-likelihood function, given by

$$\begin{aligned} \log L_{PC}(\boldsymbol{\theta}) &= n(\log\alpha+\log\beta) + \sum_{i=1}^{n}\Big[(\beta-1)\log t_i - (\alpha+1)\log(1+t_i^{\beta})\Big] \\ &\quad + \alpha\sum_{i\in S_2}\log(1+\kappa_{Li}^{\beta}), \end{aligned} \tag{3.3}$$

which we shall denote by $Q_{PC}(\boldsymbol{\theta})$, essentially serves as the pseudo-Q function in this setup, where the Q-function in the traditional EM algorithm is defined as

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)}) = E_{\boldsymbol{\theta}^{(k)}}[\log L_C(\boldsymbol{\theta})|\boldsymbol{\Gamma}],$$

with $\log L_C(\boldsymbol{\theta})$ as the complete data log-likelihood, and $\boldsymbol{\theta}^{(k)}$ as the available value of the parameter vector at the k-th stage of iteration.

To optimize $Q_{PC}(\boldsymbol{\theta})$, for fixed β, equating the first derivative of $Q_{PC}(\boldsymbol{\theta})$ with respect to α to zero we obtain

$$\alpha = \frac{n}{\sum_{i=1}^{n}\log(1+t_i^{\beta}) - \sum_{i\in S_2}\log(1+\kappa_{Li}^{\beta})} = \alpha(\beta). \tag{3.4}$$

Substituting (3.4) in (3.3), we obtain the pseudo-profile log-likelihood in β as

$$\begin{aligned} p_{PC}(\beta) &= n\log\beta + \beta\sum_{i=1}^{n}\log t_i - n\log\left\{\sum_{i=1}^{n}\log(1+t_i^{\beta}) - \sum_{i\in S_2}\log(1+\kappa_{Li}^{\beta})\right\} \\ &\quad - \sum_{i=1}^{n}\log(1+t_i^{\beta}). \end{aligned} \tag{3.5}$$

Note that maximizing $p_{PC}(\beta)$ in (3.5) is a one-dimensional optimization problem, and can be achieved by using any routine optimizer of a statistical software, for example, the maxNR() function in the "maxLik" package [12] available in R software.

The following algorithm implements the St-EM algorithm for obtaining estimates of the model parameters.

Algorithm 1: At the k-th stage of the algorithm:

Stochastic Expectation (St-E) step:

Step 1: The available parameter value is $\boldsymbol{\theta}^{(k)} = (\alpha^{(k)}, \beta^{(k)})$;

Step 2: Replace each unit in $\boldsymbol{\Gamma}$ by generating observations from $f_{T_i|T_i>y_i}(t_i|t_i>y_i;\boldsymbol{\theta}^{(k)})$ to obtain pseudo-complete data $\boldsymbol{t}_{PC}$;

Step 3: With $\boldsymbol{t}_{PC}$ thus obtained in Step 2, construct $Q_{PC}(\boldsymbol{\theta}^{(k)})$ following (3.3);

Maximization (M) step:

Step 4: Choose an initial value $\beta^{(k)}_{init}$;

Step 5: Optimize $p_{PC}(\beta)$ in (3.5) to get $\widehat{\beta^{(k+1)}}$ subject to a tolerance level;

Step 6: Using (3.4), calculate $\widehat{\alpha^{(k+1)}} = \alpha(\widehat{\beta^{(k+1)}})$;

Step 7: With the updated estimate $\boldsymbol{\theta}^{(k+1)} = (\alpha^{(k+1)}, \beta^{(k+1)})$, go back to Step 2.

These steps are iterated N times to get a sequence of estimates $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$, ..., $\boldsymbol{\theta}^{(N)}$. After discarding first M of these estimates for burn-in, the remaining ones are averaged to get estimates of α and β. As mentioned in Ye and Ng [23], sufficiently large values of N and M must be chosen for very complex data, while the values as 1000 and 100, respectively may be good enough for most problems.

The algorithm starts with an initial value for the parameter vector $\boldsymbol{\theta}^{(0)} = (\alpha^{(0)}, \beta^{(0)})$. For such optimization problems to numerically estimate the parameters, moments estimates may be used as the initial values, provided the moments of the concerned distribution exist and are easily available, for example, in closed form expressions. For the Burr XII distribution, however, moments estimates based on left truncated data are not available in closed form. A practical solution to the problem of selecting initial values for the parameters α and β in this case would be to use a two-dimensional grid search approach. However, it may be noted here that use of a two-dimensional grid search approach followed by the St-EM algorithm will be computationally costly. In this work, for given sets of true values of the parameters α and β, we have tried different arbitrarily chosen initial values for the St-EM algorithm. And we have noticed that the St-EM algorithm as described above is reasonably robust to the choice of initial values. That is, the final estimates obtained from the algorithm by using different choices of initial values are quite close. In this connection, it may be noted here that the direct method of optimization, for example, the Newton-Raphson method, is heavily dependent on the choice of initial parameters in general.

3.1. Regression

In many applications, failure-times of experimental units depend on covariates. For example, failure-time of electrical machines may depend on temperature and humidity of the place of operation; failure-time of patients may depend on their respective demographic conditions etc. In view of this, it is of interest to consider a failure-time model that can accommodate relevant covariate information.

Regression models for Burr XII distribution have been considered by some authors. Beirlant *et al.* [6], in the context of a financial application involving portfolio segmentation, discussed different strategies for accommodating covariate information in the Burr XII model through the shape and scale parameters. Regression model for the log-Burr XII distribution was considered by Silva *et al.* [19], where covariate information was modelled as a linear function of the location parameter of the log-transformed model.

In this paper, we use an approach suggested in Beirlant *et al.* [6]. However, for parameter estimation, instead of using the observed likelihood based estimation approach which may not be computationally stable for complex models, we indicate the use of stochastic EM algorithm which is more reliable for its convergence.

We allow the shape parameter β in (1.1) to vary with covariates. Thus, when the $\boldsymbol{x}$ represent the vector of covariates, we assume the model

$$\beta(\boldsymbol{x}) = \exp(\boldsymbol{\gamma}'\boldsymbol{x}), \tag{3.6}$$

where $\boldsymbol{\gamma}$ is the vector of regression parameters. Under this assumption, our model for the failure-time variable Y becomes

$$Y_i|\boldsymbol{x}_i \sim Burr(\alpha, \beta_i), \quad \text{with} \quad \beta_i = \exp(\boldsymbol{\gamma}'\boldsymbol{x}_i), \quad i = 1, ..., n. \tag{3.7}$$

The conditional distributions for generating observations are given in this case by

$$\begin{aligned} f_{T_i|T_i>y_i}(t_i|t_i > y_i; \alpha, \boldsymbol{\gamma}, \boldsymbol{x}_i) \; &= \; \alpha \exp(\boldsymbol{\gamma}'\boldsymbol{x}_i) t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)-1} (1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})^{-(\alpha+1)} \\ &\quad \times (1 + y_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})^{\alpha}, \qquad t_i > y_i. \end{aligned} \tag{3.8}$$

Corresponding to the censored failure-times, given the covariates, observations are generated from (3.8). Based on the pseudo-complete data, the pseudo-complete likelihood function is

$$\begin{aligned} L_{PC}(\alpha, \boldsymbol{\gamma}) \; &= \; \prod_{i=1}^{n} \left\{ \alpha \exp(\boldsymbol{\gamma}'\boldsymbol{x}_i) t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)-1} (1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})^{-(\alpha+1)} \right\} \\ &\quad \times \prod_{i \in S_2} \left\{ (1 + \kappa_{Li}^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})^{\alpha} \right\}. \end{aligned} \tag{3.9}$$

The corresponding pseudo-complete log-likelihood function is given by

$$\begin{aligned} \log L_{PC}(\alpha, \boldsymbol{\gamma}) \; &= \; n \log \alpha + \sum_{i=1}^{n} \boldsymbol{\gamma}'\boldsymbol{x}_i + \sum_{i=1}^{n} (\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i) - 1) \log t_i \\ &\quad - \; (\alpha + 1) \sum_{i=1}^{n} \log(1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) + \alpha \sum_{i \in S_2} \log(1 + \kappa_{Li}^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}), \end{aligned} \tag{3.10}$$

which is also the pseudo-Q function, denoted by $Q_{PC}(\alpha, \boldsymbol{\gamma})$. Equating the first derivative of (3.10) with respect to α to zero, we have

$$\alpha = \frac{n}{\sum_{i=1}^{n} \log(1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) - \sum_{i \in S_2} \log(1 + \kappa_{Li}^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)})} = \alpha(\boldsymbol{\gamma}). \tag{3.11}$$

Substituting (3.11) in (3.10), the profile-likelihood in $\boldsymbol{\gamma}$ is obtained as

$$\begin{aligned} p_{PC}(\boldsymbol{\gamma}) &= \sum_{i=1}^{n} \boldsymbol{\gamma}'\boldsymbol{x}_i + \sum_{i=1}^{n} \exp(\boldsymbol{\gamma}'\boldsymbol{x}_i) \log t_i - \sum_{i=1}^{n} \log(1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) \\ &\quad - n \log\left\{ \sum_{i=1}^{n} \log(1 + t_i^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) - \sum_{i \in S_2} \log(1 + \kappa_{Li}^{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}) \right\}. \end{aligned} \tag{3.12}$$

The profile log-likelihood in the regression parameters $\boldsymbol{\gamma}$ can be maximized first using some numerical approach such as Newton-Raphson or Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Then, the α can be estimated using (3.11). An algorithm similar to Algorithm 1 can be easily constructed for this purpose.

3.2. Asymptotic confidence intervals

For obtaining asymptotic confidence intervals for the parameters, we use the missing information principle of Louis [15] that says

$$\text{Observed Information} = \text{Complete Information} - \text{Missing Information}. \tag{3.13}$$

For the traditional EM algorithm, Louis' principle is used to obtain the asymptotic variances of the estimates. For the St-EM algorithm also, an adaptation of the Louis' principle is possible, see Ye and Ng [23]. Mitra *et al.* [16] used the same approach in connection to the Lehmann family of distributions. The approach of Ye and Ng [23] is as follows.

Let $S(\boldsymbol{\theta}, \boldsymbol{t}_{PC})$ and $H(\boldsymbol{\theta}, \boldsymbol{t}_{PC})$ denote the first, and negative of the second derivatives of $Q_{PC}(\boldsymbol{\theta})$ given in (3.3) with respect to $\boldsymbol{\theta}$. Then, by the missing information principle, following Ye and Ng [23], the observed information matrix is given by

$$I(\boldsymbol{\theta}) = E[H(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}] - E[S^2(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}] + \{E[S(\boldsymbol{\theta}, \boldsymbol{t})|\boldsymbol{y}]\}^2. \tag{3.14}$$

For evaluating $I(\boldsymbol{\theta})$ in (3.14) for a given a LTRC data, multiple samples $\boldsymbol{\Gamma}_{PC}^{(m)}$, $m = 1, ..., M$ are imputed corresponding to the censored data $\boldsymbol{\Gamma} = \{i : \delta_i = 0\}$, thus obtaining multiple pseudo-complete datasets $\boldsymbol{t}_{PC}^{(m)}$, $m = 1, ..., M$. Then, $I(\boldsymbol{\theta})$ is estimated as

$$\widehat{I(\boldsymbol{\theta})} = \frac{1}{M} \sum_{m=1}^{M} H(\boldsymbol{\theta}, \boldsymbol{t}_{PC}^{(m)}) - \frac{1}{M} \sum_{m=1}^{M} [S(\boldsymbol{\theta}, \boldsymbol{t}_{PC}^{(m)})]^2 + \left[\frac{1}{M} \sum_{m=1}^{M} S(\boldsymbol{\theta}, \boldsymbol{t}_{PC}^{(m)}) \right]^2 \Bigg|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}}. \tag{3.15}$$

Finally, the asymptotic variance-covariance matrix of the estimates is obtained by inverting $\widehat{I(\boldsymbol{\theta})}$, and the asymptotic confidence intervals for the parameters can be constructed using the asymptotic variances.

A second approach we use here is based on the bootstrap procedure. Bootstrap confidence intervals [11] are widely used in statistical literature. These intervals are particularly of interest for LTRC data, as the presence of truncation and censoring often tend to bias the estimates of parameters. We use the following algorithm to obtain parametric bootstrap confidence intervals. Here, SP is the starting point, and TP is the termination point of units.

ALGORITHM 2:

STEP 1: Based on the given LTRC data, obtain the estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta})$;

STEP 2: Construct empirical distribution of SPs for left truncated units;

STEP 3: Construct empirical distribution of SPs for untruncated units;

STEP 4: To get a bootstrap sample (preserving proportion of truncation):

- STEP 4.1: Sample SPs for truncated units from empirical distribution of Step 2;
- STEP 4.2: Sample SPs for untruncated units from empirical distribution of Step 3;
- STEP 4.3: Generate failure-times from Burr$(\widehat{\alpha}, \widehat{\beta})$;
- STEP 4.4: Add failure-times to SPs, to obtain corresponding TPs;
- STEP 4.5: Determine censoring status of units according to their TPs;

STEP 5: For this bootstrap sample, obtain bootstrap estimate $\widehat{\boldsymbol{\theta}}^* = (\widehat{\alpha}^*, \widehat{\beta}^*)$;

STEP 6: Repeat Steps 4 and 5 B times, to obtain $\widehat{\boldsymbol{\theta}}_1^*, \widehat{\boldsymbol{\theta}}_2^*, ..., \widehat{\boldsymbol{\theta}}_B^*$.

The value of B, i.e., the number of bootstrap samples, should be sufficient to stabilize the estimated bootstrap bias and variance of $\widehat{\boldsymbol{\theta}}$. For constructing parametric bootstrap confidence intervals based on LTRC data from the Burr XII distribution, we recommend using $B \geq 200$.

A $100(1-\delta)\%$ parametric bootstrap confidence interval for α is then given by

$$\left(\widehat{\alpha} - b_\alpha - z_{\delta/2}\sqrt{v_\alpha},\ \widehat{\alpha} - b_\alpha + z_{\delta/2}\sqrt{v_\alpha}\right),$$

where b_α and v_α are the bootstrap bias and bootstrap variance, respectively. Here, z_δ is the upper δ-percentile point of standard normal distribution. The $100(1-\delta)\%$ parametric bootstrap confidence intervals for β are constructed in a similar way.

3.3. Direct optimization of observed likelihood

Considering the four different types of units with different combinations of left truncation and right censoring as mentioned in Figure 3 in Section 2, the observed likelihood for LTRC data is given by

$$\begin{aligned} L(\boldsymbol{\theta}|DATA) = & \prod_{i\in S_1} \left\{f(t_i;\boldsymbol{\theta})\right\}^{\delta_i}\left\{1-F(t_i;\boldsymbol{\theta})\right\}^{1-\delta_i} \\ & \times \prod_{i\in S_0} \left\{\frac{f(t_i;\boldsymbol{\theta})}{1-F(\kappa_{Li};\boldsymbol{\theta})}\right\}^{\delta_i}\left\{\frac{1-F(t_i;\boldsymbol{\theta})}{1-F(\kappa_{Li};\boldsymbol{\theta})}\right\}^{1-\delta_i}. \end{aligned} \tag{3.16}$$

By plugging in the PDF and CDF of the Burr XII distribution in (3.16), we get the specific likelihood for the Burr XII distribution based on LTRC data. The likelihood (or the corredponding log-likelihood) function may then be maximized using routine functions in statistical software. The performance of the estimates obtained by the St-EM algorithm, and those obtained by the direct method of optimization based on observed likelihood can then be compared through Monte Carlo simulations. In this paper, we have used the "maxLik" package in R software for maximizing the observed likelihood provided in (3.16). In particular, we have used the Newton-Raphson method for direct numerical optimization; the Newton-Raphson method may be employed by using the maxNR() routine available in the maxLik package. Details of the numerical results are presented in the next Section.

4. NUMERICAL EXPERIMENTS

The methods of inference are assessed through Monte Carlo simulations using the R software. For simulating LTRC data, the following process is followed. We consider lifetime data at the yearly scale. Left truncation and right censoring points are fixed at the years 2000 and 2004, respectively, without loss of generality. That is, in connection to the notations used in Section 2, we set $L = 2000$, and $R = 2004$. The total sample size n is fixed; here, we consider several values of n, namely, 50, 100, 200, 300, and 500.

First, a truncation percentage p ($0 \leq p \leq 1$) is specified; this implies that in the sample, there will be np left truncated units, and $n(1-p)$ untruncated units. For this simulation study, p is chosen as 20, and 30. Two arbitrary sets of years as installation points (IPs), say W_{LT} and W_{NT}, respectively, are chosen; W_{LT} corresponds to the left truncated group (i.e., less than 2000), and W_{NT} corresponds to the untruncated group (i.e., more than or equal to 2000), are taken as reference frames from sampling. Then, through equal probability sampling, two sets of samples of IPs are generated from W_{LT} and W_{NT} according to the pre-specified truncation percentage; these samples represent the left truncated and the untruncated groups, respectively. For example, for $n = 100$, and $p = 20$, a sample of 20 IPs is taken from W_{LT}, and a sample of 80 IPs is taken from W_{NT}.

For generating lifetimes from the Burr XII distributions, two sets of values for the model parameters are used. A LTRC dataset of size n is generated as follows. Corresponding to each IP ω_i, with $\omega_i \in W$, $i = 1, ..., n$, where $W = W_{LT} \bigcup W_{NT}$, a failure-time y_i is generated from the Burr XII distribution in (1.2), and is added to ω_i to obtain the respective termination point (TP). Left truncation and right censoring, corresponding to $L = 2000$ and $R = 2004$, are incorporated into the generated data through the following mechanism. For $i \in W_{LT}$, if $\omega_i + y_i < 2000$ for a unit, that unit is completely discarded, and is replaced by a new set of values for ω_i, and y_i; this process of discarding continues until for that unit we have $\omega_i + y_i > 2000$. This ensures that all units have to cross a threshold to be included in the study, as required by left truncation. For a unit $i \in W$, if $\omega_i + y_i > 2004$, it is a right-censored unit; otherwise, it is not censored. The chosen values of the parameters of the Burr XII distribution ensures that there are enough censored units.

The bias, and mean squared error (MSE) of the point estimates for different simulation parameter settings are reported in Tables 1–6. The coverage probability and average length of the asymptotic confidence intervals for the model parameters are also reported in these tables.

Table 1: Performance of point and interval estimates for truncation percentage 20, corresponding to true parameter value $(\alpha, \beta) = (2, 0.5)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.051	0.123	0.037	0.121	0.960	1.298	0.960	1.335
	β	0.014	0.004	0.012	0.004	0.954	0.253	0.956	0.262
100	α	0.028	0.054	0.021	0.053	0.942	0.907	0.950	0.915
	β	0.005	0.002	0.004	0.002	0.950	0.177	0.946	0.179
200	α	0.009	0.025	0.005	0.025	0.958	0.633	0.958	0.639
	β	0.003	0.001	0.002	0.001	0.958	0.124	0.952	0.125
300	α	0.013	0.018	0.011	0.018	0.952	0.516	0.946	0.519
	β	0.004	0.001	0.004	0.001	0.956	0.102	0.956	0.102
500	α	0.002	0.011	0.001	0.011	0.932	0.396	0.948	0.400
	β	0.001	0.000	0.001	0.000	0.954	0.078	0.954	0.079

Table 2: Performance of point and interval estimates for truncation percentage 30, corresponding to true parameter value $(\alpha, \beta) = (2, 0.5)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.026	0.103	0.012	0.102	0.966	1.313	0.964	1.342
	β	0.017	0.005	0.015	0.005	0.956	0.263	0.956	0.273
100	α	0.013	0.054	0.006	0.053	0.956	0.919	0.950	0.931
	β	0.009	0.002	0.008	0.002	0.962	0.183	0.960	0.186
200	α	0.018	0.030	0.015	0.030	0.940	0.649	0.950	0.650
	β	0.003	0.001	0.002	0.002	0.952	0.128	0.952	0.128
300	α	0.012	0.019	0.010	0.019	0.962	0.527	0.958	0.529
	β	0.000	0.001	−0.000	0.001	0.942	0.104	0.950	0.104
500	α	0.004	0.012	0.003	0.012	0.932	0.401	0.942	0.411
	β	0.000	0.000	0.000	0.000	0.934	0.080	0.934	0.081

Table 3: Performance of point and interval estimates for truncation percentage 50, corresponding to true parameter value $(\alpha, \beta) = (2, 0.5)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.027	0.125	0.015	0.124	0.956	1.410	0.956	1.436
	β	0.015	0.005	0.013	0.005	0.968	0.287	0.968	0.304
100	α	0.026	0.074	0.019	0.073	0.930	0.989	0.934	1.000
	β	0.006	0.003	0.005	0.003	0.948	0.199	0.962	0.204
200	α	0.014	0.030	0.010	0.030	0.946	0.693	0.952	0.701
	β	0.001	0.001	0.001	0.001	0.956	0.139	0.956	0.141
300	α	0.012	0.020	0.010	0.020	0.946	0.563	0.952	0.569
	β	0.001	0.001	0.001	0.001	0.970	0.114	0.964	0.114
500	α	0.008	0.012	0.006	0.012	0.960	0.432	0.966	0.441
	β	0.001	0.001	0.001	0.001	0.962	0.088	0.958	0.088

Table 4: Performance of point and interval estimates for truncation percentage 20, corresponding to true parameter value $(\alpha, \beta) = (3, 1)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.080	0.231	0.076	0.231	0.946	1.764	0.955	1.875
	β	0.024	0.012	0.023	0.012	0.957	0.425	0.955	0.441
100	α	0.032	0.098	0.029	0.098	0.958	1.225	0.952	1.265
	β	0.014	0.006	0.014	0.006	0.946	0.298	0.946	0.305
200	α	0.015	0.052	0.014	0.052	0.952	0.862	0.950	0.875
	β	0.004	0.003	0.004	0.003	0.948	0.209	0.944	0.211
300	α	0.014	0.030	0.013	0.030	0.962	0.703	0.964	0.705
	β	0.006	0.002	0.006	0.002	0.944	0.171	0.952	0.172
500	α	−0.005	0.016	−0.005	0.016	0.960	0.541	0.964	0.543
	β	0.002	0.001	0.002	0.001	0.958	0.132	0.954	0.132

Table 5: Performance of point and interval estimates for truncation percentage 30, corresponding to true parameter value $(\alpha, \beta) = (3, 1)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.110	0.290	0.106	0.209	0.947	1.772	0.937	1.866
	β	0.021	0.012	0.020	0.012	0.968	0.428	0.958	0.446
100	α	−0.002	0.105	−0.004	0.105	0.944	1.209	0.942	1.239
	β	0.010	0.006	0.009	0.006	0.964	0.301	0.958	0.308
200	α	0.013	0.046	0.012	0.046	0.964	0.858	0.966	0.869
	β	0.005	0.003	0.005	0.003	0.942	0.211	0.944	0.214
300	α	−0.004	0.035	−0.005	0.035	0.946	0.697	0.946	0.701
	β	0.004	0.002	0.004	0.002	0.956	0.173	0.948	0.174
500	α	0.008	0.017	0.008	0.017	0.962	0.542	0.968	0.542
	β	0.002	0.001	0.002	0.001	0.958	0.133	0.956	0.133

Table 6: Performance of point and interval estimates for truncation percentage 50, corresponding to true parameter value $(\alpha, \beta) = (3, 1)$. Coverage probability (CP) and average length (AL) are reported for 95% confidence intervals corresponding to missing information principle (MI) and parametric bootstrap (PB).

n	Parm	St-EM		NR		MI		PB	
		Bias	MSE	Bias	MSE	CP	AL	CP	AL
50	α	0.053	0.234	0.050	0.234	0.946	1.810	0.950	1.873
	β	0.027	0.015	0.026	0.015	0.948	0.465	0.954	0.491
100	α	0.027	0.112	0.026	0.112	0.944	1.269	0.948	1.284
	β	0.008	0.007	0.007	0.007	0.950	0.323	0.958	0.332
200	α	−0.004	0.057	−0.004	0.057	0.944	0.889	0.928	0.891
	β	0.013	0.003	0.013	0.003	0.964	0.230	0.970	0.232
300	α	0.009	0.030	0.009	0.030	0.942	0.729	0.944	0.728
	β	0.002	0.002	0.002	0.002	0.944	0.185	0.940	0.187
500	α	0.005	0.020	0.005	0.020	0.940	0.564	0.944	0.562
	β	0.004	0.001	0.004	0.001	0.946	0.144	0.944	0.145

The Monte Carlo estimate of coverage probability of an asymptotic confidence interval corresponding to a nominal level of confidence (say, 95%) is the proportion of times the asymptotic confidence interval includes the true parameter value out of the total number of Monte Carlo runs of the experiment. The average length of an asymptotic confidence interval is the mean length of the interval, averaged over the lengths obtained in the Monte Carlo runs.

From Tables 1–6, we notice that the estimates obtained by the St-EM algorithm are quite efficient in general, with respect to their bias and MSE. As one would expect, with increase in sample size, the bias and MSE of the estimates reduce. Truncation percentage does not seem to have a significant effect on the point estimates, as the bias and MSE values do not change much with change in truncation percentage.

It may also be of interest to compare the results of the St-EM algorithm with that of the Newton-Raphson method. It is observed from Tables 1–6 that for parameter α, the biases corresponding to the St-EM algorithm and the Newton-Raphson method are to some extent different for smaller sample sizes (i.e., $n = 50$ and 100). However, with increase in sample size (i.e., for $n = 200, 300, 500$), the biases become very close. For parameter β, the biases of the estimates corresponding to the two methods are close for all simulation settings considered here. Finally, the MSE of the estimates of both α and β are always quite close for the two methods.

Tables 1–6 also report coverage probabilities (CP) and average lengths (AL) for asymptotic 95% confidence intervals. The coverage probability and average length are two important criteria for assessing the performance of confidence intervals. For a confidence interval to be reasonable, its coverage probability should be close to the nominal confidence level, and its average length should not be large. It may be noted that the coverage probabilities corresponding to the missing information principle are always very close to the nominal level. The coverage probabilities corresponding to the parametric bootstrap are also close to the nominal level. With respect to average length of the intervals, both methods perform closely. It is also observed that with increase in sample size, though their average lengths reduce as expected, but the confidence intervals are able to retain the coverage probability close to the nominal level.

5. THE THREE PARAMETER BURR XII DISTRIBUTION

Considering a scale parameter λ along with the two shape parameters α and β, the PDF of the three parameter Burr XII distribution is given by (see [24])

$$f_Y(y;\lambda,\alpha,\beta) = \frac{\alpha\beta}{\lambda}\left(\frac{y}{\lambda}\right)^{\beta-1}\left(1+\left(\frac{y}{\lambda}\right)^{\beta}\right)^{-(\alpha+1)}, \quad y>0, \quad \lambda,\alpha,\beta>0.$$

For the St-EM algorithm, the conditional distributions for generating random observations is given by

$$f_{T_i|T_i>y_i}(t_i|t_i>y_i;\lambda,\alpha,\beta) = \frac{\alpha\beta}{\lambda}\left(\frac{t_i}{\lambda}\right)^{\beta-1}\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right)^{-(\alpha+1)}\left(1+\left(\frac{y_i}{\lambda}\right)^{\beta}\right)^{\alpha}, \quad t_i>y_i. \tag{5.1}$$

After replacing the right censored failure-times by randomly generated observations from (5.1), the pseudo-Q function is obtained as

$$\begin{aligned} Q_{PC}(\boldsymbol{\theta}) &= n(\log\alpha + \log\beta - \beta\log\lambda) + (\beta-1)\sum_{i=1}^{n}\log t_i - \sum_{i=1}^{n}\log\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right) \\ &\quad - \alpha\left[\sum_{i=1}^{n}\log\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right) - \sum_{i\in S_2}\log\left(1+\left(\frac{\kappa_{Li}}{\lambda}\right)^{\beta}\right)\right], \end{aligned} \tag{5.2}$$

with $\boldsymbol{\theta} = (\lambda, \alpha, \beta)$. For fixed λ and β, equating the first derivative of $Q_{PC}(\boldsymbol{\theta})$ with respect to α to zero, we obtain

$$\alpha = \frac{n}{W_{PC}(\lambda,\beta)}, \tag{5.3}$$

where

$$W_{PC}(\lambda,\beta) = \sum_{i=1}^{n}\log\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right) - \sum_{i\in S_2}\log\left(1+\left(\frac{\kappa_{Li}}{\lambda}\right)^{\beta}\right).$$

Substituting (5.3) in (5.2), the profile log-likelihood in λ and β is obtained as

$$p_{PC}(\lambda,\beta) = n(\log\beta - \beta\log\lambda - \log W_{PC}(\lambda,\beta)) + (\beta-1)\sum_{i=1}^{n}\log t_i - \sum_{i=1}^{n}\log\left(1+\left(\frac{t_i}{\lambda}\right)^{\beta}\right), \tag{5.4}$$

which can then be maximized by a routine two-parameter optimizer.

Starting with an initial value for the parameter vector as $\boldsymbol{\theta}^{(0)} = (\lambda^{(0)}, \alpha^{(0)}, \beta^{(0)})$, to the choice of which the St-EM algorithm is quite robust, the following are the steps of the St-EM algorithm for the three-parameter Burr XII distribution based on LTRC data.

Algorithm 3: At the k-th stage of the algorithm:

Stochastic Expectation (St-E) step:

Step 1: The available parameter value is $\boldsymbol{\theta}^{(k)} = (\lambda^{(k)}, \alpha^{(k)}, \beta^{(k)})$;

Step 2: Replace each unit in $\boldsymbol{\Gamma}$ by generating observations from $f_{T_i|T_i>y_i}(t_i|t_i>y_i;\boldsymbol{\theta}^{(k)})$ to obtain pseudo-complete data $\boldsymbol{t}_{PC}$;

Step 3: With $\boldsymbol{t}_{PC}$ from Step 2, construct $Q_{PC}(\boldsymbol{\theta}^{(k)})$ following (5.2);

Maximization (M) step:

Step 4: Choose initial values $\lambda_{init}^{(k)}$ and $\beta_{init}^{(k)}$ based on the pseudo-complete data $\boldsymbol{t}_{PC}$;

Step 5: Optimize $p_{PC}(\lambda,\beta)$ in (5.4) to get $\widehat{\lambda^{(k+1)}}$ and $\widehat{\beta^{(k+1)}}$ subject to a tolerance level;

Step 6: Using (5.3), calculate $\widehat{\alpha^{(k+1)}} = \frac{n}{W_{PC}(\lambda^{(k+1)},\beta^{(k+1)})}$;

Step 7: With updated estimate $\boldsymbol{\theta}^{(k+1)} = (\lambda^{(k+1)}, \alpha^{(k+1)}, \beta^{(k+1)})$, go back to Step 2.

From a sequence of estimates $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, ..., \boldsymbol{\theta}^{(N)}$, after discarding first M estimates for burn-in, the remaining ones are averaged to get estimates $\widehat{\lambda}$, $\widehat{\alpha}$, and $\widehat{\beta}$. The confidence intervals for the model parameters may be obtained by similar processes as described in Section 3 for the two-parameter Burr XII distribution.

It may be mentioned here that in case of LTRC data from the three-parameter Burr XII distribution, based on some limited simulations, we have noticed an indication that the St-EM algorithm performs better compared to the direct optimization of observed likelihood, in terms of bias and MSE of the estimates. However, implementation of the St-EM algorithm in this case is very challenging due to its computational cost; the running time of the St-EM algorithm for the three-parameter case is significantly longer than that of the direct optimization method.

6. PREDICTION OF EXPECTED NUMBER OF FAILURES IN A FUTURE INTERVAL

Consider the right censored units with the i-th unit having right censored lifetime y_i, $i \in S_{cen}$. Consider a future interval $(\tau_1, \tau_2]$ with $t_{max} < \tau_1$, where $t_{max} = \text{Max}\{t_i; i \in S_{cen}\}$. The probability that the i-th unit fails in this interval $(\tau_1, \tau_2]$ is given by

$$\pi_i = P(\tau_1 < T_i \leq \tau_2 | T_i > y_i) = \frac{S(\tau_1; \boldsymbol{\theta}) - S(\tau_2; \boldsymbol{\theta})}{S(y_i; \boldsymbol{\theta})}, \tag{6.1}$$

where $S(t) = P(T > t)$ is the survival function of the underlying failure-time variable T. Note that the expression for this probability remains same regardless of the truncation status of the i-th unit, $i \in S_{cen}$. We are interested in obtaining the expected number of failures in the future interval $(\tau_1, \tau_2]$.

Let us define random variables U_i, $i \in S_{cen}$, such that

$$U_i = \begin{cases} 1, & \text{if } i\text{-th item fails in } (\tau_1, \tau_2] \\ 0, & \text{otherwise.} \end{cases}$$

Note that $E[U_i] = P(U_i = 1) = \pi_i$. We want to obtain the expected number of failures in the future interval (τ_1, τ_2), given by

$$\zeta = E\left[\sum_{i \in S_{cen}} U_i\right] = \sum_{i \in S_{cen}} \pi_i.$$

Now, using the expression for π_i in (6.1), we obtain

$$\begin{aligned} \zeta &= \{(1+\tau_1^{\beta})^{-\alpha} - (1+\tau_2^{\beta})^{\alpha}\} \sum_{i \in S_{cen}} (1+y_i^{\beta})^{\alpha} \\ &= h(\boldsymbol{\theta}) \quad \text{(say)}. \end{aligned} \tag{6.2}$$

Clearly, an estimate $\widehat{\zeta}$ of the expected number of failures ζ can be obtained by simply plugging-in the estimated parameters in (6.2). It is also possible to provide an asymptotic confidence interval for ζ by a straightforward application of the delta-method, by using the asymptotic normality and the delta-method. That is, using the fact that $\sqrt{n}(\widehat{\zeta} - \zeta) \xrightarrow{D} N(0, \mathrm{Var}(\widehat{\zeta}))$, where the variance can be estimated as

$$(6.3)\qquad \widehat{\mathrm{Var}(\widehat{\zeta})} = \left(\left(\frac{\partial h}{\partial \alpha}\right)^2 \mathrm{Var}(\widehat{\alpha}) + 2\left(\frac{\partial h}{\partial \alpha}\right)\left(\frac{\partial h}{\partial \beta}\right) \mathrm{Cov}(\widehat{\alpha}, \widehat{\beta}) + \left(\frac{\partial h}{\partial \beta}\right)^2 \mathrm{Var}(\widehat{\beta}) \right)\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} .$$

Finally, using the estimated variance, an asymptotic $100(1-\gamma)\%$ confidence interval for ζ can be easily obtained.

7. ILLUSTRATIVE DATA ANALYSIS

The Channing House data involves lives of residents of a retirement centre in Palo Alto, California. The dataset contains lifetimes of residents of the centre since it started operations in 1965 till July, 1975. A person had to be at least 60 years of age to be a resident of the centre; this fact incorporated left truncation in the data. In fact, due to this restriction on the entry of individuals to the centre, the entire data (i.e., 100% of the observations) is left truncated according to the notion of left truncation followed in here.

Some individuals died as residents of the centre, while some other were still alive when the collection ended in July, 1975. This incorporated right censoring in the data. The dataset contains lives of total 462 residents. Out of 462, the number of observed failures is only 176, and the rest of the units are right censored. A summary of the dataset is presented in Table 7.

Table 7: Summary of Channing House Data.

Group	Total number	Right censoring	Mean lifetime (Years)	SD lifetimes (Years)
Male	97	52.58%	82.63	6.14
Female	365	64.38%	82.04	6.15
Combined	462	61.90%	82.17	6.15

Before analyzing the dataset, we change the origin and scale of this data by subtracting 720 from each of the lifetimes (and left truncation times), and by dividing them by 200; this change of origin and scale of the data will not impact the inferential results in any way. We assume that the underlying failure-time variable follow a two-parameter Burr XII distribution. The results of point and interval estimation by the proposed methods are given in Table 8. The estimated parameters can then be used in further analyses, for example, in predicting future failures as described below.

Table 8: Point and interval estimates of model parameters for Channing House Data.

Parameter	Point Estimate	Interval Estimate	
		MI	BB
α	0.508	$(0.354, 0.662)$	$(0.354, 0.655)$
β	3.976	$(2.915, 5.038)$	$(2.764, 5.002)$

As the estimates of the parameters turn out to be $\widehat{\alpha} = 0.508$ and $\widehat{\beta} = 3.976$, using these, we can obtain the asymptotic variance-covariance matrix by using the missing information principle as described in Section 3.2 as

$$\begin{pmatrix} \mathrm{Var}(\widehat{\alpha}) & \mathrm{Cov}(\widehat{\alpha,\beta}) \\ & \mathrm{Var}(\widehat{\beta}) \end{pmatrix} = \begin{pmatrix} 0.0061 & -0.0369 \\ & 0.2934 \end{pmatrix}.$$

The maximum value among the transformed right censored lifetimes is 2.435. Suppose, we are interested in predicting the expected number of failures in the interval $(2.5, 2.7]$. By plugging-in the estimates $\widehat{\alpha}$ and $\widehat{\beta}$ in (6.2), we get the expected number of failures in the above interval as $\widehat{\zeta} = 13.4000$. Finally, upon estimating the variance of $\widehat{\zeta}$ by (6.3) to be 0.0490, a 95% confidence interval for the expected number of failures in this future interval is obtained as $(12.9660, 13.8339)$.

Suppose it is of interest to select an appropriate model for this dataset among many candidate models. One way to achieve this would be fit different models to the dataset, and then to choose the model for which the value of the maximized log-likelihood, evaluated at the MLE, is the largest. Naturally, the distributions which are frequently used to model lifetime data would be the candidate models. As suggested by a reviewer, here, we consider Weibull, Gompertz, and Lomax distributions as the candidate models, along with the Burr XII model. It may be mentioned here that Weibull, Gompertz, and Lomax distributions belong to a family of distributions known as the Lehmann family of distributions [16].

Table 9 gives the results of the model selection. We fit the Burr XII model to the Channing House data by using the St-EM algorithm; we also fit Weibull, Gompertz, and Lomax distributions to the data by using St-EM algorithm. Then, we evaluate the log-likelihood functions corresponding to the four distributions at the respective MLEs, the log-likelihood being constructed by using the LTRC data structure. It may be mentioned here that since all the models considered here have same number of parameters, the process of using the maximized log-likelihood is essentially equivalent to using the Akaike's information criterion (AIC) for model selection.

Table 9: Maximized log-likelihood for different models.

Model	Distribution Function	Maximized log-likelihood
Weibull	$F_W(t;\lambda,\alpha) = 1 - e^{-\alpha t^{\lambda}}, \quad t > 0$	-155.9704
Gompertz	$F_G(t;\lambda,\alpha) = 1 - e^{-\alpha(e^{\lambda t}-1)}, \quad t > 0$	-152.9099
Lomax	$F_L(t;\lambda,\alpha) = 1 - \left[\frac{1}{1+\lambda t}\right]^{\alpha}, \quad t > 0$	-189.7542
Burr XII	$F_B(t;\alpha,\beta) = 1 - (1+t^{\beta})^{-\alpha}, \quad t > 0$	-181.7247

It turns out that the maximized log-likelihood is the largest for the Gompertz distribution based on this data. Therefore, the Gompertz distribution turns out to be the most suitable model for the Channing House data by the above criterion.

8. CONCLUSION AND FUTURE WORK

In this article, statistical inferential procedures for the Burr XII distribution based on LTRC data are discused. the two- and the three-parameter Burr XII models are considered. Detailed steps of the stochastic EM algorithm based on LTRC data are developed for obtaining point estimates of the model parameters. Two methods for constructing asymptotic confidence intervals of the parameters are discussed: one by using the missing information principle, and the other by using a parametric bootstrap approach. A method for including covariates in the Burr XII model in this setup is also discussed. An application of the estimated parameters in predictive number of failures in a future interval is presented.

From the numerical results of a detailed Monte Carlo simulation study, it is observed that the stochastic EM algorithm performs reasonably well in estimating the model parameters. The approaches for constructing confidence intervals also perform satisfactorily, as the coverage probabilities of the confidence intervals remain always close to the nominal confidence level of 95%. It is also observed that the performance of the St-EM algorithm is close to that of the Newton-Raphson method.

While parametric inference can generate accurate results when the assumptions regarding the underlying distribution of data are appropriate, it may be of interest to verify whether the distributional assumptions are reasonable or not. In view of this, it will be of interest to develop a test for goodness of fit for the Burr XII distribution based on LTRC data.

Another problem of interest would be to study Bayesian inference for the Burr XII distribution based on LTRC data. The Bayesian methods can provide significant information regarding a model, especially when the prior assumptions are appropriate. In particular, the Bayesian methods may outperform classical inferential methods when the sample size is not very large, provided meaningful prior assumptions are made. However, the most critical task of performing Bayesian inference would be the elicitation of the prior distributions.

ACKNOWLEDGMENTS

The author expresses his sincere thanks to the two anonymous reviewers who made many constructive comments that helped to signficantly improve an earlier version of this manuscript.

REFERENCES

[1] Balakrishnan, N. and Mitra, D. (2011). Likelihood inference for log-normal data with left truncation and right censoring with an illustration, *Journal of Statistical Planning and Inference*, **141**, 3536–3553.

[2] Balakrishnan, N. and Mitra, D. (2012). Left truncated and right censored Weibull data and likelihood inference with an illustration, *Computational Statistics and Data Analysis*, **56**, 4011–4025.

[3] Balakrishnan, N. and Mitra, D. (2013). Likelihood inference based on left truncated and right censored data from a gamma distribution, *IEEE Transactions on Reliability*, **62**, 679–688.

[4] Balakrishnan, N. and Mitra, D. (2014a). Some further issues concerning likelihood inference for left truncated and right censored lognormal data, *Communications in Statistics – Simulation and Computation*, **43**, 400–416.

[5] Balakrishnan, N. and Mitra, D. (2014b). EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination (with discussions), *South African Statistical Journal*, **48**, 125–204.

[6] Beirlant, J.; Goegebeur, Y.; Verlaak, R. and Vynckier, P. (1998). Burr regression and portfolio segmentation, *Insurance: Mathematics and Economics*, **23**, 231–250.

[7] Bordes, L. and Chauveau, D. (2014). Comments: EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination, *South African Statistical Journal*, **48**, 197–200.

[8] Burr, I.W. (1942). Cumulative frequency functions, *Annals of Mathematical Statistics*, **13**, 215–232.

[9] Celeux, G.; Chauveau, D. and Diebolt, J. (1996). Stochastic versions of the em algorithm: an experimental study in the mixture case, *Journal of Statistical Computation and Simulation*, **55**, 287–314.

[10] 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, Series B*, **39**, 1–38.

[11] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

[12] Henningsen, A. and Toomet, O. (2011). maxLik: a package for maximum likelihood estimation in R, *Computational Statistics*, **26**, 443–458.

[13] Hong, Y.; Meeker, W.Q. and McCalley, J.D. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data, *The Annals of Applied Statistics*, **3**, 857–879.

[14] Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, Second Edition, Springer-Verlag, New York.

[15] 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.

[16] Mitra, D.; Kundu, D. and Balakrishnan, N. (2021). Likelihood analysis and stochastic EM algorithm for left truncated right censored data and associated model selection from the Lehmann family of life distributions, *Japanese Journal of Statistics and Data Science*, **4**, 1019–1048.

[17] Nielsen, S.F. (2000). The stochastic EM algorithm: Estimation and asymptotic results, *Bernoulli*, **6**, 457–489.

[18] Ng, H.K.T. and Ye, Z. (2014). Comments: EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination, *South African Statistical Journal*, **48**, 177–180.

[19] Silva, G.O.; Edwin, M.M.O.; Cancho, V.G. and Barreto, M.L. (2008). Log-Burr XII regression models with censored data, *Computational Statistics & Data Analysis*, **52**, 3820–3842.

[20] Silva, G.O.; Edwin, M.M.O. and Paula, G.A. (2011). Residuals for log-Burr XII regression models in survival analysis, *Journal of Applied Statistics*, **38**, 1435–1445.

[21] Soliman, A.A. (2005). Estimation of parameters of life from progressively censored data using Burr-XII model, *IEEE Transaction on Reliability*, **54**, 34–42.

[22] Tadikamalla, P.R. (1980). A look at the Burr and related distributions, *International Statistical Review*, **48**, 337–344.

[23] Ye, Z. and Ng, H.K.T. (2014). On analysis of incomplete field failure data, *The Annals of Applied Statistics*, **8**, 1713–1727.

[24] Zimmer, W.J.; Keats, J.B. and Wang, F.K. (1998). The Burr XII distribution in reliability analysis, *Journal of Quality Technology*, **30**, 386–394.