

Artículo de investigación

Burr XII Distribution with Truncated and Censored Data: Maximum Likelihood Estimation Based on Newton-Raphson Method

Distribución de Burr XII con datos truncados y censurados: Estimación de máxima verosimilitud basada en Método de Newton-Raphson

Ramón Giraldo ¹ ✉, Armando Zarruk², Edwin Riveros²

¹Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Departamento de Estadística

²Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Departamento de Matemáticas

Recepción: 2-sep-2022 **Aceptado:** 08-marzo-2024 **Publicado:** 23-jul-2024

Cómo citar: Giraldo, R., Zarruk, A., & Riveros, E. (2024). Distribución de Burr XII con datos truncados y censurados: Estimación de máxima verosimilitud basada en Método de Newton-Raphson. *Ciencia En Desarrollo*, 15(2). <https://doi.org/10.19053/01217488.v15.n2.2024.14858>

Abstract

This work shows a methodology to estimate by maximum likelihood (ML) the parameters of a Burr XII distribution when data are simultaneously left-truncated and right-censored. Given that ML equations do not have a definitive solution under these conditions, an iterative procedure based on the Newton-Raphson method is considered; percentile matching is used to set initial values to the algorithm. Simulation and real data analysis results indicate that the alternative proposed performs well.

Keywords: Burr XII distribution, censored and truncated data, maximum likelihood.

Resumen

Este trabajo muestra una metodología para estimar por máxima verosimilitud (ML) los parámetros de una distribución Burr XII cuando los datos son simultáneamente truncados por la izquierda y censurados por la derecha. Dado que las ecuaciones de ML no tienen solución definitiva en estas condiciones, se considera un procedimiento iterativo basado en el método de Newton-Raphson. La coincidencia de percentiles se utiliza para establecer valores iniciales en el algoritmo. Los resultados basados en simulaciones y análisis de datos reales indican que la alternativa propuesta tiene un buen desempeño.

Palabras Clave: Distribución Burr XII, datos censurados y truncados, máxima verosimilitud.

1 Introduction

The Burr distribution was formally presented in 1942, as part of a family of useful distributions for fitting data [1]. Among the initial list of twelve parametric functions given as solution examples of the differential equation being analysed, the last one, usually called the Burr XII or simply the Burr distribution, is the most popular. Part of the advantages of the three-parameter Burr XII distribution, sometimes also called Singh-Maddala distribution, arises from the flexibility that it provides, including that some other practical and well know distributions such as the Pareto, Weibull, Log-logistic, and Paralogistic can be obtained as a special case of the Burr XII model.

There are different applications of the Burr distribution to model data in a wide range of areas. In actuarial science, it is used to model the size of insurance claims, especially for heavy-tailed distributions [2, 3, 4, 5]. In reliability analysis, this distribution is used to model failure rates of components [6, 7]. In survival analysis, it is used to model lifetime variables [8, 9]. In economics, it is used to model income distribution [10]. It has also been used in finance [11] and engineering [12, 13].

Several authors have used ML to estimate the Burr Type XII distribution parameters under different scenarios of incomplete censored data. Among those, we can mention type II censored data [14], progressively censored data [14], multiple-censored and singly-censored data (Type I censoring or Type II censoring) [15], random censored data [16] and middle-censored data [17]. Similarly to data modifications by censoring, but to a lesser extent, other authors have analyzed the Burr XII distribution under truncated data conditions [18, 19]. Some authors have also studied parameter estimation under data of various lifetime distributions with some censoring and truncation simultaneously [20, 21, 22, 23].

In addition to estimation developments under data modifications, research has also focused on extended models of the three-parameter Burr XII distribution, including its properties and related applications. Some of these extensions can be obtained by adding parameters; included are a four-parameter model called the Weibull Burr XII distribution [24], a five-parameter distribution called the Kumaraswamy Burr XII distribution [25], a six-parameter generalized Burr XII distribution [26] and even a seven-parameter Burr distribution [27]. Additional extended models can also be obtained using other mathematical techniques [28, 29, 30, 31] or definitions of the density function [32].

This paper focuses on estimating the three-parameter Burr XII distribution parameters via MLE when data are simultaneously left-truncated and right-censored. Section 2 describes some properties of the Burr distribution and introduces the definition of left-truncated and right-censored data. Section 3 presents the ML function for the Burr distribution with truncated and censored data. In Section 4, the MLE equations are solved, and the observed Fisher information matrix is obtained. Section 5 describes how to estimate confidence intervals for the parameters. A simulation study is implemented in section 6. Applications to real data are presented in Section 7. The paper ends with a short discussion and suggestions for further research in Section 8.

2 Background

This section describes the Burr XII distribution and its properties and reviews the definitions of truncated and censored data.

2.1 Burr XII Distribution

The parametrization for the Burr XII distribution follows [3] and corresponds to the way it is usually presented in the actuarial literature, where it has gained significant attention in the last two decades. A random variable X is said to have a three-parameter $\Theta = (\alpha, \gamma, \theta)$ Burr XII distribution, or simply a Burr XII distribution,

if its density and distribution functions are given by

$$f(x; \Theta) = \frac{\alpha \gamma \left(\frac{x}{\theta}\right)^\gamma}{x \left[1 + \left(\frac{x}{\theta}\right)^\gamma\right]^{\alpha+1}}, \text{ and} \tag{1}$$

$$F(x; \Theta) = 1 - \left[1 + \left(\frac{x}{\theta}\right)^\gamma\right]^{-\alpha}, \text{ with } x, \theta, \gamma, \text{ and } \alpha > 0, \text{ respectively,} \tag{2}$$

where α and γ are shape parameters and θ is a scale parameter. In order to illustrate its flexibility, some forms of the density and distribution functions are shown in Figure 1.

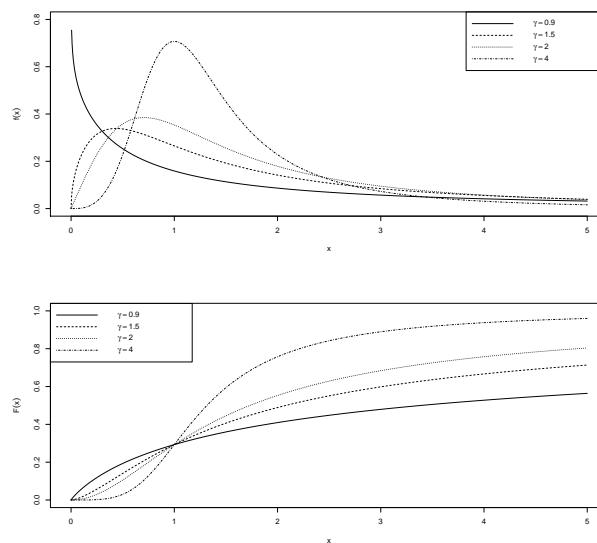


Figure 1: Densities and distributions functions of Burr XII distributions with parameters $\alpha = 0.5, \theta = 1$ and $\gamma = (0.9, 1.5, 2, 4)$, respectively.

Given a random sample x_1, \dots, x_n , of a Burr XII distribution, the log-likelihood function can be written as

$$l(\Theta) = n \ln \alpha + n \ln \gamma + (\gamma - 1) \sum_{i=1}^n \ln x_i - n \gamma \ln \theta - (\alpha + 1) \sum_{i=1}^n \ln [1 + (x_i/\theta)^\gamma] \tag{3}$$

Taking the derivative of $l(\Theta)$ in (3) with respect to the parameters α, θ and γ and setting them equal to zero we have

$$\begin{aligned} \frac{\partial}{\partial \alpha} l(\Theta) &= \frac{n}{\alpha} - \sum_{i=1}^n \ln [1 + (x_i/\theta)^\gamma] = 0 \\ \hat{\alpha} &= \frac{n}{\sum_{i=1}^n \ln [1 + (x_i/\theta)^\gamma]}, \\ \frac{\partial}{\partial \theta} l(\Theta) &= -\frac{n \gamma}{\theta} + (\alpha - 1) \sum_{i=1}^n \frac{\gamma x_i^\gamma \theta^{-(\gamma+1)}}{1 + (x_i/\theta)^\gamma} = 0, \\ \frac{\partial}{\partial \gamma} l(\Theta) &= \frac{n}{\gamma} + \sum_{i=1}^n \ln x_i - n \ln \theta - (\alpha + 1) \sum_{i=1}^n \frac{(x_i/\theta)^\gamma \ln(x_i/\theta)}{1 + (x_i/\theta)^\gamma} = 0. \end{aligned}$$

There is not an explicit solution for θ and γ above, and therefore some numerical method such as Newton-Raphson needs to be used.

2.2 Left Truncated and Right Censored Data

In practice situations it is usual that complete information is not available and the researcher is only disposed incomplete data. In survival and reliability studies the interest is on the non-negative random variable representing the time until the occurrence of an event, usually the time of death or failure of a component. A common example of incomplete information is when the failure time data includes information from individuals who did not fail in the observation period or were only observed up until certain point of time; this information is said to be right censored and it is incomplete since the survival time goes beyond the observed point, but the exact time of failure is not known.

A second source for incomplete information arises when only individuals who experience some event are observed; that is, individuals are followed in the study only if they fulfill a required condition which screens them in order to be included in the analysis. This produces a biased sample, and those who are observed are said to be left truncated. It is usual to have left truncated data when individuals have a delayed entry or when they do not survive until the beginning of the observation period.

In insurance, the main variables of interest are the number of claims and the size or value of those claims. Left truncation occurs if the insurer pays only when the value of the claim is more than an agreed amount d , called the deductible. Right censoring arises with a policy limit u , where even if the loss exceeds the pre-established amount u , benefits are paid only up to u .

In survival analysis, as well as in insurance, it is very common to have data that is left truncated and right censored (LTRC) simultaneously. An observation is *right censored* (or censored from above) at u if when it is at or above u it is recorded as being equal to u , but when it is below u it is recorded at its observed value. On the other hand, an observation is *left truncated* (or truncated from below) at d if when it is at or below d it is not recorded, but when it is above d it is recorded at its observed value [3].

3 Log-Likelihood Function for the Burr XII Distribution with LTRC Data

It is well known that under complete individual data, the contribution to the likelihood function from an individual with exact observed lifetime of x_i is the corresponding density function, notated by $f(x_i; \Theta)$ where $\Theta = (\theta_1, \theta_2, \dots, \theta_n)$ is the parameter vector of the distribution. However, when dealing with incomplete data, the construction of the likelihood function needs to be carefully considered in order to account properly for the information given in each sampled observation. If the observation comes from a population with left truncation at d , then it should be noted that it was observed given that its value exceeded d , and therefore the (conditional) contribution to the likelihood function from a left truncated observation is $\frac{f(x_i; \Theta)}{1 - F(d; \Theta)}$.

For a right censored observation at u , all we know is that its value exceeded u , thus the contribution to the likelihood function from a right censored observation is simply $(1 - F(u; \Theta))$, that corresponds to the probability of exceeding u . Consider a sample of $n + m$ data left truncated at d , where m of them are also right censored at u . The likelihood function, for a data set with these characteristics, is given

by

$$L(\Theta) = \prod_{i=1}^{n+m} f(x_i, \Theta) = \prod_{i=1}^n f_1(x_i; \Theta) \times \prod_{i=n+1}^{n+m} f_2(x_i; \Theta) \\ = \prod_{i=1}^n \frac{f(x_i; \Theta)}{1 - F(d; \Theta)} \times \prod_{i=n+1}^{n+m} \frac{1 - F(u; \Theta)}{1 - F(d; \Theta)} = \frac{[1 - F(u; \Theta)]^m}{[1 - F(d; \Theta)]^{n+m}} \prod_{i=1}^n f(x_i; \Theta), \tag{4}$$

with

$$f_1(x; \Theta) = P(X = x | X > d; \Theta) = \frac{P(X = x \cap X > d; \Theta)}{P(X > d; \Theta)} = \frac{f(x, \Theta)}{1 - F(d; \Theta)}, \tag{5}$$

and

$$f_2(x; \Theta) = P(X > u | X > d; \Theta) = \frac{P(X > u \cap X > d; \Theta)}{P(X > d; \Theta)} = \frac{1 - F(u; \Theta)}{1 - F(d; \Theta)} \tag{6}$$

Taking logarithms in (4) we obtain the log-likelihood function as follows

$$l(\Theta) = \ln L(\Theta) = m \ln [1 - F(u; \Theta)] - (m + n) \ln [1 - F(d; \Theta)] + \sum_{i=1}^n \ln f(x_i; \Theta) \tag{7}$$

From (1)

$$\ln f(x; \Theta) = \ln \alpha + \ln \gamma + \gamma \ln x - \gamma \ln \theta - \ln x - (\alpha + 1) \ln \left(1 + \frac{x}{\theta}\right)^\gamma \\ = \ln \alpha + \ln \gamma + (\gamma - 1) \ln x - \gamma \ln \theta - (\alpha + 1) \ln \left(1 + \frac{x}{\theta}\right)^\gamma. \tag{8}$$

From (2)

$$1 - F(u; \Theta) = \left[1 + \left(\frac{u}{\theta}\right)^\gamma\right]^{-\alpha}, \\ \ln [1 - F(u; \Theta)] = -\alpha \ln \left[1 + \left(\frac{u}{\theta}\right)^\gamma\right], \text{ and} \tag{9}$$

$$\ln [1 - F(d; \Theta)] = -\alpha \ln \left[1 + \left(\frac{d}{\theta}\right)^\gamma\right]. \tag{10}$$

Replacing (8), (9) and (10) in (7) we obtain the log-likelihood function for the Burr XII distribution with censored and truncated data. This is given by

$$l(\Theta) = -m\alpha \ln [1 + (u/\theta)^\gamma] + (n + m)\alpha \ln [1 + (d/\theta)^\gamma] + n \ln \alpha + n \ln \gamma \tag{11}$$

$$+ (\gamma - 1) \sum_{i=1}^n \ln x_i - n\gamma \ln \theta - (\alpha + 1) \sum_{i=1}^n \ln [1 + (x_i/\theta)^\gamma] \\ = -\alpha S_2 + \alpha S_3 + n \ln \alpha + n \ln \gamma + (\gamma - 1) \sum_{i=1}^n \ln x_i - n\gamma \ln \theta - (\alpha + 1) S_1,$$

with

$$S_1 = \sum_{i=1}^n \ln [1 + (x_i/\theta)^\gamma], \\ S_2 = m \ln [1 + (u/\theta)^\gamma], \text{ and} \\ S_3 = (n + m) \ln [1 + (d/\theta)^\gamma].$$

Notation S_1, S_2, S_3 is given in [15].

4 Parameter Estimation of the Burr XII Distribution with LTRC Data

Taking the derivative in Equation (11) with respect to α , and setting it equal to zero we have

$$\hat{\alpha} = \frac{n}{S_1 + S_2 - S_3}.$$

Defining

$$S_{1\theta} = \frac{\partial S_1}{\partial \theta} = -\sum_{i=1}^n \frac{\gamma x_i^\gamma \theta^{-(\gamma+1)}}{1 + (x_i/\theta)^\gamma}, \quad S_{2\theta} = \frac{\partial S_2}{\partial \theta} = -\frac{m\gamma u^\gamma \theta^{-(\gamma+1)}}{1 + (u/\theta)^\gamma},$$

$$S_{3\theta} = \frac{\partial S_3}{\partial \theta} = -\frac{(n+m)\gamma d^\gamma \theta^{-(\gamma+1)}}{1 + (d/\theta)^\gamma}, \quad S_{1\gamma} = \frac{\partial S_1}{\partial \gamma} = \sum_{i=1}^n \frac{(x_i/\theta)^\gamma \ln(x_i/\theta)}{1 + (x_i/\theta)^\gamma}$$

$$S_{2\gamma} = \frac{\partial S_2}{\partial \gamma} = \frac{m(u/\theta)^\gamma \ln(u/\theta)}{1 + (u/\theta)^\gamma}, \quad S_{3\gamma} = \frac{\partial S_3}{\partial \gamma} = \frac{(n+m)(d/\theta)^\gamma \ln(d/\theta)}{1 + (d/\theta)^\gamma},$$

the likelihood equations for θ and γ are

$$g_1(\theta, \gamma) = \frac{\partial}{\partial \theta} l(\Theta) = -\hat{\alpha} S_{2\theta} + \hat{\alpha} S_{3\theta} - \frac{n\gamma}{\theta} - (\hat{\alpha} + 1) S_{1\theta} = 0$$

$$g_2(\theta, \gamma) = \frac{\partial}{\partial \gamma} l(\Theta) = -\hat{\alpha} S_{2\gamma} + \hat{\alpha} S_{3\gamma} + \frac{n}{\gamma} + \sum_{i=1}^n \ln x_i - n \ln \theta - (\hat{\alpha} + 1) S_{1\gamma} = 0.$$

The equations for θ and γ can be written in matrix notation as

$$G(\theta, \gamma) = \begin{bmatrix} g_1(\theta, \gamma) \\ g_2(\theta, \gamma) \end{bmatrix} = 0.$$

The MLE of θ and γ are the values $\hat{\theta}$ and $\hat{\gamma}$ that simultaneously solve this non-linear equation system. An option in this scenario is to use the iterative Newton-Raphson method [33]. According to this approach, the iteration $(k + 1)$ to solve the system is

$$\begin{pmatrix} \hat{\theta}_{k+1} \\ \hat{\gamma}_{k+1} \end{pmatrix} = \begin{pmatrix} \hat{\theta}_k \\ \hat{\gamma}_k \end{pmatrix} - G'(\hat{\theta}_k, \hat{\gamma}_k)^{-1} G(\hat{\theta}_k, \hat{\gamma}_k), \quad (12)$$

$$\text{with } G'(\theta_k, \gamma_k) = \begin{pmatrix} \frac{\partial g_1}{\partial \theta} & \frac{\partial g_1}{\partial \gamma} \\ \frac{\partial g_2}{\partial \theta} & \frac{\partial g_2}{\partial \gamma} \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}.$$

Equation (12) can be simplified as

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \frac{g_1(\hat{\theta}_k, \hat{\gamma}_k) g_{22}(\hat{\theta}_k, \hat{\gamma}_k) - g_2(\hat{\theta}_k, \hat{\gamma}_k) g_{12}(\hat{\theta}_k, \hat{\gamma}_k)}{g_{11}(\hat{\theta}_k, \hat{\gamma}_k) g_{22}(\hat{\theta}_k, \hat{\gamma}_k) - g_{12}(\hat{\theta}_k, \hat{\gamma}_k) g_{21}(\hat{\theta}_k, \hat{\gamma}_k)},$$

$$\hat{\gamma}_{k+1} = \hat{\gamma}_k - \frac{g_2(\hat{\theta}_k, \hat{\gamma}_k) g_{11}(\hat{\theta}_k, \hat{\gamma}_k) - g_1(\hat{\theta}_k, \hat{\gamma}_k) g_{21}(\hat{\theta}_k, \hat{\gamma}_k)}{g_{11}(\hat{\theta}_k, \hat{\gamma}_k) g_{22}(\hat{\theta}_k, \hat{\gamma}_k) - g_{12}(\hat{\theta}_k, \hat{\gamma}_k) g_{21}(\hat{\theta}_k, \hat{\gamma}_k)},$$

where $\hat{\alpha}_0$, $\hat{\theta}_0$, and $\hat{\gamma}_0$ are values for the starting points. A good choice of these is important to improve convergence. Generally the method of moments (MM) or percentile matching method (PM) are used to set the initial values. From a computational point of view, it is preferable the second option [3] since the MM also requires to solve a system of equations using numerical methods. The PM consist of replacing theoretical percentiles with empirical percentiles of the random sample. In the case of a Burr XII distribution the quantile function is defined as

$$F^{-1}(p) = \theta[(1-p)^{-1/\alpha} - 1]^{1/\gamma}, \quad 0 < p < 1.$$

For example, let us assume that \hat{p}_1 , \hat{p}_2 and \hat{p}_3 are estimations of percentiles $p_1 = 25\text{th}$, $p_2 = 50\text{th}$ and $p_3 = 75\text{th}$, respectively. It can

be shown that $\hat{\alpha}_0$, $\hat{\gamma}_0$ and $\hat{\theta}_0$ are obtained by solving the following equations

$$\frac{\ln \left[\frac{(1-p_1)^{-1/\alpha-1}}{(1-p_2)^{-1/\alpha-1}} \right]}{\ln \left[\frac{(1-p_2)^{-1/\alpha-1}}{(1-p_3)^{-1/\alpha-1}} \right]} - \frac{\ln(\hat{p}_1/\hat{p}_2)}{\ln(\hat{p}_2/\hat{p}_3)} = 0$$

$$\hat{\gamma}_0 = \frac{\ln \left[\frac{(1-p_1)^{-1/\hat{\alpha}_0-1}}{(1-p_2)^{-1/\hat{\alpha}_0-1}} \right]}{\ln(\hat{p}_1/\hat{p}_2)}, \text{ and}$$

$$\hat{\theta}_0 = \frac{\hat{p}_1}{[(1-p_1)^{-1/\hat{\alpha}_0} - 1]^{1/\hat{\gamma}_0}}.$$

5 Covariance Matrix and Confidence Intervals

ML estimators are asymptotically normal estimators with variance and covariance given by the inverse of the Fisher information matrix $I(\Theta)$. When certain regularity conditions are met $I(\Theta)$ can be approximated by the observed Fisher information matrix [34], notated $I(\hat{\Theta})$ and given by

$$I(\hat{\Theta}) = - \begin{pmatrix} \frac{\partial^2 l(\Theta)}{\partial \alpha^2} & \frac{\partial^2 l(\Theta)}{\partial \alpha \partial \theta} & \frac{\partial^2 l(\Theta)}{\partial \alpha \partial \gamma} \\ \frac{\partial^2 l(\Theta)}{\partial \theta \partial \alpha} & \frac{\partial^2 l(\Theta)}{\partial \theta^2} & \frac{\partial^2 l(\Theta)}{\partial \theta \partial \gamma} \\ \frac{\partial^2 l(\Theta)}{\partial \gamma \partial \alpha} & \frac{\partial^2 l(\Theta)}{\partial \gamma \partial \theta} & \frac{\partial^2 l(\Theta)}{\partial \gamma^2} \end{pmatrix}_{\Theta=\hat{\Theta}}$$

From Equations (5) and (6) it can be seen that for a fixed truncation point d and a censoring point u the contribution of LTRC to the log-likelihood function preserve the functional form of the log-likelihood function of a Burr XII distribution with complete data; therefore, the existence of the Fisher information matrix for LTRC data is guaranteed. To verify the later we refer the reader to papers such as [35] and [36]. The approximate covariance matrix can be estimated as

$$I^{-1}(\hat{\Theta}) = \begin{pmatrix} \widehat{Var}(\hat{\alpha}) & \widehat{Cov}(\hat{\alpha}, \hat{\theta}) & \widehat{Cov}(\hat{\alpha}, \hat{\gamma}) \\ \widehat{Cov}(\hat{\theta}, \hat{\alpha}) & \widehat{Var}(\hat{\theta}) & \widehat{Cov}(\hat{\theta}, \hat{\gamma}) \\ \widehat{Cov}(\hat{\gamma}, \hat{\alpha}) & \widehat{Cov}(\hat{\gamma}, \hat{\theta}) & \widehat{Var}(\hat{\gamma}) \end{pmatrix},$$

where

$$\widehat{Var}(\hat{\alpha}) = (bc - f^2)/(cd^2 - 2edf + af^2 + be^2 - abc),$$

$$\widehat{Var}(\hat{\theta}) = (ac - e^2)/(cd^2 - 2edf + af^2 + be^2 - abc),$$

$$\widehat{Var}(\hat{\gamma}) = (ab - d^2)/(cd^2 - 2edf + af^2 + be^2 - abc),$$

with (see appendix)

$$a = -\frac{\partial^2 l(\Theta)}{\partial \alpha^2}, \quad b = -\frac{\partial^2 l(\Theta)}{\partial \theta^2}, \quad c = -\frac{\partial^2 l(\Theta)}{\partial \gamma^2},$$

$$d = -\frac{\partial^2 l(\Theta)}{\partial \alpha \partial \theta}, \quad e = -\frac{\partial^2 l(\Theta)}{\partial \alpha \partial \gamma}, \quad f = -\frac{\partial^2 l(\Theta)}{\partial \theta \partial \gamma},$$

all of these evaluated at the estimated parameters $\hat{\alpha}$, $\hat{\theta}$ and $\hat{\gamma}$. Based on asymptotic normality of ML estimators, confidence intervals of $100(1 - \rho)\%$ for α, θ, γ are defined as

$$\hat{\alpha} \pm z_{1-\frac{\rho}{2}} \sqrt{\widehat{Var}(\hat{\alpha})},$$

$$\hat{\theta} \pm z_{1-\frac{\rho}{2}} \sqrt{\widehat{Var}(\hat{\theta})}, \text{ and}$$

$$\hat{\gamma} \pm z_{1-\frac{\rho}{2}} \sqrt{\widehat{Var}(\hat{\gamma})},$$

where $z_{1-\frac{\rho}{2}}$ is a quantile of a standard normal distribution.

6 Simulation Study

Initially, in Section 6.1, we graphically establish the goodness of fit of the proposed methodology. Subsequently, in Sections 6.2 and 6.3, a more intensive computational study is carried out to study empirically the convergence of the estimators. The R code used in the simulations is given in Section 9 and is also available at the web page <https://sites.google.com/a/unal.edu.co/ramon-giraldo-webpage/r-code> (see RCodeBurrXII.R).

6.1 Goodness of fit test for the Burr XII ($\alpha = 2, \gamma = 3, \theta = 4$) distribution

Using the R library actuar [37] 1000 data from a Burr XII distribution ($\alpha = 2, \gamma = 3, \theta = 4$) are simulated (histogram in Figure 2). A large data set is used to have a good approximation to the population distribution. The data generated are posteriorly left truncated at $d = 1.51$ and right censored at $u = 5.17$ (dashed vertical lines in Figure 2) obtaining 796 observations in the interval $(d, u]$. Using these partition of the data and the iterative procedure described in Section 4, the MLE are obtained. The 25th, 50th, and 75th percentiles of the

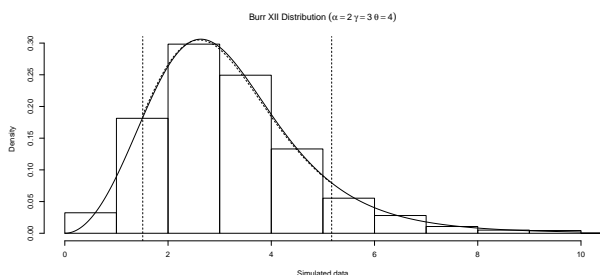


Figure 2: Histogram density estimation of 1000 simulated data from a Burr XII distribution with $\alpha = 2, \gamma = 3$, and $\theta = 4$. Data are left truncated and right censored at 1.51 and 5.17, respectively (vertical dashed lines), The black line corresponds to the Burr XII density, and the dashed curve (between 1.51 and 5.17) to the estimation of a Burr XII left truncated and right censored.

generated data are 2.29, 2.98, and 3.76, respectively. The starting points are calculated using these values and the PM method (see Section 4). These are $\hat{\alpha}_0=2.20, \hat{\gamma}_0=3.71$, and $\hat{\theta}_0=3.89$, respectively. The final estimates are $\hat{\alpha}=1.93, \hat{\gamma}=2.98$, and $\hat{\theta}=3.93$. The estimated density function (dashed curve in the interval $[1.51, 5.17]$ in Figure 2) is very close to the reference density (continuous line in Figure 2) and consequently suggests that the approach implemented is satisfactory.

6.2 Distribution of the estimators

10000 simulations size $n = 1000$ of a Burr XII distribution ($\alpha = 2, \gamma = 3, \theta = 4$) were run. Each random sample simulated is trimmed at $d=1.51$ and $u=5.17$. At each iteration the Newton-Raphson algorithm is carried out (based on 100 iterations) and the parameters are estimated. If the starting points are not properly defined the method fails. Approximately 70% of the runs allow obtaining the estimates of the parameters. Kernel density estimations obtained for $\hat{\alpha}, \hat{\gamma}$ and $\hat{\theta}$ are shown in Figures 3, 4, and 5. These plots suggest that is reasonable assuming asymptotic normality for the distributions of $\hat{\gamma}$ and $\hat{\theta}$. The distribution of $\hat{\alpha}$ is slightly biased. A large sample size n could be required in this case.

The means of the corresponding estimations (vertical dashed lines) are 1.98, 2.98, and 3.93. These values are very similar to the parameters. The curves in Figures 3 to 5, indicate that in general the methodology used presents a good performance.

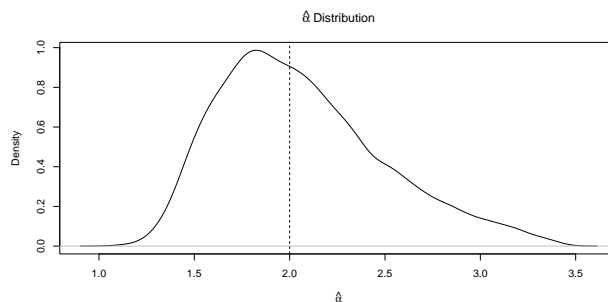


Figure 3: Kernel density estimation of the $\hat{\alpha}$ distribution. Values obtained from 10000 size $n = 1000$ Monte Carlo simulations of $X \sim$ Burr XII ($\alpha = 2, \gamma = 3$ and $\theta = 4$). Dashed vertical line allows identify the parameter α .

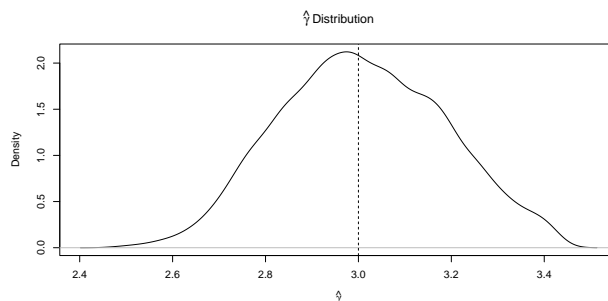


Figure 4: Kernel density estimation of the $\hat{\gamma}$ distribution. Values obtained from 10000 Monte Carlo simulations of $X \sim$ Burr XII ($\alpha = 2, \gamma = 3$ and $\theta = 4$). Dashed line allows identify the parameter γ .

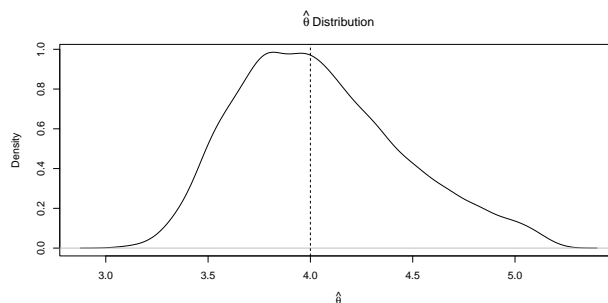


Figure 5: Kernel density estimation of the $\hat{\theta}$ distribution. Values obtained from 10000 size $n = 1000$ Monte Carlo simulations from $X \sim$ Burr XII ($\alpha = 2, \gamma = 3$ and $\theta = 4$). Dashed line allows identify the parameter θ .

6.3 Comparison of results varying n and θ

A computationally intensive simulation procedure was conducted. 10000 simulations size $n = 50, 100, 200, 500$, and 1000 of Burr XII distributions were generated. At each case following the methodology proposed were estimated the parameters (α, θ, γ) . Three levels of censor and truncation (5%, 10%, 20%) are considered. Summary statistics (mean, median, coefficient of variation, relative bias, root mean square error) were calculated based on the estimations. As in Section 6.2 the final number of estimations depends on the convergence rate of the method which is around 70%. The results are shown in Tables 1 and 2. Various points can be highlighted from these Tables. In general, the greater the number of censored and truncated

observations the greater the relative bias and the root mean square error. There are differences between means and medians which indicate the presence of some atypical estimated values. As expected, according with the root mean square errors (RMS), the more data for model fitting the better the estimations, i.e., the estimators are asymptotically unbiased.

7 Applications

We now consider two examples with real data. Although both data sets have been used to fit a Burr XII distribution, they are modified to adjust them to left truncation and right censoring conditions. An R [38] code is used to compute ML estimations with LTRC data. The code allows solving the adjusted ML equations (given in section 3) using the Newton-Raphson algorithm. The *starting points* are obtained from the PM with the 25th, 50th, and 75th percentiles. 95% confidence intervals for the parameters are obtained. Plots comparing empirical and fitted distributions are presented. A KS test, adjusted to incomplete data, is used to assess the goodness of fit.

7.1 Breast Cancer Data

The information corresponds to breast cancer survival data of 935 patients [39]. The study was carried out between 2009 and 2013. 705 people were still alive at the end of 2013 (data are censored at $u = 1826$). The frequency histogram is shown in Figure 6. A log transformation of frequencies is used to reduce the scale. We observe a high frequency on the right of the distribution which is due to the censored information. The data set is artificially trimmed in order to left truncate it by considering a truncation point at $d = 30$ days. Thus, a new data set is obtained by eliminating the 9 points that are below 30 from the adjusted set, resulting in 221 patients with exact observed times and 705 with censored information. Here we take into account censored and truncated data to estimate the parameters.

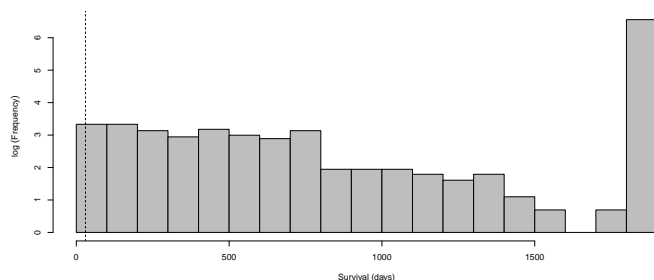


Figure 6: Frequency histogram of breast cancer survival data. A log transformation of frequencies is used to reduce the scale. High frequency at the end of the distribution is due to censored data. Data are left truncated at $d = 30$ (dashed line) and right censored at $u = 1826$ (last interval).

From available information, initial estimates ($\hat{\alpha}_0 = 0.26$, $\hat{\gamma}_0 = 2.84$, and $\hat{\theta}_0 = 201.51$) of the parameters were obtained by using the PM. The final estimates after using the methodology proposed are $\hat{\alpha} = 0.08$, $\hat{\gamma} = 1.57$, $\hat{\theta} = 182.96$.

A comparison between the empirical and estimated distribution functions is shown in Figure 7. The estimated distribution is obtained from a simulation size 10000 of a Burr XII ($\hat{\alpha} = 0.077$, $\hat{\gamma} = 1.566$, $\hat{\theta} = 182.96$) model. We can observe graphically that the model fitted seems appropriate. The p-value of a KS goodness of fit test conducted to compare the distributions in Figure 7 is 0.83. According to this value there is not statistical evidence to reject the hypothesis that breast cancer survival data can be fitted with the Burr XII distribution estimated. 95% confidence intervals for α , γ , and θ were obtained. These are respectively [0.072, 0.079], [1.413, 1.719], and [140.301,

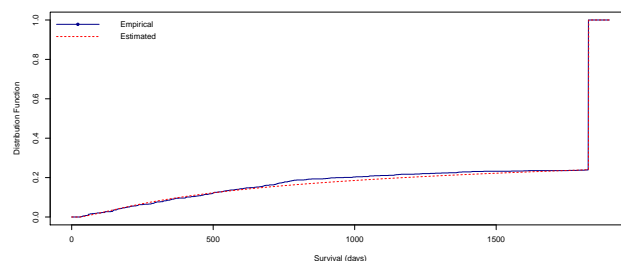


Figure 7: Distribution functions (empirical and estimated) for survival data (breast cancer data)

225.629].

7.2 Bladder Cancer Data

The data correspond to remission times from patients with bladder cancer as described in [40]. A subset of this data was used in [41] to fit a log-logistic distribution (a particular case of a Burr XII distribution). The frequency histogram is shown in Figure 8. This data set is truncated from below at $d = 2$ (dashed line in Figure 8) and censored from above at $u = 30$. The ML estimates for the parameters of the

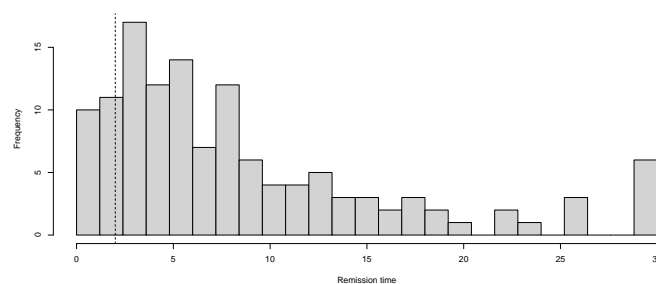


Figure 8: Frequency histogram of bladder cancer remission times. Data are left truncated at $d = 2$ (dashed line) and right censored at $u = 30$ (last interval).

Burr XII model are $\hat{\alpha} = 2.33$, $\hat{\gamma} = 1.23$ and $\hat{\theta} = 12.50$. A comparison between the empirical distribution function (generated from the data recorded) and the theoretical distribution function (obtained from 10000 simulated data of a Burr XII model $(\hat{\alpha}, \hat{\gamma}, \hat{\theta})$) is presented in Figure 9. This graph validates from a descriptive point of view the methodology used.

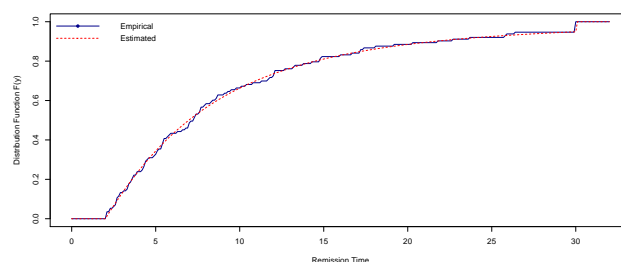


Figure 9: Distribution functions (empirical and estimated) for remission times (bladder cancer data)

Table 1: The parameters are $\alpha = 2$, $\theta = 4$, and $\gamma = 3$. n : sample size. %: Percentage of censoring and truncation. For example 5% means 2.5% of censoring and 2.5% of truncation. $(\bar{\alpha}, \bar{\theta}, \bar{\gamma})$ and $(\tilde{\alpha}, \tilde{\theta}, \tilde{\gamma})$ are the means and medians of the estimations, respectively. CV: Coefficient of variation of the estimations. RB: Relative bias %. RMS: Root mean square error.

n	%	$\bar{\alpha}$	$\tilde{\alpha}$	CV	RB	RMS	$\bar{\theta}$	$\tilde{\theta}$	CV	RB	RMS	$\bar{\gamma}$	$\tilde{\gamma}$	CV	RB	RMS
50	5	1.9	1.3	1.0	-5	1.9	3.7	3.4	0.4	9	1.5	4.5	3.6	1.1	49	5.3
	10	1.7	1.2	1.0	-13	1.7	3.5	3.2	0.4	12	1.5	4.7	3.8	1.0	57	4.9
	20	1.8	1.0	1.4	-9	2.5	3.4	3.1	0.5	14	1.8	6.2	4.0	1.3	105	8.0
100	5	2.1	1.7	0.7	5	1.4	4.0	3.7	0.3	17	1.1	3.1	3.1	0.2	10	0.6
	10	2.3	1.6	0.9	13	2.1	4.0	3.6	0.4	16	1.5	3.7	3.2	0.2	14	0.8
	20	2.5	1.4	1.3	21	3.3	4.1	3.4	0.5	-1	1.2	6.1	3.6	0.4	25	1.6
200	5	2.8	1.9	0.5	8	1.1	4.1	3.9	0.2	12	0.9	3.1	3.0	0.1	4	0.4
	10	3.0	1.8	0.7	14	1.6	4.1	3.8	0.3	12	1.2	3.3	3.1	0.2	6	0.5
	20	2.7	1.7	1.2	37	3.2	4.3	3.7	0.4	2	1.9	4.7	3.3	0.2	11	0.8
500	5	2.1	2.0	0.3	6	0.7	4.1	4.0	0.1	3	0.6	3.0	3.0	0.1	1	0.3
	10	2.2	2.0	0.4	9	0.9	4.1	4.0	0.2	6	0.8	3.1	3.0	0.1	1	0.3
	20	2.7	1.9	0.9	33	2.4	4.4	3.9	0.4	4	1.6	3.3	3.0	0.2	3	0.5
1000	5	2.1	2.0	0.2	3	0.4	4.1	4.0	0.1	1	0.4	3.0	3.0	0.1	0	0.2
	10	2.1	2.0	0.3	5	0.6	4.1	4.0	0.1	2	0.5	3.0	3.0	0.1	0	0.2
	20	2.2	2.0	0.5	16	1.2	4.2	4.0	0.2	4	1.0	3.3	3.0	0.1	1	0.4

Table 2: The parameters are $(\alpha = 2, \theta = 4000, \gamma = 3)$. n : sample size. %: Percentage of censoring and truncation. For example % = 5% means 2.5% of censoring and 2.5% of truncation. $(\bar{\alpha}, \bar{\theta}, \bar{\gamma})$ and $(\tilde{\alpha}, \tilde{\theta}, \tilde{\gamma})$ are the means and medians of the estimations, respectively. CV: Coefficient of variation of the estimations. RB: Relative bias %. RMS: Root mean square error. For simplicity in the presentation RMS values corresponding to the estimation of θ are divided by 1000 and rounded to one decimal.

n	%	$\bar{\alpha}$	$\tilde{\alpha}$	CV	RB	RMS	$\bar{\theta}$	$\tilde{\theta}$	CV	RB	RMS	$\bar{\gamma}$	$\tilde{\gamma}$	CV	RB	RMS
50	5	2.2	1.7	0.8	-25	2.0	3046	2439	0.7	-24	2.4	1.9	1.8	0.2	25	0.6
	10	2.2	1.5	1.0	-28	2.3	2911	2162	0.8	-27	2.6	2.0	1.9	0.3	34	0.7
	20	2.1	1.2	1.2	-30	2.7	2000	1934	0.9	-30	2.8	2.2	2.0	0.4	49	1.2
100	5	2.9	2.2	0.8	-3	2.3	3808	2956	0.7	5	2.7	1.7	1.7	0.2	12	0.3
	10	2.9	1.9	1.0	-5	2.8	3796	2678	0.9	6	3.3	1.8	1.7	0.2	17	0.4
	20	3.0	1.6	1.3	-1	3.8	3992	2321	1.1	3	4.3	1.9	1.8	0.3	24	0.6
200	5	3.1	2.5	0.6	4	2.0	4119	3390	0.6	3	2.3	1.6	1.6	0.1	6	0.2
	10	3.2	2.3	0.8	5	2.6	4155	3168	0.7	4	3.0	1.6	1.6	0.1	8	0.3
	20	3.6	2.1	1.2	19	4.3	4626	2855	1.1	16	4.9	1.7	1.6	0.2	13	0.4
500	5	3.2	2.9	0.4	7	1.2	4269	3819	0.4	7	1.7	1.5	1.5	0.1	1	0.1
	10	3.2	2.8	0.6	12	4.0	4440	3723	0.5	11	2.4	1.5	1.5	0.1	2	0.2
	20	4.0	2.6	1.1	34	4.4	5173	3505	0.9	29	5.0	1.6	1.5	0.1	4	0.2
1000	5	3.2	3.0	0.3	6	0.9	4202	3973	0.3	5	1.1	1.5	1.5	0.1	0	0.1
	10	3.3	2.9	0.4	9	1.3	4340	3913	0.4	8	1.6	1.5	1.5	0.1	0	0.1
	20	3.9	2.9	0.8	29	3.2	5052	3840	0.7	26	3.7	1.5	1.5	0.1	1	0.2

A KS test was conducted to formally compare these cumulative distributions. The p-value obtained (0.997) corroborate the graphical results described. 95% confidence intervals for α, γ , and θ are respectively, [1.95, 2.72], [0.90, 1.57], [8.86, 16.15].

8 Conclusion

The Burr XII distribution is useful to model data in several areas, including applications in survival analysis, actuarial science, and reliability theory. Its flexibility and closed form, among other statistical properties, are some advantages when it is compared with other distributions. In many real situations it can be required to work with left truncated and right censored data. In this work we propose a methodology to estimate the parameters of the Burr XII distribution in this context. We use numerical methods to estimate the parameters by ML. We also give an approach to estimate the variances and covariances matrix and therefore to calculate confidence intervals. In order to illustrate the methodology, a simulation study and two real data analysis are carried out. In both scenarios, the results indicate that the methodology proposed has a good performance. This work can be extended to include other types of data modifications that also arise in practice; in addition it can be applied to other family of distributions where LTRC data are naturally found.

9 Appendix. R code

```
# Data simulation

library(actuar)
nm=3000
alpha=2
thet=4
gamm=3
d=qburr(0.10, shape1=alpha, shape2=gamm, scale=thet)
u=qburr(0.90, shape1=alpha, shape2=gamm, scale=thet)
x1=sort(rburr(nm, shape1=alpha, shape2=gamm, scale=thet))

# Limit

x2=rep(u, nm); x3=rep(1, nm)
for(i in 1:nm){ if (x1[i]<u) {(x2[i]=x1[i])&(x3[i]=0)} }
m=sum(x3); mp=m; x4=rep(0, nm-m)
for(i in 1:(nm-m)){ if (x1[i]<u) {x4[i]=x1[i]} }

# Deductible

x4=sort(x4, decreasing=T); l=length(x4); x5=rep(0, l)
for(i in 1:l){ if (x4[i]>d) {x5[i]=1} }
o=sum(x5); x=rep(0, o)
for(i in 1:o){ if (x4[i]>d) {x[i]=x4[i]} }

# Starting points

q1=qburr(0.25, shape1=2, shape2=3, scale=4)
q2=qburr(0.50, shape1=2, shape2=3, scale=4)
q3=qburr(0.75, shape1=2, shape2=3, scale=4)
p1=quantile(x, 0.25, names=FALSE)
p2=quantile(x, 0.50, names=FALSE)
p3=quantile(x, 0.75, names=FALSE)
q1; q2; q3; p1; p2; p3
a0=2.2015
g0=log( (.75^(-1/a0)-1)/(.5^(-1/a0)-1) )/log(p1/p2)
t0=p1*(.75^(-1/a0)-1)^(-1/g0)

# Maximum likelihood of a Triparametric BURR XII
distribution

n=length(x)
itera=50
theta=rep(t0, itera)
```

```
gamma=rep(g0, itera)
for(i in 1:(itera-1))
{
  t=theta[i]; g=gamma[i]
  s1=sum( log(1+(x/t)^g) )
  s2=m*log(1+(u/t)^g)
  s3=(n+m)*log(1+(d/t)^g)
  s1t=-sum( g*(x^g)*(t^(-g-1))/(1+(x/t)^g) )
  s2t=-m*g*(u^g)*(t^(-g-1))/(1+(u/t)^g)
  s3t=-((n+m)*g*(d^g)*(t^(-g-1)))/(1+(d/t)^g)
  s1g=sum( (x/t)^g*log(x/t)/(1+(x/t)^g) )
  s2g=m*(u/t)^g*log(u/t)/(1+(u/t)^g)
  s3g=(n+m)*(d/t)^g*log(d/t)/(1+(d/t)^g)
  s1tt=sum( (g*(g+1)*x^g*t^(-g-2)*(1+(x/t)^g)
    -(g*x^g*t^(-g-1))^2)/(1+(x/t)^g)^2 )
  s2tt=m*( g*(g+1)*u^g*t^(-g-2)*(1+(u/t)^g)
    -(g*u^g*t^(-g-1))^2 )/(1+(u/t)^g)^2
  s3tt=(n+m)*( g*(g+1)*d^g*t^(-g-2)*(1+(d/t)^g)
    -(g*d^g*t^(-g-1))^2 )/(1+(d/t)^g)^2
  s1tg=-((1/t)*sum( ((x/t)^g+g*(x/t)^g*log(x/t))*(1+(x/t)^g)
    -g*(x/t)^(2*g)*log(x/t))/(1+(x/t)^g)^2 )
  s2tg=-((m/t)*((u/t)^g+g*(u/t)^g*log(u/t))*(1+(u/t)^g)
    -g*(u/t)^(2*g)*log(u/t) )/(1+(u/t)^g)^2
  s3tg=-((n+m)/t)*((d/t)^g+g*(d/t)^g*log(d/t))*(1+(d/t)^g)
    -g*(d/t)^(2*g)*log(d/t) )/(1+(d/t)^g)^2
  s1gg=sum( ((x/t)^g*log(x/t))^2*(1+(x/t)^g)
    -((x/t)^g*log(x/t))^2)/(1+(x/t)^g)^2 )
  s2gg=m*( (u/t)^g*log(u/t))^2*(1+(u/t)^g)
    -((u/t)^g*log(u/t))^2 )/(1+(u/t)^g)^2
  s3gg=(n+m)*((d/t)^g*log(d/t))^2*(1+(d/t)^g)
    -((d/t)^g*log(d/t))^2 )/(1+(d/t)^g)^2

  g1=-n*s2t/(s1+s2-s3)+n*s3t/(s1+s2-s3)-n*g/t-(n/(s1+s2-s3)+1)*s1t
  g2=-n*s2g/(s1+s2-s3)+n*s3g/(s1+s2-s3)+n/g+sum(log(x))
    -n*log(t)-(n/(s1+s2-s3)+1)*s1g
  g11=-n*(s2tt*(s1+s2-s3)-s2t*(s1t+s2t-s3t))/(s1+s2-s3)^2
    +n*(s3tt*(s1+s2-s3)-s3t*(s1t+s2t-s3t))/(s1+s2-s3)^2
    -s1t*(s1t+s2t-s3t)/(s1+s2-s3)^2-s1tt
  g12=-n*(s2tg*(s1+s2-s3)-s2t*(s1g+s2g-s3g))/(s1+s2-s3)^2
    +n*(s3tg*(s1+s2-s3)-s3t*(s1g+s2g-s3g))/(s1+s2-s3)^2-n/t-n
    *(s1tg*(s1+s2-s3)-s1t*(s1g+s2g-s3g))/(s1+s2-s3)^2-s1tg
  g21=-n*(s2tg*(s1+s2-s3)-s2g*(s1t+s2t-s3t))/(s1+s2-s3)^2
    +n*(s3tg*(s1+s2-s3)-s3g*(s1t+s2t-s3t))/(s1+s2-s3)^2
    -n/t-n*(s1tg*(s1+s2-s3)-s1g*(s1t+s2t-s3t))/(s1+s2-s3)^2-s1tg
  g22=-n*(s2gg*(s1+s2-s3)-s2g*(s1g+s2g-s3g))/(s1+s2-s3)^2
    +n*(s3gg*(s1+s2-s3)-s3g*(s1g+s2g-s3g))/(s1+s2-s3)^2-n/g^2
    -n*(s1gg*(s1+s2-s3)-s1g*(s1g+s2g-s3g))/(s1+s2-s3)^2-s1gg
  theta[i+1]=t-(g1*g22-g2*g12)/(g11*g22-g12*g21); gamma[i+1]=
    g-(g2*g11-g1*g21)/(g11*g22-g12*g21)}

thetaest=theta[itera]
gammaest=gamma[itera]
alphaest=n/(sum(log(1+(x/thetaest)^gammaest))
  +m*log(1+(u/thetaest)^gammaest)
  -(n+m)*log(1+(d/thetaest)^gammaest))

# Variances and confidence intervals

a=-n/alphaest
b=-alphaest*s2tt+alphaest*s3tt+n*gammaest/(thetaest^2)
  -(alphaest+1)*s1tt
c=-alphaest*s2gg+alphaest*s3gg-n/(gammaest^2)-(alphaest+1)*s1gg
d=-s2t+s3t-s1t
e=-s2g+s3g-s1g
f=-alphaest*s2tg+alphaest*s3tg-n/thetaest-(alphaest+1)*s1tg
varalpha=-((f^2-b*c)/(c*d^2-2*e*d*f+a*f^2+b*e^2-a*b*c)
vartheta=-((e^2-a*c)/(c*d^2-2*e*d*f+a*f^2+b*e^2-a*b*c)
vargamma=-((d^2-a*b)/(c*d^2-2*e*d*f+a*f^2+b*e^2-a*b*c)
z=qnorm(0.975, 0, 1)
```


ICa=c(alphaest-z*sqrt(varalpha), alphaest+z*sqrt(varalpha))
 ICt=c(thetaest-z*sqrt(vartheta), thetaest+z*sqrt(vartheta))
 ICg=c(gammaest-z*sqrt(vargamma), gammaest+z*sqrt(vargamma))
 ICa
 ICt
 ICg

References

- [1] Burr I. Cumulative frequency functions. *Annals of Mathematical Statistics*. 1942; 2(13): 215-232.
- [2] Hogg R, Klugman S. *Loss Distributions*. New York: John Wiley & Sons; 2009.
- [3] Klugman S, Panjer H, Willmot G. *Loss Models: From Data to Decisions*. New York: John Wiley & Sons; 2012.
- [4] Embrechts P, Klüppelberg C, Mikosch T. *Modelling Extremal Events: for Insurance and Finance*. New York: Springer Science & Business Media; 2013.
- [5] Albrecher H, Beirlant J, Teugels J. *Actuarial and Statistical Aspect*. New York: John Wiley & Sons; 2017.
- [6] Zimmer W, Keats J, Wang, F. The Burr XII distribution in reliability analysis. *Journal of Quality Technology*. 1998; 30(4):386-394.
- [7] Rao G, Aslam M, Kundu D. Burr-XII distribution parametric estimation and estimation of reliability of multicomponent stress-strength. *Communications in Statistics-Theory and Methods*. 2015; 44(23):4953-4961.
- [8] Wingo D. Maximum likelihood methods for fitting the Burr type XII distribution to life test data. *Biometrical Journal*. 1983; 25(1):77-84.
- [9] Shao Q, Zhou X. A new parametric model for survival data with long-term survivor. *Statistics in Medicine*. 2004; 23(22):3525-3543.
- [10] Kleiber C, Kotz S. *Statistical Size Distributions in Economics and Actuarial Sciences*. New York: John Wiley & Sons; 2003.
- [11] Beirlant J, Goegebeur Y, Verlaak R, Vynckier P. Burr regression and portfolio segmentation. *Insurance: Mathematics and Economics*. 1998; 23(3):231-250.
- [12] Yourstone S, Zimmer W. Non-normality and the design of control charts for averages. *Decision Sciences*. 1992; 23(5):1099-1113.
- [13] Chou C, Chen C, Liu H. Acceptance control charts for non-normal data. *Journal of Applied Statistics*. 2005; 32(1):25-36.
- [14] Wingo D. Maximum likelihood methods for fitting the Burr type XII distribution to multiply (progressively) censored life test data. *Metrika*. 1993; 40(1):203-210.
- [15] Wang F, Keats J, Zimmer W. Maximum likelihood estimation of the Burr XII parameters with censored and uncensored data. *Microelectronics Reliability*. 1996; 36(3):359-362.
- [16] Ghitany M, Al-Awadhi S. Maximum likelihood estimation of Burr XII distribution parameters under random censoring. *Journal of Applied Statistics*. 2002; 29(7):955-965.
- [17] A. Abuzaid. The estimation of the Burr-XII parameters with middle-censored data. *SpringerPlus* vol. 4, no. 1, pp 1-10, 2015.
- [18] Aslam M, Mahmood Y, Lio Y, Tsai T, Khan M. Double acceptance sampling plans for Burr type XII distribution percentiles under the truncated life test. *Journal of the Operational Research Society*. 2012; 63(7):1010-1017.
- [19] Tsai T, Lio Y, Jiang N, Lin Y, Fan Y. Economical sampling plans with warranty based on truncated data from Burr type XII distribution. *Journal of the Operational Research Society*. 2015; 66(9):1511-1518.
- [20] Balakrishnan N, Mitra D. Left truncated and right censored Weibull data and likelihood inference with an illustration. *Computational Statistics & Data Analysis*. 2012; 56(12):4011-4025.
- [21] Balakrishnan N, Mitra D. EM-based likelihood inference for some lifetime distributions based on left truncated and right censored data and associated model discrimination. *South African Statistical Journal*. 2014; 48(12):125-171.
- [22] Kundu D, Mitra D. A. Analysis of left truncated and right censored competing risks data. *Computational Statistics & Data Analysis*. 2017; 108:12-26.
- [23] Kundu D, Mitra D. Statistical inference based on left truncated and interval censored data from log-location-scale family of distributions. *Communications in Statistics Simulation and Computation*. 2021; 50(4):1073-1093.
- [24] Afify A, Cordeiro G, Ortega E, Yousof H, Butt N. The four-parameter Burr XII distribution: Properties, regression model, and applications. *Communications in Statistics-Theory and Methods*. 2018; 47(11):2605-2624.
- [25] Paranaíba P, Ortega E, Cordeiro G, Pascoa M. The Kumaraswamy Burr XII distribution: theory and practice. *Journal of Statistical Computation and Simulation*. 2021; 83(11):2117-2143.
- [26] Olapade A. On a six-parameter generalized Burr XII distribution. *arXiv preprint 0806.1579:1-8*.
- [27] El-Bassiouny A, Abdo N. Reliability properties of seven parameters Burr XII distribution. *Computational Methods in Science and Technology*. 2010; 16:127-133.
- [28] Adeyemi S, Adebajji A. Moments of order statistics from doubly truncated Burr XII distribution: A complementary note with applications. *Journal of Statistical Research*. 2015; 41(1):37-50.
- [29] Gomes A, da-Silva C, Cordeiro, G. Two extended Burr models: Theory and practice. *Journal of Applied Statistics*. 2015; 44(8):1706-1734.
- [30] Mead M. A new generalization of Burr XII distribution. *Journal of Statistics: Advances in Theory and Applications*. 2014; 12(2):53-73.
- [31] Lanjoni B, Ortega E, Cordeiro G. Extended Burr XII regression models: theory and applications. *Journal of Agricultural, Biological, and Environmental Statistics*. 2016; 21(1):203-224.
- [32] Shakhatareh M, Al-Masri A. On the weighted Burr XII distribution: theory and practice. *Electronic Journal of Applied Statistical Analysis*. 2015; 13(1):229-255.
- [33] Süli E, Mayers, D. *An Introduction to Numerical Analysis*. 2003, Cambridge University Press.
- [34] Efron B, Hinkley, D. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*. 1978; 65(3): 457-483
- [35] Watkins, A. A note on expected fisher information for the Burr XII distribution. *Microelectronics Reliability*. 1997; 37(12), 1849-1852.

- [36] Brazauskas, V. Information matrix for Pareto (IV), Burr, and related distributions. *Communications in Statistics-Theory and Methods*. 2003; 32(2), 315-325.
- [37] Dutang C, Goulet V, Pigeon M. actuar: An R package for actuarial science, *Journal of Statistical Software*. 2008; 25(7):1-37.
- [38] R Core Team. R. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2018; <https://www.R-project.org/>.
- [39] Okasha M, Matter M. On the three-parameter Burr type XII distribution and its application to heavy tailed lifetime data. *Journal of Statistics: Advances in Theory and Applications*. 2014; 12(2):53-73.
- [40] Lee E, Wang, J. *Statistical Methods for Survival Data Analysis*. 2003, John Wiley & Sons.
- [41] Al-Shomrani A, Shawky A, Arif O, Aslam, M. Log-logistic Distribution for Survival Data Analysis Using MCMC. *Springer-Plus*. 2016; 5:1-16.