Unit Modified Burr-III Distribution: Estimation, Characterizations and Validation Test

Muhammad Ahsan ul Haq
College of Statistical & Actuarial Sciences, University of the Punjab, Lahore, Pakistan.
Quality Enhancement Cell, National College of Arts, Lahore-Pakistan.
E-mail: ahsanshani36@gmail.comm

Sharqa Hashmi
Lahore College for Women University (LCWU) Lahore Pakistan
College of Statistical & Actuarial Sciences, University of the Punjab, Lahore, Pakistan.
Email: sharqa1972@yahoo.com

Khaoula AIDI
Laboratory of probability and statistics LaPS, University Badji Mokhtar- Annaba- Algeria.
E-mail: khaoula.aidi@yahoo.fr

Pedro Luiz Ramos
Institute of Mathematical Science and Computing, University of São Paulo, São Carlos, SP, Brazil.
E-mail: pedrolramos@usp.br

Francisco Louzada
Institute of Mathematical Science and Computing, University of São Paulo, São Carlos, SP, Brazil.
E-mail: louzada@icmc.usp.br

Abstract

In this paper, a new three-parameter unit probability distribution is proposed. The new model is a generalization of Burr III distribution, and it is more flexible than some existing well-known distribution due to its different shapes of the hazard function and probability density functions. The mathematical properties of this distribution are presented, including moments, reliability measures, mean residual life, and characterizations, and we also propose a modified chi-squared goodness-of-fit test based on Nikulin–Rao–Robson statistic Y^2 in the presence of complete and censored data. The parameters related to the proposed distribution are estimated using well-known estimation methods. A numerical simulations study is conducted for reinforcement of the results. In the end, we considered two real datasets to illustrate the applicability of the proposed model.

Keywords: Burr III distribution, moments, estimation, characterization, goodness-of-fit test.

1. Introduction

The use of unit distributions plays a crucial role in modeling proportions that are usually observed in industry, medical applications and risk analysis to list a few. The two-parameter distribution that is extensively used for modeling bounded data is the Beta distribution (Gupta & Nadarajah 2004). In this context, several probability distributions have been proposed for handling bounded data sets in different fields. Notable among them are Johnson SB distribution (Johnson 1949), Unit-Logistic distribution proposed Menezes et al. 2018, Topp-Leone distribution (Topp and Leone 1955), Kumaraswamy distribution (Kumaraswamy 1980), Unit-Gompertz distribution (Mazucheli et al. 2019), Unit-Birnbaum-Saunders distribution (Mazucheli et al., 2018), Unit-Weibull distribution by (Mazucheli et al. 2018) and Unit-inverse Gaussian distribution by (Ghitany et al. 2019).

The modified Bur III (MBIII) distribution was proposed by Ali et al. (2015) to model the lifetime data. The authors derived the properties of the new model and discussed its application. Bhatti et al. (2018) characterized MBIII distribution based on two truncated moments, elasticity function, and reversed hazard function. Important generalization for the Bur III can be seen in Usman and Haq (2019) and Chakraborty et al. (2020). On the other hand, some generalizations of MBIII distribution have been attempted by researchers. For example, Ali and Ahmad (2015) introduced transmuted modified Burr III distribution. Haq et al. (2019) introduced generalized odd Burr III-G family of distributions, and Mukhtar et al. (2019) introduced McDonald modified Burr-III distribution.

Let X be a non-negative random variable that follows a MBIII distribution, thenthe probability density function (pdf) and cumulative distribution function (cdf) are given

$$f(x) = \alpha\beta x^{-\beta-1}\left(1+\gamma x^{-\beta}\right)^{-\frac{\alpha}{\gamma}-1}, \quad x > 0 \tag{1}$$

$$F(x) = \left(1+\gamma x^{-\beta}\right)^{-\frac{\alpha}{\gamma}}, \tag{2}$$

where$\alpha, \beta, \gamma > 0$ are shape parameters.

In this paper, we introduce a new unit probability distribution with three parametersbased on a unit transformation of the MBIII distribution. The new distribution named Unit-Modified Burr III (UMBIII) distribution is very flexible and can be used to describe different datasets where the range is included between 0 and 1. Further, we also derive the mathematical properties of the UMBIII and inferential procedures to estimate the parameters under different estimation methods. Two real datasets are used to illustrate the suitability of the proposed model over other well-known models. The originality of this study comes from the fact that this is the first unit modification based on the Modified Burr-III distribution. In this sense, our three-parameter version is very flexible when compared with the standard unit distributions. Additionally, the new mathematical properties and inference for the parameters of the proposed model have not been presented previously. A simulation study is conducted which selects the maximum product spacing (MPS) estimators as the best estimation procedure among the proposed methods. Finally, we proposed a modified Chi-square goodness of fit test to verify the quality of fit for the empirical data.

The study is organized as follows. We derived the proposed distribution and its description in Section 2. Mathematical properties, such as quantile function, moments, mgf, reliability measures, mean residual life, and characterizations related to truncated moments and failure rate function, are presented in Section 3. The distributional parameters are estimated by the method of maximum likelihood for censored and complete data in Section 4. An extensive simulation study is carried out considering different estimation methods in Section 5. Section 6 is devoted to introducing a modified Chi-square goodness-of-fit test for the new model assuming complete and censored data.The proposed methodology is illustrated in two realapplications which are presented in Section 7 and the final comments are given in Section 8.

2. Unit Modified Burr-III distribution

Suppose that Y follows a MBIII distribution, by the transformation $X = Y/(1+Y)$we obtain a new unit modified Burr-III distribution. In this case, the cdf can be defined as:

$$F(x) = \left[1 + \gamma\left(\frac{x}{1-x}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}, \quad 0 < x < 1, \tag{3}$$

where $\alpha, \beta, \gamma > 0$ and its corresponding pdf is given by

$$f(x) = \alpha\beta x^{-2}\left(\frac{1-x}{x}\right)^{\beta-1}\left[1 + \gamma\left(\frac{1-x}{x}\right)^{\beta}\right]^{-\frac{\alpha}{\gamma}-1}. \tag{4}$$

Using binomial expansion pdf can be written as

$$f(x) = \alpha\beta\sum_{i=0}^{\infty}(-1)^i(\gamma)^i\binom{\frac{\alpha}{\gamma}+i}{i}x^{-2}\left(\frac{1-x}{x}\right)^{\beta(i+1)-1}. \tag{5}$$

The behaviour of the shapes of the UMBIII distribution has different forms. Firstly, near $x = 0$ the curve decreases from infinity, or starts from a particular point on the vertical axis, or it starts increasing near to the origin. We obtain the solution of the following limit:

$$\lim_{x\to 0} f(x) = \lim_{x\to 0}\alpha\beta x^{-2}\left(\frac{1-x}{x}\right)^{\beta-1}\left[1 + \gamma\left(\frac{1-x}{x}\right)^{\beta}\right]^{-\frac{\alpha}{\gamma}-1}.$$

The limiting behavior of the pdf is achieved by considering the L'Hopital rule.

$$\lim_{x\to 0} f(x) = \lim_{x\to 0}\frac{\left(\frac{1-x}{x}\right)^{-(\beta-1)}\alpha\left\{x^{\beta+1}(\beta-1)(1-x)^{-\beta}+(1-x)^{1-\beta}x^{\beta}(\beta+1)\right\}\left[1+\gamma\left(\frac{1-x}{x}\right)^{\beta}\right]^{\frac{\alpha}{\gamma}+2}}{\gamma\left(1+\frac{\alpha}{\gamma}\right)\left(\frac{1-x}{x^2}+\frac{1}{x}\right)},$$

$$\lim_{x\to 0} f(x) = \begin{cases} 0 & \beta \geq 1, \alpha \geq \gamma \\ \infty & \beta < 1, \alpha \leq \gamma\,. \end{cases}$$

There is one more possibility that the density curve starts at a specific value on the vertical axis and reach the x-axis at zero or goes upwards when approaches to infinity for $\beta > 1, \alpha < \gamma$.

Figure 1 presents the shapes of the pdf of UMBIII distribution for different parameter values.

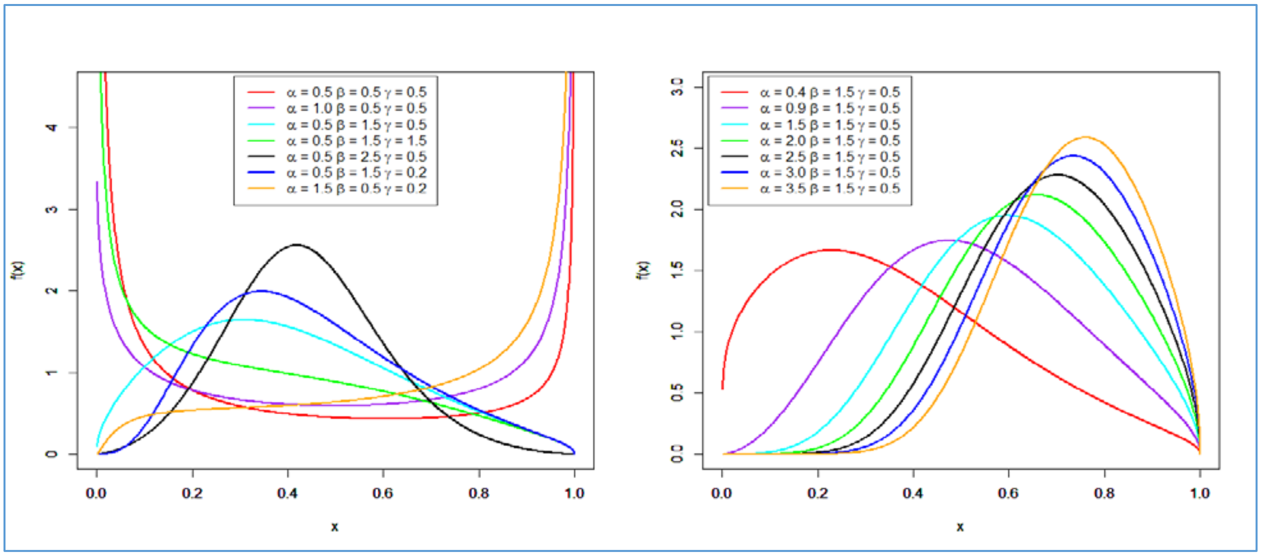

Figure 1: Pdf curves of UMBIII distribution for some parameter values.

From Figure 1, we observe that the pdf has many different shapes such as bathtub shape, left-skewed (negative skewness) as well as right-skewed (positive skewness). Hence, the proposed model is very flexible to fit unit data.

The survival function related to the new model as well as the hazard function are given by

$$S(x) = 1 - \left[1 + \gamma\left(\frac{x}{1-x}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}, \tag{6}$$

$$h(x) = \frac{\alpha\beta x^{-2}\left(\frac{1-x}{x}\right)^{\beta-1}\left[1 + \gamma\left(\frac{1-x}{x}\right)^{\beta}\right]^{-\frac{\alpha}{\gamma}-1}}{1 - \left[1 + \gamma\left(\frac{x}{1-x}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}}. \tag{7}$$

Figure 2 presents the different shapes for the hazard function of UMBIII distribution under different parameter values.

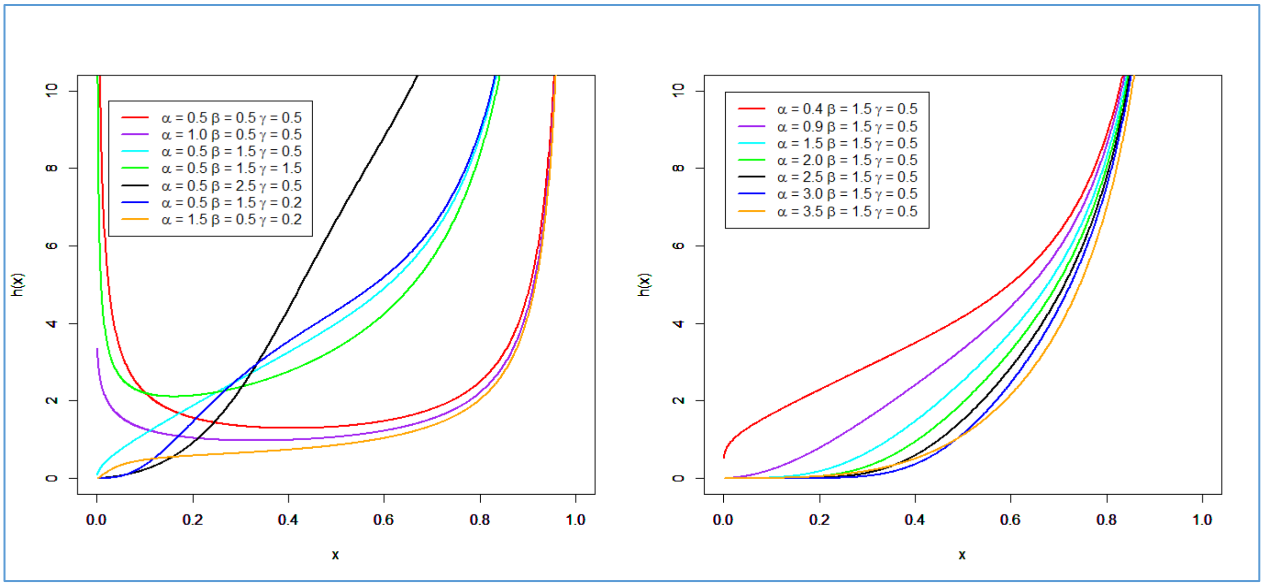

Figure 2: Hrf curves of UMBIII distribution for some parameter values.

3. Mathematical Properties

The quantile function of UMBIII distribution can be obtained by inverting Equation (3) which has the form

$$x = \frac{1}{1 + \left(-\frac{1-p^{-\frac{\gamma}{\alpha}}}{\gamma}\right)^{\frac{1}{\beta}}}. \tag{8}$$

Suppose that the random variable X follows an UMBIII distributed, then its rth moment around zero is given by

$$\mu_r' = \alpha\beta \sum_{i=0}^{\infty} (-1)^i (\gamma)^i \binom{\frac{\alpha}{\gamma}+i}{i} B[r - \beta(i+1), \beta(i+1)] \tag{9}$$

where $B(\cdot,\cdot)$ is the beta function.

Another important function is the moment generating function of UMBIII distribution which can be obtained as

$$M_X(t) = \alpha\beta \sum_{r=0}^{\infty} \frac{(-1)^r t^r}{r!} \sum_{i=0}^{\infty} (-1)^i (\gamma)^i \binom{\frac{\alpha}{\gamma}+i}{i} B[r - \beta(i+1), \beta(i+1)]. \tag{10}$$

The mean residual life (MRL) function is the second most important function used to represent lifetime distributions. It determines the remaining life of a component or unit of a system that has survived up to a particular point in time. That is, it measures the life expectancy of a component or unit that has survived up to time x. Therefore, we obtain the rth moment of the residual life of X via the general formula

$$M_r = E((X-t)^r | X > t) = \frac{1}{1-F(t)} \int_t^{\infty} (x-t)^r f(x) dx, \qquad r = 1,2, \ldots.$$

Using binomial series and Eq. (5)

$$M_r = \frac{1}{1-F(t)}\alpha\beta\sum_{i=0}^{\infty}(-1)^i(\gamma)^i\binom{\frac{\alpha}{\gamma}+i}{i}\sum_{n=0}^{r}(-1)^n\binom{r}{n}t^{r-n}\int_t^1 x^{n-\beta(i+1)-1}(1-x)^{\beta(i+1)-1}dx,$$

the rth moment of the residual life of the UMBIII distribution as

$$M_r = \frac{1}{1-F(t)}\alpha\beta\sum_{i=0}^{\infty}(-1)^i(\gamma)^i\binom{\frac{\alpha}{\gamma}+i}{i}\sum_{n=0}^{r}(-1)^n\binom{r}{n}t^{r-n}B[n-\beta(i+1),\beta(i+1)]\,. \quad (11)$$

The mean residual life (MRL) function is expressed by $M_1 = E\big((X-t)\big|X>t\big)$ and it can be obtained by setting $r=1$ in (11).

In the cases where a new stochastic function is going to be introduced, it is important to check the necessary conditions of a specific underlying model. The study of characterizations can fulfill this requirement. Many characterization techniques have been presented by Glänzel (1987, 1990), Glänzel & Hamedani (2001) and Hamedani (1993, 2002). The characterizations of UMBIII distribution considering the ratio of two truncated moments are discussed. In order to present the characterization of distribution, we considered the theorem presented in Glänzel (1987).

Proposition 3.1: Assume thatX:$\Omega \to (0,1)$ and is distributed following the pdf (4) and

$$q_1(x) = \left\{1+\left(\frac{1-x}{x}\right)^{\beta}\gamma\right\}^{1+\frac{\alpha}{\gamma}}$$
$$q_2(x) = q_1(x)\;\;\alpha\left(\frac{1}{x}-1\right)^{\beta},\quad x>0. \quad (12)$$

Then, the rv X follows UMBIII if and only if the function η presented in the Theorem of Glänzel (1987)has the form

$$\eta(x) = \frac{1}{2}\left(\frac{1}{x}-1\right)^{\beta}\alpha. \quad (13)$$

Proof: Let $0 < x < 1$, we have that

$$\big(1-F(x)\big)E[q_1(X)|X\geq x] = \int_x^1 (1-x)^{-1+\beta}x^{-1-\beta}\alpha\beta\,dx = \alpha\left(\frac{1}{x}-1\right)^{\beta}.$$

Similarly,

$$\big(1-F(x)\big)E[q_2(X)|X\geq x] = \frac{1}{2}\left(\frac{1}{x}-1\right)^{2\beta}\alpha^2.$$

Now

$$\eta(x) = \frac{E[q_2(X)|X\geq x]}{E[q_1(X)|X\geq x]} = \frac{1}{2}\left(\frac{1}{x}-1\right)^{\beta}\alpha$$

and $\eta(x)q_1(x) - q_2(x) = q_1(x)\left[\eta(x) - \alpha\left(\frac{1}{x}-1\right)^{\beta}\right]$

$$= -\frac{1}{2}\alpha\left(\frac{1}{x}-1\right)^{\beta}\left\{1+\left(\frac{1-x}{x}\right)^{\beta}\gamma\right\}^{1+\frac{\alpha}{\gamma}} < 0$$

$$\acute{s}\,(x) = \frac{\acute{\eta}(x)q_1(x)}{\eta(x)q_1(x)-q_2(x)} = \frac{\beta}{\left(\frac{1}{x}-1\right)x^2}$$

and hence

$$s(x) = -\ \ln\left(\alpha\left(\frac{1}{x}-1\right)^{\beta}\right). \tag{14}$$

Now by Proposition (3.1), X has a density (4).

Corollary 3.1: Assume that $X\colon \Omega \to (0,1)$is a continuous rv and let $q_1(x)$ be the same as the one presented in Proposition (3.1). Then we have that the pdf of X is of the form (4) if and only if there exist functions $q_2(x)$ and η presented in the Theorem of Glänzel (1987) satisfying the following differential equation

$$\frac{\acute{\eta}(x)q_1(x)}{\eta(x)q_1(x)-q_2(x)} = \frac{-\frac{\left(\frac{1}{x}-1\right)^{\beta-1}\alpha\beta}{x^2}}{\left(\frac{1}{x}-1\right)^{\beta}\alpha}. \tag{15}$$

The solution of the differential equation above given by

$$\eta(x) = \left[\left(\frac{1}{x}-1\right)^{\beta}\alpha\right]^{-1}\left[\int_0^x x^{-2}\left(\left(\frac{1}{x}-1\right)^{\beta-1}\alpha\beta\right)[q_2(x)\{q_1(x)\}^{-1}]dx + D\right],$$

where D is a constant. The family of functions satisfying the differential equation (15) is presented in Proposition (3.1) with $D = 0$. It is worth mentioning that there are other triplets (q_1, q_2, η) that can satisfies the condition of the Theorem of Glänzel (1987)

Here, we also present the characterizations of UMBIII distribution in term of its the hazard function. An important fact is that the hazard function satisfies the differential equation

$$\frac{\acute{f}(x)}{f(x)} = \frac{\acute{h}(x)}{h(x)} - h(x).$$

Proposition 3.2: Let X be a r. v. with pdf (4) then the hrf satisfies the differential equation

$$\acute{h}(x) - (\beta+1)x^{-1}h(x) = \frac{(1-x)^{\beta-2}x^{-(\beta+2)}\alpha\beta\left(1+\left(-1+\frac{1}{x}\right)^{\beta}\gamma\right)^{-\left(\frac{\alpha}{\gamma}+2\right)}\left(1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right)^{\alpha/\gamma}}{\left(-1+\left(1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right)^{\alpha/\gamma}\right)^{2}} \times$$

$$\left[-x(\beta-1)\left(1+\left(\frac{1}{x}-1\right)^{\beta}\gamma\right)\left(\left(1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right)^{\alpha/\gamma}-1\right)\right.$$

$$+\beta\left\{-\left(\frac{1}{x}-1\right)^{\beta}(\alpha+\gamma)+\frac{\alpha\left(1+\left(\frac{1}{x}-1\right)^{\beta}\gamma\right)}{\left(\frac{x}{1-x}\right)^{\beta}+\gamma}+\left(\frac{1}{x}-1\right)^{\beta}(\alpha\right.$$

$$\left.\left.+\gamma)\left(1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right)^{\alpha/\gamma}\right\}\right] \tag{16}$$

assuming the boundary conditions $h(0)\geq 0$.

Proof: Assuming that X has the hazard function (7) then

$$\acute{h}(x)=\frac{1}{\left(\left(\left(\frac{x}{1-x}\right)^{\beta}+\gamma\right)\left(\left(1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right)^{\alpha/\gamma}-1\right)^{2}\right)}$$

$$\times\left[(1-x)^{\beta-2}x^{-(\beta+2)}\alpha\beta\left(1+\left(\frac{1}{x}-1\right)^{\beta}\gamma\right)^{-2-\frac{\alpha}{\gamma}}\left(1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right)^{\alpha/\gamma}\left\{\gamma\right.\right.$$

$$-\left\{\left(\frac{x}{1-x}\right)^{\beta}\right\}\left\{1+\left(\frac{1}{x}-1\right)^{\beta}\gamma\right\}\left\{\left(1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right)^{\alpha/\gamma}-1\right\}$$

$$+2x\left\{\left(\frac{x}{1-x}\right)^{\beta}+\gamma\right\}\left\{1+\left(\frac{1}{x}-1\right)^{\beta}\gamma\right\}\left\{\left(\frac{x}{1-x}\right)^{-\beta}\gamma-1\right\}^{\alpha/\gamma}$$

$$+\beta\left\{\left(\frac{x}{1-x}\right)^{\beta}+\alpha-\left(\frac{1}{x}-1\right)^{\beta}\left(\frac{x}{1-x}\right)^{\beta}\alpha+\gamma\right.$$

$$\left.\left.\left.+\left\{\left(\frac{1}{x}-1\right)^{\beta}\alpha-1\right\}\left(\left(\frac{x}{1-x}\right)^{\beta}+\gamma\right)\left\{1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right\}^{\alpha/\gamma}\right\}\right\}\right]$$

and by substitution the above result in $\acute{h}(x)-(\beta+1)x^{-1}h(x)$ and the result follows.

Conversely, if Eq.(16) holds then

$$\frac{d}{dx}\left[x^{\beta+1}h(x)\right]=\frac{d}{dx}\left[(1-x)^{\beta-1}\alpha\beta\left\{1+\left(\frac{1}{x}-1\right)^{\beta}\gamma\right\}^{-\frac{\alpha+\gamma}{\gamma}}\times\left(1+\frac{1}{-1+\left(1+\left(\frac{x}{1-x}\right)^{-\beta}\gamma\right)^{\alpha/\gamma}}\right)\right]$$

$$h(x)=\frac{(1-x)^{-1+\beta}x^{-1-\beta}\alpha\beta(1+(\frac{1-x}{x})^{\beta}\gamma)^{-1-\frac{\alpha}{\gamma}}}{1-(1+(\frac{x}{1-x})^{-\beta}\gamma)^{-\frac{\alpha}{\gamma}}}+C$$

which implies $C=0$.

4 Estimation

In this section, we estimate the unknown parameters using seven different estimation methods such asthe maximum likelihood estimation, ordinary least square (OLS), weighted least square (WLS), percentile (PE), maximum product spacing (MPS), Cramer-von-Mises (CVM) and Anderson Darling (AD). These estimation methods have been considered by many authors for other distributions (Louzada et al., 2020; Ramos et al. 2019). A brief description of these methods is given below.

4.1 Maximum likelihood

Let $x_1, x_2, \dots x_n$ be random samples distributed according to the UMBIII distribution, the likelihood function is obtained from

$$L=\prod_{i=1}^{n} f(x_i,\alpha,\beta,\gamma).$$

Using the pdf (4), we have

$$L=\prod_{i=1}^{n}\alpha\beta x_i^{-2}\left(\frac{1-x_i}{x_i}\right)^{\beta-1}\left[1+\gamma\left(\frac{1-x_i}{x_i}\right)^{\beta}\right]^{-\frac{\alpha}{\gamma}-1}.$$

By taking the natural logarithm, the log-likelihood function is obtained as;

$$\log L=n\log(\alpha\beta)-2\sum_{i=1}^{n}\log(x_i)+(\beta-1)\sum_{i=1}^{n}\log\left(\frac{1-x_i}{x_i}\right)-\left(\frac{\alpha}{\gamma}+1\right)\sum_{i=1}^{n}\log\left[1+\gamma\left(\frac{1-x_i}{x_i}\right)^{\beta}\right].$$

For our model, we considered that $u_i=\frac{1-x_i}{x_i}$. Hence, the maximum likelihood estimators $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\beta}$ of the unknown parameters α, γ, and β are derived from the nonlinear following score equations:

$$\frac{\partial L}{\partial\alpha}=\frac{\mathrm{n}}{\alpha}-\frac{1}{\gamma}\sum_{i=1}^{n}\ln\left(1+\gamma u_i^{\beta}\right),$$

$$\frac{\partial L}{\partial\beta}=\frac{\mathrm{n}}{\beta}+\sum_{i=1}^{n}\ln(u_i)-\gamma\left(\frac{\alpha}{\gamma}+1\right)\sum_{i=1}^{n}\frac{u_i^{\beta}\ln(u_i)}{1+\gamma u_i^{\beta}},$$

$$\frac{\partial L}{\partial\gamma}=\frac{\alpha}{\gamma^2}\sum_{i=1}^{n}\ln\left(1+\gamma u_i^{\beta}\right)-\left(\frac{\alpha}{\gamma}+1\right)\sum_{i=1}^{n}\frac{u_i^{\beta}}{1+\gamma u_i^{\beta}}.$$

Although the second and the third linear equations do not have closed-form expression, we can obtain the estimate from α by considering:

$$\alpha = \frac{\mathrm{n}}{\frac{1}{\gamma}\sum_{i=1}^{n}\ln\left(1+\gamma u_i^{\beta}\right)}.$$

4.2 Ordinary Least Square Estimators

Other common estimation methods are the least-square and weighted least square estimators. Let $X_{(1)} < X_{(2)}, \ldots, X_{(n)}$ be the order statistics related to a random sample obtained from the UMBIII distribution, then the OLSEs of α, β and γ, denoted by $\hat{\alpha}_{OLSE}$, $\hat{\beta}_{OLSE}$ and $\hat{\gamma}_{OLSE}$ are obtained by minimizing

$$S(\alpha,\beta,\gamma) = \sum_{i=1}^{n}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}} - \frac{i}{n+1}\right]^2,$$

in terms of its parameters, i.e., they can be obtained by solving

$$\sum_{i=1}^{n}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}} - \frac{i}{n+1}\right]\omega_1\left(x_{(i)}\middle|\alpha,\beta,\gamma\right) = 0,$$

$$\sum_{i=1}^{n}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}} - \frac{i}{n+1}\right]\omega_2\left(x_{(i)}\middle|\alpha,\beta,\gamma\right) = 0,$$

$$\sum_{i=1}^{n}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}} - \frac{i}{n+1}\right]\omega_3\left(x_{(i)}\middle|\alpha,\beta,\gamma\right) = 0,$$

where

$$\omega_1\left(x_{(i)}\middle|\alpha,\beta,\gamma\right) = \frac{\left(1+\left(\frac{1}{x_{(i)}}-1\right)^{\beta}\gamma\right)^{-\frac{\alpha}{\gamma}}\log\left(1+\gamma\left(\frac{1}{x}-1\right)^{\beta}\right)}{\gamma}, \quad (17)$$

$$\omega_2\left(x_{(i)}\middle|\alpha,\beta,\gamma\right) = \alpha\left(\frac{1}{x}-1\right)^{\beta}\left(1+\left(\frac{1}{x}-1\right)^{\beta}\gamma\right)^{-\frac{\alpha}{\gamma}-1}\log\left[\frac{1}{x}-1\right], \quad (18)$$

$$\omega_3\left(x_{(i)}\middle|\alpha,\beta,\gamma\right) = \left(1+\left(\frac{1}{x}-1\right)^{\beta}\gamma\right)^{-\frac{\alpha}{\gamma}}\left(\frac{\alpha\left(\frac{1}{x}-1\right)^{\beta}}{\gamma\left(1+\gamma\left(-1+\frac{1}{x}\right)^{\beta}\right)} - \frac{\alpha\log\left[1+\gamma\left(-1+\frac{1}{x}\right)^{\beta}\right]}{\gamma^2}\right). \quad (19)$$

The WLSEs of to α, β and γ, i.e.$\hat{\alpha}_{OLSE}$, $\hat{\beta}_{OLSE}$ and $\hat{\gamma}_{OLSE}$ can be obtained by minimizing in term of its parameters the following function:

$$W(\alpha,\beta,\gamma)=\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}-\frac{i}{n+1}\right]^2.$$

In the same way of the OLS the WLS estimators can be obtained by solving

$$\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}-\frac{i}{n+1}\right]\omega_1\left(x_{(i)}\middle|\alpha,\beta,\gamma\right)=0$$

$$\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}-\frac{i}{n+1}\right]\omega_2\left(x_{(i)}\middle|\alpha,\beta,\gamma\right)=0$$

$$\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}-\frac{i}{n+1}\right]\omega_3\left(x_{(i)}\middle|\alpha,\beta,\gamma\right)=0.$$

4.3 Percentile Estimators

The percentile methodology was originally proposed by Kao (1959) to estimate the parameters of a probability distribution parameters that has the quantile function in a closed-form expression. Since the quantile function of the new distribution has closed-form, then the estimators of α, β and γ are derived by minimizing with respect to α, β and γ the following expression

$$\sum_{i=1}^{n}\left[\log\left(\frac{i}{n+1}\right)-\left(1+\left(-\frac{1-(\hat{p}_i)^{-\frac{\gamma}{\alpha}}}{\gamma}\right)^{\frac{1}{\beta}}\right)^{-1}\right]^2.$$

4.4 Cramer-von Mises Minimum

Based on Goodness of fit measures the Cramer-von Mises estimators (CVM) for the parameters α, β and γ can be obtained by minimizing the function

$$CV(\alpha,\beta,\gamma)=\frac{1}{12n}+\sum_{i=1}^{n}\left[1-\left[1+\gamma\left(\frac{x_{(i)}}{1-x_{(i)}}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}-\frac{2i-1}{2n}\right]^2 \qquad (20)$$

with respect to α, β and γ.

4.5 Maximum Product Spacing

The Maximum Product Spacing (MPS) method is a powerful estimation procedure used as an alternative of the MLE method to achieve estimates of the parameters in continuous distributions (Cheng and Amin 1983). The uniform spacings of a random sample obtained from the UMBIII distribution is given by

$$D_i(\alpha,\beta,\gamma)=F\left(x_{(i)}|\alpha,\beta,\gamma\right)-F\left(x_{(i-1)}|\alpha,\beta,\gamma\right),\quad i=1,2,\dots,n+1,$$

where$F\left(x_{(0)}|\alpha,\beta,\gamma\right)=0$, $F\left(x_{(n+1)}|\alpha,\beta,\gamma\right)=1$. From the definition above we have that $\sum_{i=1}^{n+1}D_i(\alpha,\beta,\gamma)=1$

The MPS estimator are achieved by maximizing w.r.t. the parameters the geometric mean (GM) of the spacings

$$\mathrm{GM}=\prod_{i=1}^{n+1}\left(D_i(\alpha,\beta,\gamma)\right)^{\frac{1}{n+1}} \tag{21}$$

The MPS estimates of α, β and γcan be obtained by the directed maximization the logarithm of the GM. In this case, we have to consider numerical methods to obtain the solution of the equation above.

4.6 Right tail Anderson Darling

The Right Anderson and Darling estimators (ADEs) is another estimation method obtained from goodness of fit measures. The estimates for α, β and γ are obtained by minimizing the following function,

$$\mathrm{RAD}(\alpha,\beta,\gamma)=\frac{n}{2}+2\sum_{i=1}^{n}F(x_{i:n}|\alpha,\beta,\gamma)-\frac{1}{n}\sum_{i=1}^{n}(2i-1)\log\bar{F}(x_{n+1-i:n}|\alpha,\beta,\gamma). \tag{22}$$

with respect the parameters of interesting.

5 Simulation Studies:

For the support of this research, we conduct a comprehensive simulation study for both cases considering the different estimation methods. To compare the estimation methods we considered the mean relative estimate (MRE) and mean square error (MSE) that are computed by

$$\mathrm{MRE}\left(\theta_j\right)=\frac{1}{N}\sum_{i=1}^{n}\frac{\hat{\theta}_{i,j}}{\theta_j}\quad\text{and}\quad\mathrm{MSE}\left(\theta_j\right)=\frac{1}{N}\sum_{i=1}^{n}\left(\hat{\theta}_{i,j}-\theta_j\right)^2,$$

where $\theta=(\alpha,\beta,\gamma)$ and $\hat{\theta}=\left(\hat{\alpha},\hat{\beta},\hat{\gamma}\right)$ are estimates obtained from the samples Our simulation study was conducted using $N=10{,}000$ samples with sample sizes ofn=20, 30... 400. The following parametric values was considered:

α =2, β =0.5 and γ=1.5 for Figure 3.

α =0.5, β =0.5 and γ=2 for Figure 4.

α =0.5, β =1 and γ=0.5 for Figure 5.

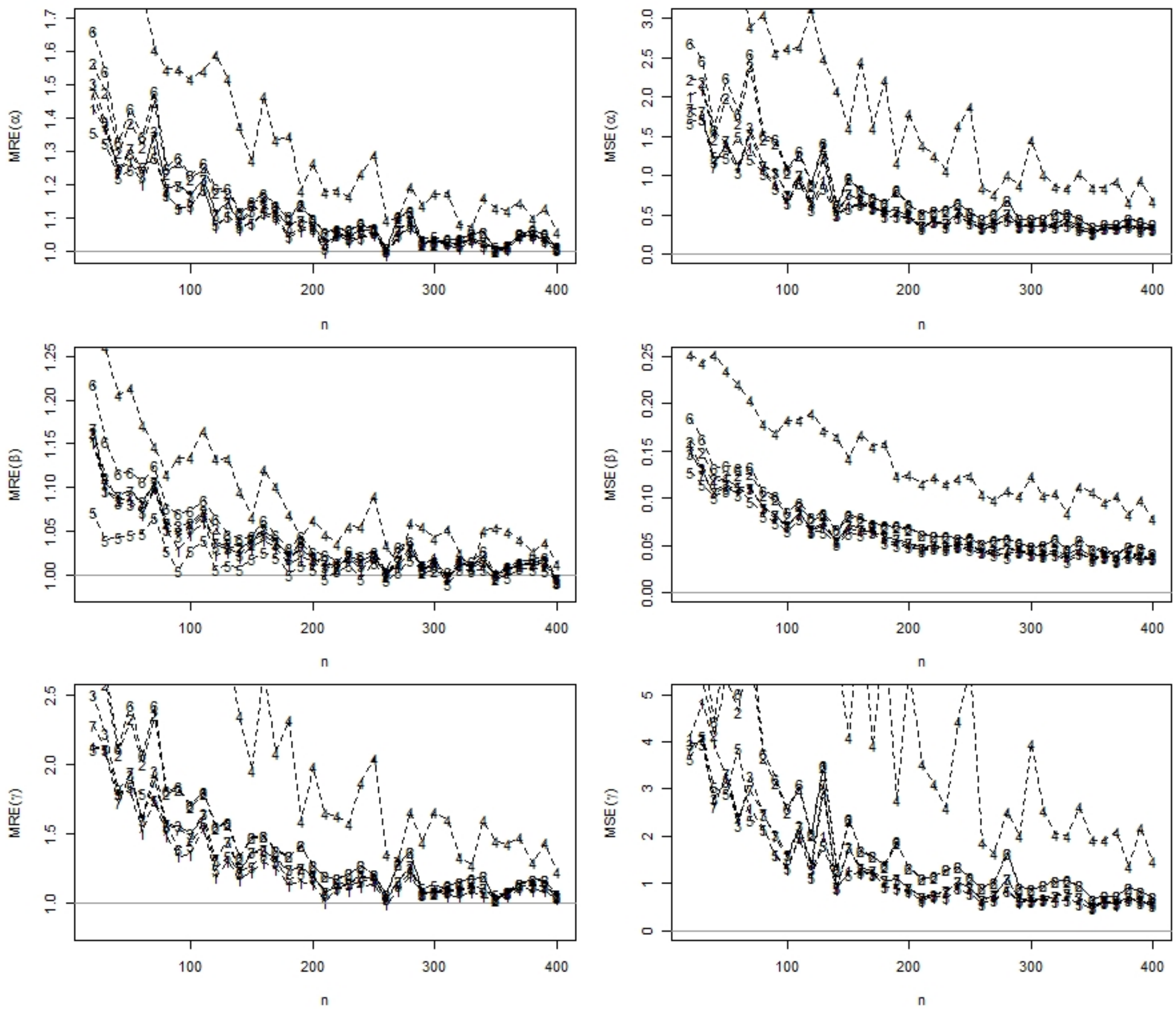

Figure 3: Graphs of MRE and MSE for parameters (α =2, β =0.5 and γ=1.5).

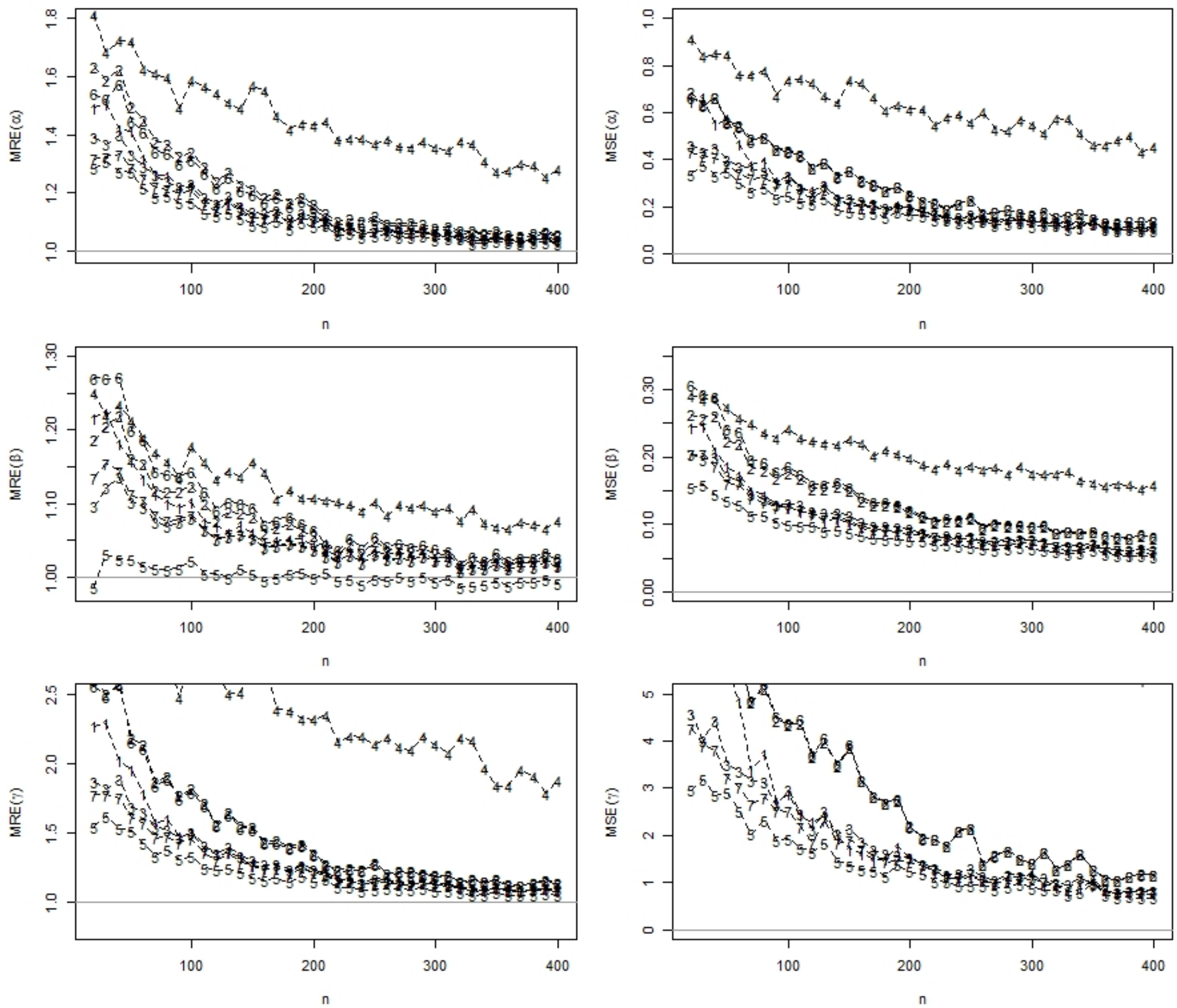

Figure 4: Graphs of MRE and MSE for parameters (α =0.5, β =0.5 and γ=2).

The simulation study was conducted using the software R and the codes are available upon request. The legends from the different estimation methods are 1- MLE, 2 - LSE, 3 - WLSE, 4 - PER, 5 - MPS, 6 - CME, 7 – RDA. Under this scenario, we expect that the best estimation method will return the MRE closer to one with smaller MSE. The results are presented in Figures 3-5.

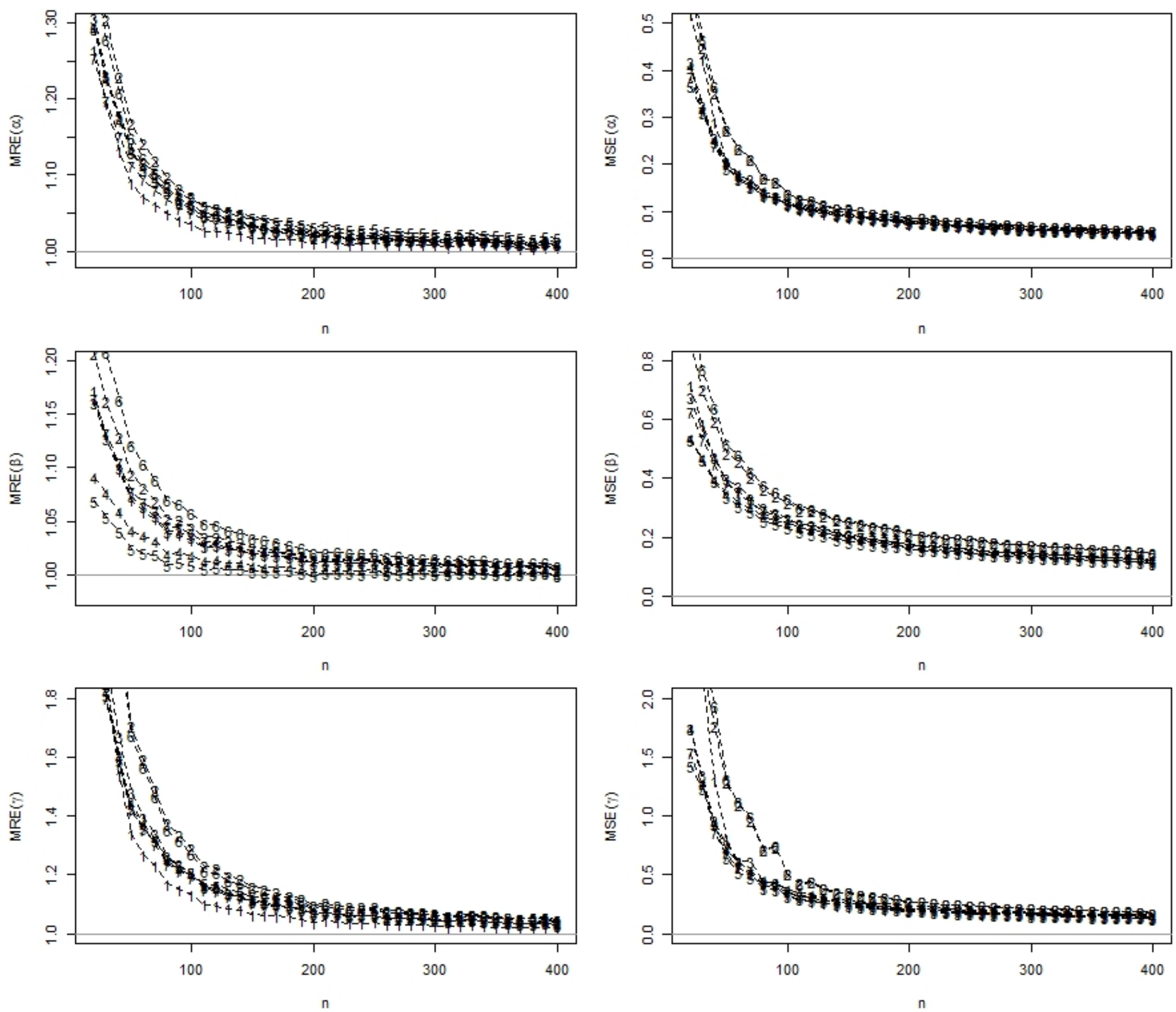

Figure 5: Graphs of MRE and MSE for parameters (α =0.5, β =1 and γ=0.5).

As can be seen in Figure 3 – 5 we observe that the MREs are closer to one for the MPS, this method also returns the minimum MSE for most of the scenarios. On the other hand, the CME returned poor estimates when compared with other estimation methods. Hence, overall the MPS returned best estimates among the estimation methods. Another interesting fact is that the MPS has many interesting properties such as the estimator is invariant under one-to-one transformation, asymptotic efficiency, and, more importantly, the consistency of the MPS holds under more general conditions than for MLEs.

6 Goodness-of-fit test

Some techniques that exist in literature are commonly used to verify if real data can be fitted by statistical models. Generally Chi-square statistics are the most used.

6.1 Estimation under right-censored data

The hypotheses test will be discussed under complete and censored data, however, the MPS is only defined for complete data, since the MLE is usually considered for right-censored data, Let us consider $X_1, X_2, \dots, X_n$ a random right censored sample obtained from the Unit Modified Burr-III distribution with the parameter vector $\theta = (\alpha, \beta, \gamma)^T$. The censoring time τ is fixed. So, the observation X_i is equal to $X_i = (x_i, \delta_i)$ where

$$\delta_i = \begin{cases} 0 & if x_i isacensoringtime \\ 1 & if x_i isafailuretime \end{cases}$$

In this case, the log-likelihood is obtained as follow

$$L_n(\theta) = \sum_{i=1}^{n} \delta_i \, lnh(x_i, \theta) + \sum_{i=1}^{n} ln\, S(x_i, \theta)$$

$$L_n(\theta) = \sum_{i=1}^{n} \delta_i \begin{bmatrix} \ln(\alpha\beta) - 2\ln(x_i) - \left(\frac{\alpha}{\gamma} + 1\right)\ln\left(1 + \gamma u_i^{\beta}\right) \\ -(\beta - 1)ln(u_i) - \ln\left(1 - \left(1 + \gamma u_i^{\beta}\right)^{-\alpha/\gamma}\right) \end{bmatrix} + \sum_{i=1}^{n} \ln\left(1 - \left(1 + \gamma u_i^{\beta}\right)^{-\alpha/\gamma}\right).$$

The maximum likelihood estimators $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\beta}$ of the unknown parameters α, γ, and β are derived from the nonlinear following score equations:

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{n} \delta_i \left[\frac{1}{\alpha} - \frac{\ln\left(1 + \gamma u_i^{\beta}\right)}{\gamma} - \frac{\alpha\left(1 + \gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}} \ln\left(1 + \gamma u_i^{\beta}\right)}{\gamma\left(1 - \left(1 + \gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\right)}\right] + \frac{\alpha}{\gamma}\sum_{i=1}^{n} \frac{\left(1 + \gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}} \ln\left(1 + \gamma u_i^{\beta}\right)}{1 - \left(1 + \gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}}$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} \delta_i \left[\frac{1}{\beta} + \ln(u_i) - \left(\frac{\alpha}{\gamma} + 1\right)\frac{\gamma u_i^{\beta}\ln(u_i)}{1 + \gamma u_i^{\beta}} - \frac{\alpha u_i^{\beta}\ln(u_i)\left(1 + \gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma} - 1}}{1 - \left(1 + \gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}}\right]$$

$$+ \sum_{i=1}^{n} \frac{\alpha u_i^{\beta}\ln(u_i)\left(1 + \gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma} - 1}}{1 - \left(1 + \gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}}$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{n} \delta_i \left[\frac{\alpha\ln\left(1 + \gamma u_i^{\beta}\right)}{\gamma^2} - \left(\frac{\alpha}{\gamma} + 1\right)\frac{u_i^{\beta}}{1 + \gamma u_i^{\beta}} - \frac{\alpha}{\gamma^2}\frac{\gamma u_i^{\beta}\ln\left(1 + \gamma u_i^{\beta}\right) - \gamma u_i^{\beta} + \ln\left(1 + \gamma u_i^{\beta}\right)}{\left(1 + \gamma u_i^{\beta}\right)\left(1 - \left(1 + \gamma u_i^{\beta}\right)^{\frac{\alpha}{\gamma}}\right)}\right]$$

$$+ \frac{\alpha}{\gamma^2}\sum_{i=1}^{n} \frac{\gamma u_i^{\beta}\ln\left(1 + \gamma u_i^{\beta}\right) - \gamma u_i^{\beta} + \ln\left(1 + \gamma u_i^{\beta}\right)}{\left(1 + \gamma u_i^{\beta}\right)\left(1 - \left(1 + \gamma u_i^{\beta}\right)^{\frac{\alpha}{\gamma}}\right)}.$$

Monte Carlo technique or other iterative methods can be used to determine the values of $\hat{\alpha}$,$\hat{\gamma}$ and $\hat{\beta}$.

6.2 Test statistic for right-censored data

Let $X_1,\dots,X_n$ be n i.i.d. random variables grouped into r classes I_i. To assess the adequacy of a parametric model F_0

$$H^0: \quad P(X_i \le x \mid H^0) = F^0(x;\theta), x \ge 0,\ \ \theta = (\theta_1,\dots,\theta_s)^T \in \Theta \subset R^s$$

when data are right-censored and the parameter vector θ is unknown, Bagdonavičius and Nikulin (2011) proposed a statistic test Y^2 based on the vector

$$Z_j = \frac{1}{\sqrt{n}}(U_j - e_j)\,, \quad j = 1,2,\dots,r \ \ , \qquad withr \succ s.$$

This one represents the differences between observed and expected numbers of failures ($U_j$and e_j) to fall into these grouping intervals $I_j = (a_{j-1}, a_j]$ with $a_0 = 0, a_r = \tau$, where τ is considered to be finite. The authors considered a_j as random data functions such as the r intervals chosen have equal expected numbers of failurese_j.

The statistic test Y^2 is defined by

$$Y^2 = Z^T\hat{\Sigma}^{-}Z \ = \sum_{i=1}^{r}\frac{\left(U_j - e_j\right)^2}{U_j} + Q$$

where$Z = (Z^1,\dots,Z_k)^T$ and $\hat{\Sigma}^-$ is a generalized inverse of the covariance matrix $\hat{\Sigma}$ and

$$Q = W^T\hat{G}^{-}W, \hat{A}_J = \frac{U_j}{n}, U_j = \sum_{i:X_i\in I_j}\delta_i$$

$$W = (W_1,\dots,W_s)^T, \hat{G} = [\hat{g}_{ll'}]_{s\times s}, \hat{g}_{ll'} = \hat{\imath}_{ll'} - \sum_{j=1}^{r}\hat{C}_{lJ}\hat{G}_{l'J}\hat{A}_J^{-1}$$

$$\hat{C}_{lj} = \frac{1}{n}\sum_{i:X_i\in I_j}\delta_i\frac{\partial lnh(x_i,\hat{\theta})}{\partial\theta}, \hat{\imath}_{ll'} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{\partial lnh(x_i,\hat{\theta})}{\partial\theta_l}\frac{\partial lnh(x_i,\hat{\theta})}{\partial\theta_{l'}}$$

$$\widehat{W}_l = \sum_{j=1}^{r}\hat{C}_{lJ}\hat{A}_J^{-1}Z_j\ l, l' = 1,\dots,s$$

$\hat{\theta}$is the maximum likelihood estimator of θ on initial non-grouped data.

Under the null hypothesis H_0, the limit distribution of the statistic Y^2 is a chi-square with $r = rank(\Sigma)$ degrees of freedom. The description and applications of modified Chi-square tests are discussed in Bagdonavičius et al. (2013).

The interval limits a_j for grouping data into j classes I_j are considered as data functions and defined by

$$\hat{a}_j = H^{-1}\left(\frac{E_j - \sum_{l=1}^{i-1}H(x_l,\hat{\theta})}{n-i+1}, \hat{\theta}\right),\ \ \hat{a}_j = \max\,(X_{(n)}, \tau)$$

such as the expected failure times e_j to fall into these intervals are $e_j = \frac{E_r}{r}$for any j, with $E_r = \sum_{i=1}^{n} H(x_i, \theta)$. The distribution of this statistic test $Y_n^2$is chi-square (see Bagdonavičius et al. (2013)).

5.2 Criteria test for Unit Modified Burr-III distribution

To verify if data can be described by the Unit Modified Burr-III model, we propose the construction of a modified chi-squared using the statistic Y^2 . Suppose that observed data are grouped into $r > s$ sub-intervals$I_j = (a_{j-1},\ a_j]$of $[0, \tau]$ where τ is a finite time. As limit intervals a_j are defined such as we obtain the same expected numbers of failures in each interval I_j, so the expected numbers of failures e_j are obtained as

$$E_j = -\frac{j}{r-1}\sum_{i=1}^{n} \ln\left(1-\left[1+\gamma\left(\frac{x}{1-x}\right)^{-\beta}\right]^{-\frac{\alpha}{\gamma}}\right), \quad j = 1, \ldots\ldots, r-1$$

Estimated matrix $\widehat{W}$ et $\widehat{C}$

The components of the estimated matrix $\widehat{\boldsymbol{W}}$ are obtained from the estimated matrix $\widehat{\boldsymbol{C}}$ which is given by:

$$\hat{C}_{1j} = \frac{1}{n}\sum_{i:x_i\in I_j}^{n} \delta_i\left[\frac{1}{\alpha} - \frac{\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma} - \frac{\alpha\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma\left(1-\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\right)}\right]$$

$$\hat{C}_{2j} = \frac{1}{n}\sum_{i:x_i\in I_j}^{n} \delta_i\left[\frac{1}{\beta} + \ln(u_i) - \left(\frac{\alpha}{\gamma}+1\right)\frac{\gamma u_i^{\beta}\ln(u_i)}{1+\gamma u_i^{\beta}} - \frac{\alpha u_i^{\beta}\ln(u_i)\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}-1}}{1-\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}}\right]$$

$$\hat{C}_{3j} = \frac{1}{n}\sum_{i:x_i\in I_j}^{n} \delta_i\left[\frac{\alpha\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma^2} - \left(\frac{\alpha}{\gamma}+1\right)\frac{u_i^{\beta}}{1+\gamma u_i^{\beta}} - \frac{\alpha}{\gamma^2}\frac{\gamma u_i^{\beta}\ln\left(1+\gamma u_i^{\beta}\right) - \gamma u_i^{\beta} + \ln\left(1+\gamma u_i^{\beta}\right)}{\left(1+\gamma u_i^{\beta}\right)\left(1-\left(1+\gamma u_i^{\beta}\right)^{\frac{\alpha}{\gamma}}\right)}\right]$$

and

$$\widehat{W}_l = \sum_{j=1}^{r} \hat{C}_{lJ}\hat{A}_J^{-1}Z_j\ l, l' = 1,2,3, \quad j = 1, \ldots, r$$

Estimated Matrix $\widehat{\boldsymbol{G}}$

The estimated matrix $\hat{G} = [\hat{g}_{ll'}]_{3\times3}$is defined by

$$\hat{g}_{ll'} = \hat{\imath}_{ll'} - \sum_{j=1}^{r} \hat{C}_{lJ}\hat{G}_{l'J}\hat{A}_J^{-1}$$

where

$$\hat{\imath}_{ll'} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \frac{\partial lnh(x_i,\hat{\theta})}{\partial\theta_l}\frac{\partial lnh(x_i,\hat{\theta})}{\partial\theta_{l'}}\, l, l' = 1,2,3$$

or

$$\hat{\imath}_{11} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{1}{\alpha} - \frac{\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma} - \frac{\alpha\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma\left(1-\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\right)}\right]^2$$

$$\hat{\imath}_{22} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{1}{\beta} + \ln(u_i) - \left(\frac{\alpha}{\gamma}+1\right)\frac{\gamma u_i^{\beta}\ln(u_i)}{1+\gamma u_i^{\beta}} - \frac{\alpha u_i^{\beta}\ln(u_i)\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}-1}}{1-\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}}\right]^2$$

$$\hat{\imath}_{33} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{\alpha\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma^2} - \left(\frac{\alpha}{\gamma}+1\right)\frac{u_i^{\beta}}{1+\gamma u_i^{\beta}} - \frac{\alpha}{\gamma^2}\frac{\gamma u_i^{\beta}\ln\left(1+\gamma u_i^{\beta}\right) - \gamma u_i^{\beta} + \ln\left(1+\gamma u_i^{\beta}\right)}{\left(1+\gamma u_i^{\beta}\right)\left(1-\left(1+\gamma u_i^{\beta}\right)^{\frac{\alpha}{\gamma}}\right)}\right]^2$$

$$\hat{\imath}_{12} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{1}{\alpha} - \frac{\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma} - \frac{\alpha\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma\left(1-\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\right)}\right]$$

$$\times\left[\frac{1}{\beta} + \ln(u_i) - \left(\frac{\alpha}{\gamma}+1\right)\frac{\gamma u_i^{\beta}\ln(u_i)}{1+\gamma u_i^{\beta}} - \frac{\alpha u_i^{\beta}\ln(u_i)\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}-1}}{1-\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}}\right]$$

$$\hat{\imath}_{13} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{1}{\alpha} - \frac{\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma} - \frac{\alpha\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma\left(1-\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}\right)}\right]$$

$$\times\left[\frac{\alpha\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma^2} - \left(\frac{\alpha}{\gamma}+1\right)\frac{u_i^{\beta}}{1+\gamma u_i^{\beta}} - \frac{\alpha}{\gamma^2}\frac{\gamma u_i^{\beta}\ln\left(1+\gamma u_i^{\beta}\right) - \gamma u_i^{\beta} + \ln\left(1+\gamma u_i^{\beta}\right)}{\left(1+\gamma u_i^{\beta}\right)\left(1-\left(1+\gamma u_i^{\beta}\right)^{\frac{\alpha}{\gamma}}\right)}\right]$$

$$\hat{\imath}_{23} = \frac{1}{n}\sum_{i=1}^{n}\delta_i = \left[\frac{1}{\beta} + \ln(u_i) - \left(\frac{\alpha}{\gamma}+1\right)\frac{\gamma u_i^{\beta}\ln(u_i)}{1+\gamma u_i^{\beta}} - \frac{\alpha u_i^{\beta}\ln(u_i)\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}-1}}{1-\left(1+\gamma u_i^{\beta}\right)^{-\frac{\alpha}{\gamma}}}\right]$$

$$\times\left[\frac{\alpha\ln\left(1+\gamma u_i^{\beta}\right)}{\gamma^2} - \left(\frac{\alpha}{\gamma}+1\right)\frac{u_i^{\beta}}{1+\gamma u_i^{\beta}} - \frac{\alpha}{\gamma^2}\frac{\gamma u_i^{\beta}\ln\left(1+\gamma u_i^{\beta}\right) - \gamma u_i^{\beta} + \ln\left(1+\gamma u_i^{\beta}\right)}{\left(1+\gamma u_i^{\beta}\right)\left(1-\left(1+\gamma u_i^{\beta}\right)^{\frac{\alpha}{\gamma}}\right)}\right]$$

Therefore, the quadratic form of the test statistic can be obtained easily:

$$Y_n^2(\hat{\beta}) = \sum_{j=1}^{r} \frac{\left(U_j - e_j\right)^2}{U_j} + \widehat{W}^T \left[\hat{\imath}_{ll'} - \sum_{j=1}^{r} \hat{C}_{lJ} \hat{G}_{l'J} \hat{A}_J^{-1}\right]^{-1} \widehat{W}$$

6.3 Maximum likelihood estimation using right censored samples:

In this section, $N = 10{,}000$ right censored samples are simulated from the Unit Modified Burr-III model with parameters α=0.8,$\beta = 1.5$ and $\gamma = 1.7$. Different sizes are considered n= 30, 50,150,350,500. To calculate the maximum likelihood estimates and their mean squared errors (MSE), we use Barzilai-Borwein (BB) algorithms (Ravi, 2009) . Results are presented in Table 1.

Table1. Obtained values of MLEs $\hat{\alpha}, \hat{\beta}$ and $\hat{\gamma}$ and their corresponding square mean errors

$N = 10{,}000$	$n_1 = 30$	$n_2 = 50$	$n_3 = 150$	$n_4 = 350$	$n_5 = 500$
$\hat{\alpha}$	0.9328	0.8839	0.8678	0.8320	0.8003
$MSE(\hat{\alpha})$	0.0122	0.0092	0.0088	0.0049	0.0023
$\widehat{\theta}$	1.7263	1.6736	1.6239	1.5312	1.5013
$MSE(\widehat{\theta})$	0.0102	0.0095	0.0083	0.0037	0.0027
$\widehat{\gamma}$	1.5939	1.6246	1.6523	1.6836	1.7009
$MSE(\widehat{\gamma})$	0.0073	0.0056	0.0035	0.0028	0.0017

The maximum likelihood estimates presented in Table 1, agree closely with the true parameter values.

Criteria test Y_n^2

For testing the null hypothesis H_0 that right-censored data become from Unit Modified Burr-III model, we compute the criteria statistic $Y_n^2(\theta)$ as defined above for 10,000 simulated samples from the hypothesized distribution with different sizes (30, 50,150, 350, 500).Then, we calculate empirical levels of significance, when $Y^2 > \chi_\varepsilon^2$(r), corresponding to theoretical levels of significance ($\varepsilon = 0.10, \varepsilon = 0.05, \varepsilon = 0.01$), We choose $r = 5$. Obtained results are given in table 2.

Table 2. Simulated levels of significance for$Y_n^2(\theta)$test for Unit Modified Burr-III distribution and their corresponding critical values($\varepsilon = 0.01$,0.05,0.10).

$N = 10{,}000$	$n_1 = 30$	$n_2 = 50$	$n_3 = 150$	$n_4 = 350$	$n_5 = 500$
$\varepsilon = 1\%$	0.0083	0.0087	0.0092	0.0098	0.0105
$\varepsilon = 5\%$	0.0388	0.0398	0.0402	0.0453	0.0499
$\varepsilon = 10\%$	0.0856	0.0891	0.0950	0.0989	0.1002

The null hypothesis H_0 for which simulated samples are fitted by Unit Modified Burr-III distribution is widely validated for the different levels of significance. Therefore, the test proposed in this work can be used in fitting this new model.

7 Applications

7.1 Case of complete data

Here, we considered two real datasets to illustrate the flexibility of the proposed distribution. The first data is related to the measures of petroleum rock. The data was presented by Cordeiro & dos Santos Brito (2012) and consist of 48 rock samples from a petroleum reservoir. According to the authors the data observations correspond to 12 core samples from petroleum reservoirs that were sampled by four cross-sections and our focus is in the shape perimeter by squared (area). The second data set consider the total milk production in the first birth of 107 cows from the SINDI race (Cordeiro & dos Santos Brito, 2012). The summary measures for both data sets are presented in Table 1.

Table3: Summary measures for both data sets.

Measures	Data I	Data II
Min.	0.09033	0.0168
Median	0.19886	0.4741
Mean	0.21811	0.4689
Variance	0.00697	0.0369
Max.	0.46413	0.8781

The TTT plots and box plots are given in Figures 6 and 7, respectively.

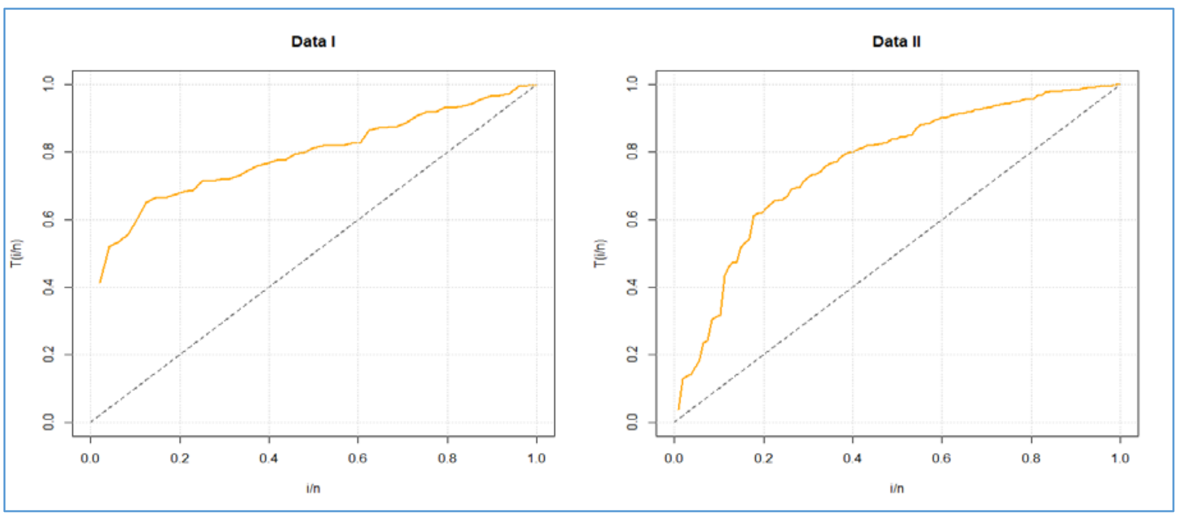

Figure 6: TTT plots for data set I and II.

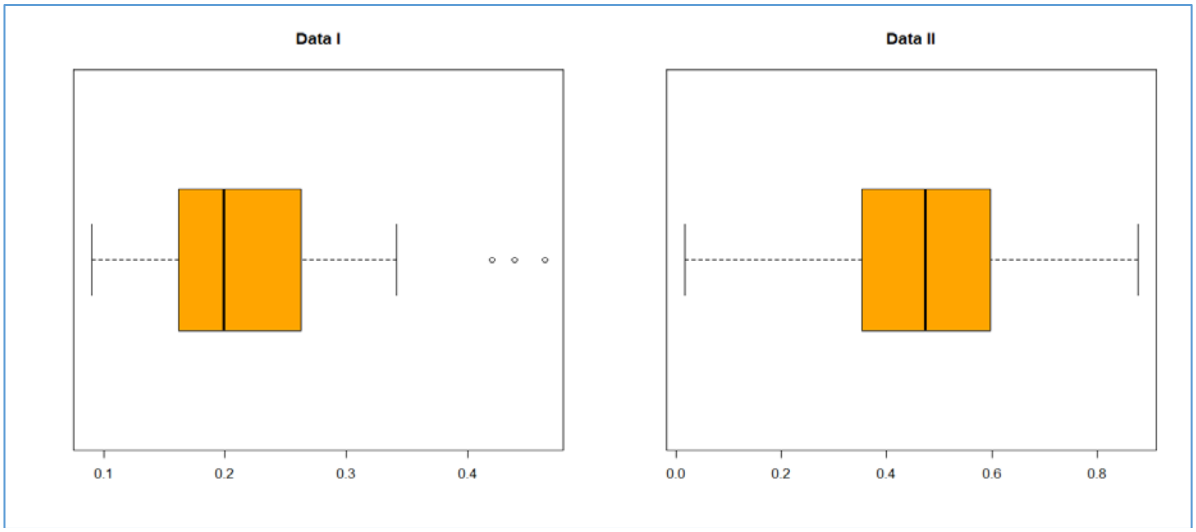

Figure 7: Boxplots for data set I and II.

Table 4: Estimates of fitted probability models data set I and II

Model	Estimates		
Data I			
UMBIII(α,β,γ)	0.012251 (0.00773)	2.978202 (0.49652)	0.003526 (0.00253)
UBIII(α,β)	0.008681 (0.00751)	86.27007 (75.6466)	-
Kw(α,β)	2.718627 (0.29347)	44.65225 (17.5699)	-
Beta(α,β)	5.941494 (1.18133)	21.20558 (4.34689)	-
Data II			
UMBIII(α,β,γ)	1.453992 (0.54590)	2.901119 (0.48203)	3.390499 (2.06368)
UBIII(α,β)	0.778232 (0.08888)	2.146778 (0.21063)	-
Kw(α,β)	2.194214 (0.22233)	3.434557 (0.58165)	-
Beta(α,β)	2.412477 (0.31449)	2.829738 (0.37443)	-

Table 5: log-likelihood, AIC, BIC, ADF, CVM, KS for data set I and II

Model	$\hat{l}$	AIC	BIC	ADF	CVM	KS
Data I						
UMBIII(α,β,γ)	58.4032	-110.806	-105.193	0.166041	0.025035	0.076755
UBIII(α,β)	26.4270	-48.8540	-45.1116	9.217786	1.888226	0.340846
Kw(α,β)	52.4915	-100.983	-97.2407	1.289244	0.205989	0.153311
Beta(α,β)	55.6002	-107.200	-103.458	0.776733	0.129985	0.142721
Data II						
UMBIII(α,β,γ)	29.3319	-52.6639	-44.6454	0.230279	0.032617	0.052212
UBIII(α,β)	26.6057	-49.2113	-43.8657	0.951273	0.161876	0.077329
Kw(α,β)	25.39467	-46.7894	-41.4437	1.003554	0.152338	0.076311
Beta(α,β)	23.77723	-43.5545	-38.2088	1.385648	0.228342	0.091009

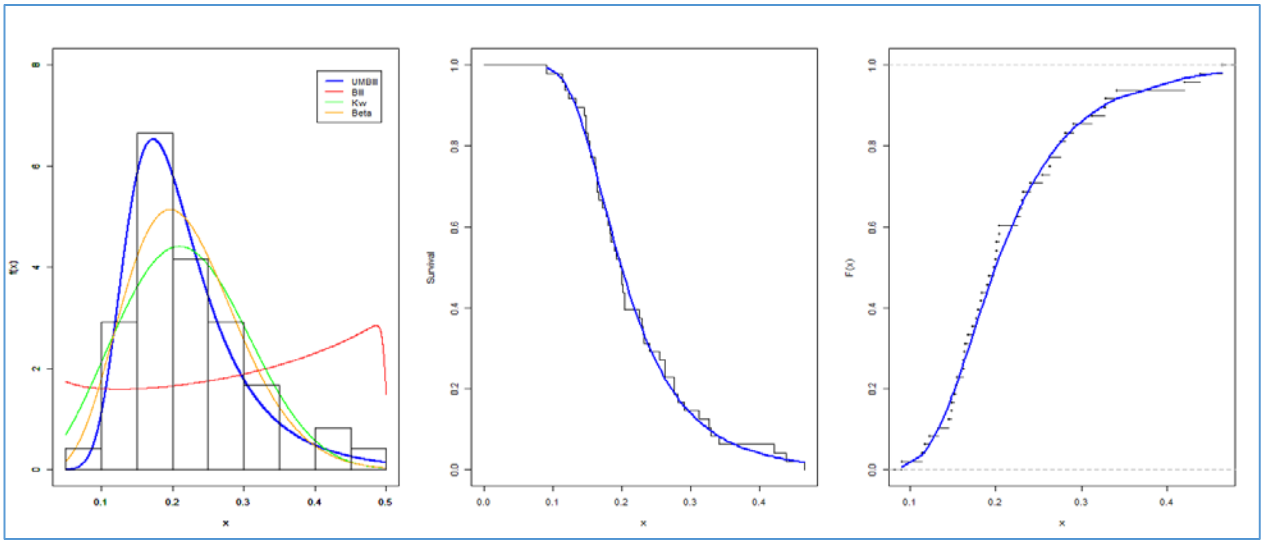

Figure 8: Fitted probability models and empirical data for data set I.

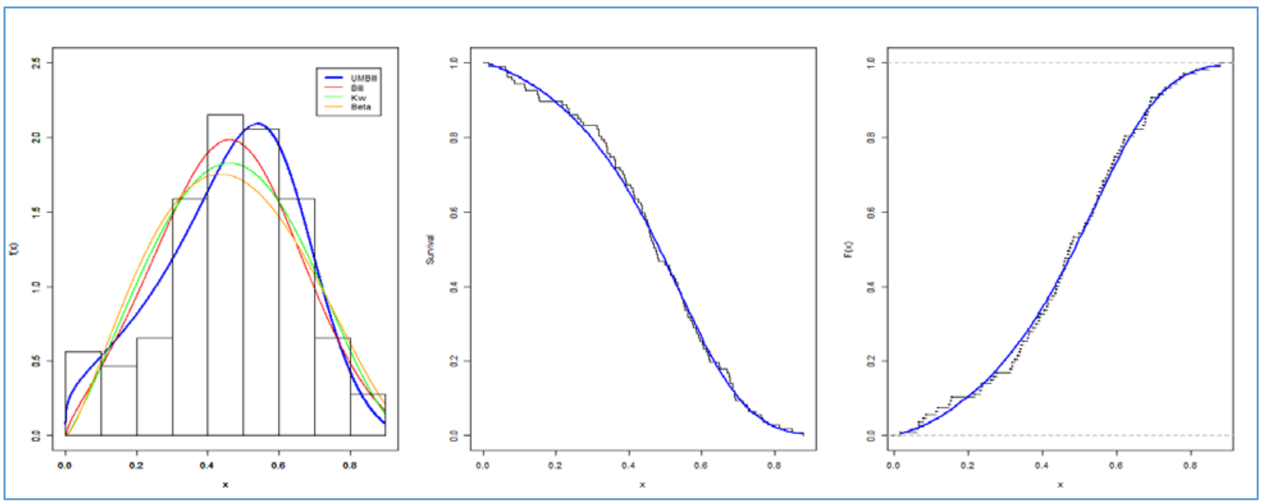

Figure 8: Fitted probability models and empirical data for data set II.

7.2 *Case of censored data*

Consider data of times to infection of kidney dialysis Patients (Nahman et al. 1992; Klein & Moeschberger, 2006):

Infection Times: 1.5, 3.5, 4.5, 4.5, 5.5, 8.5, 8.5, 9.5, 10.5, 11.5, 15.5, 16.5, 18.5, 23.5 26.5

Censored Observations: 2.5, 2.5, 3.5, 3.5, 3.5, 4.5, 5.5, 6.5, 6.5, 7.5, 7.5, 7.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 12.5, 13.5, 14.5, 14.5, 21.5, 21.5, 22.5, 22.5, 25.5, 27.5.

We make a change of variable; we divide the data by 30, to get data between 0 and 1. We use the statistic test provided above to verify if these data follow the Unit Modified Burr-III distribution, and at that end, maximum likelihood estimators are computed and the values are

$$\theta = (\alpha, \beta, \gamma)^T = (1.152\ , 1.784, 2.0154)^T.$$

Then data are grouped into $r = 5$ intervals I_j. We give the necessary calculus in Table 6.

Table 6.values of $a_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}, \hat{C}_{3j}$

a_j	0.174	0.262	0.425	0.731	0.924
U_j	10	8	11	8	6
e_j	5.4152	5.4152	5.4152	5.4152	5.4152
$\hat{C}_{1j}$	0.936	0.0836	0.7362	1.035	0.9843
$\hat{C}_{2j}$	−0.0236	0.01352	−0.9387	−1.2343	0.0214
$\hat{C}_{3j}$	0.0186	−2.3462	0.0236	−0.9363	0.0415

The statistic test value Y_n^2 is equal to

$$Y_n^2 = X^2 + Q = 4.624 + 2.936 = 7.560$$

This value $Y_n^2 = 7.560$ is less than the critical chi-square value χ_5^2= 11.0705 (for significance level ε = 0.05), so we can say that the proposed model Unit Modified Burr-IIIfit these data.

8. Conclusion

In this work, we proposed a unit distribution based on the MBIII distribution called UMBIII distribution and derived its mathematical properties. For this flexible model, we presented many important mathematical properties that allow us the application in many problems. Since the parameters are unknown, we derived different classical inferential procedures for the proposed distribution. From a simulation study, we observed that the MPS estimator returned best estimates when compared with other estimation methods. The MPS has important properties such as invariant under one-to-one transformation, asymptotic efficiency, and the consistency of the MPS holds under more general conditions than for MLEs. Additionally, we proposed a modified Chi-square goodness of fit test to verify the adequacy of the data. The tests are provided for the right-censored data, which occurs in many real problems related to survival analysis, and for complete data. We have shown the usefulness of the proposeddistribution in two real applications. Further applications of the proposed model to describe business data mining (Olson et al., 2007; Shi, 2014; Shi et al., 2011) are under investigation.

References

1. Ali, A., & Ahmad, M. (2015). The Transmuted Modified Burr III Distribution. *Journal of ISOSS*, *1*(2), 119-130.
2. Ali, A., Hasnain, S. A., & Ahmad, M. (2015). Modified Burr III Distribution, Properties and Applications. *Pak. J. Statist*, *31*(6), 697-708.
3. Bagdonavičius, V. and Nikulin, M. (2011), Chi-squared goodness-of-fit test for right-censored data, International Journal of Applied Mathematics and Statistics, vol. 24, 30-50.

4. Bagdonavičius, V., Levuliene, R., J. and Nikulin, M., (2013), Chi-squared goodness-of-fit tests for parametric accelerated failure time models, Communications in Statistics- Theory and Methods, Volume: 42 Issue: 15, 2768-2785.
5. Bhatti, F. A., Hamedani, G. G., Ali, A., & Ahmad, M. (2018). Some Characterizations of Transmuted Modified Burr III Distribution. *Asian Journal of Probability and Statistics*, 1-9.
6. Chakraborty, S., Handique, L., & Usman, R. M. (2020). A Simple Extension of Burr-III Distribution and Its Advantages over Existing Ones in Modelling Failure Time Data. Annals of Data Science, 7(1), 17-31.
7. Cordeiro, G. M., & dos Santos Brito, R. (2012). The beta power distribution. Brazilian journal of probability and statistics, 26(1), 88-112.
8. Ghitany, M. E., Mazucheli, J., Menezes, A. F. B., &Alqallaf, F. (2019). The unit-inverse Gaussian distribution: A new alternative to two-parameter distributions on the unit interval. *Communications in Statistics-Theory and Methods*, *48*(14), 3423-3438.
9. Glanzel W. A (1987), Characterization theorem based on truncated moments and its application to some distribution families.
10. Glänzel W. A. (1990), Some consequences of a characterization theorem based on truncated moments. Statistics; 21: 613-618.
11. Glanzel W, Hamedani, GG. (2001) Characterizations of univariate continuous Distributions. StudiaScientiarumMathematcarumHungarica; 37:11883.
12. Gupta, A. K., &Nadarajah, S. (2004). *Handbook of beta distribution and its applications*. CRC press.
13. Hamedani GG. (1993), Characterizations of Cauchy, normal and uniform distributions. StudiaScientiarum Mathematic arum Hungarica;28(3):243-248
14. Hamedani GG. (2002), Characterizations of univariate continuous distributions. II, StudiaScientiarum Mathematic arum Hungarica; 39:407-424.
15. Haq, M. A. U., Elgarhy, M., & Hashmi, S. (2019). The generalized odd Burr III family of distributions: properties, applications and characterizations. *Journal of Taibah University for Science*, *13*(1), 961-971.
16. John P.Klein, Melvin L. Moeschberger, Survival Analysis: Techniques for Censored and Truncated Data, Statistics for Biology and Health, 1997
17. Johnson, N. L. (1949). Bivariate distributions based on simple translation systems. *Biometrika*, *36*(3/4), 297-304.
18. Kumaraswamy, P. (1980). A generalized probability density function for double bounded random processes. Journal of Hydrology, 46, 79-88.
19. Klein, J. P., & Moeschberger, M. L. (2006). Survival analysis: techniques for censored and truncated data. Springer Science & Business Media.

20. Louzada, F., Luiz Ramos, P., & Henrique Ferreira, P. (2020). Exponential-Poisson distribution: estimation and applications to rainfall and aircraft data with zero occurrence. Communications in Statistics-Simulation and Computation, 49(4), 1024-1043.
21. Menezes, A. F. B., Mazucheli, J., & Dey, S. (2018). THE UNIT-LOGISTIC DISTRIBUTION: DIFFERENT METHODS OF ESTIMATION. PesquisaOperacional, 38(3), 555-578.
22. Mazucheli, J., Menezes, A. F., & Dey, S. (2019). Unit-Gompertz distribution with applications. Statistica, 79(1), 25-43.
23. Mazucheli, J., Menezes, A. F., & Dey, S. (2018). The unit-Birnbaum-Saunders distribution with applications. *Chilean Journal of Statistics*, *9*(1), 47-57.
24. Mazucheli, J., Menezes, A. F. B., &Ghitany, M. E. (2018). The unit-Weibull distribution and associated inference. *J. Appl. Probab. Stat*, *13*, 1-22.
25. Mukhtar, S., Ali, A., & Alya, A. M. (2019). Mc-Donald modified Burr-III distribution: properties and applications. *Journal of Taibah University for Science*, *13*(1), 184-192.
26. Olson, D. L., Shi, Y., & Shi, Y. (2007). Introduction to business data mining (Vol. 10, pp. 2250-2254). Englewood Cliffs: McGraw-Hill/Irwin.
27. Ramos, P. L., Louzada, F., Ramos, E., & Dey, S. (2019). The Fréchet distribution: Estimation and application-An overview. Journal of Statistics and Management Systems, 1-30.
28. Ravi, V. and Gilbert, P. D. (2009), BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function, J. Statist. Software, 32(4).
29. Rennolls, K., & Wang, M. (2005). A new parameterization of Johnson's SB distribution with application to fitting forest tree diameter data. Canadian journal of forest research, 35(3), 575-579.
30. Shi, Y. (2014). Big data: History, current status, and challenges going forward. The Bridge, The US National Academy of Engineering, 44(4), 6-11.
31. Shi, Y., Tian, Y., Kou, G., Peng, Y., & Li, J. (2011). Optimization based data mining: theory and applications. Springer Science & Business Media.
32. Tadikamalla, P. R., & Johnson, N. L. (1982). Systems of frequency curves generated by transformations of logistic variables. Biometrika, 69(2), 461-465.
33. Topp, C. W., & Leone, F. C. (1955). A family of J-shaped frequency functions. Journal of the American Statistical Association, 50(269), 209-219.
34. Usman, R. M., &Haq, M. A. (2019). Some remarks on odd Burr III Weibull distribution. Annals of Data Science, 6(1), 21-38.