STATISTICA, anno LXXIX, n. 3, 2019

COMPARISONS OF METHODS OF ESTIMATION FOR A NEW PARETO-TYPE DISTRIBUTION

Ali Saadati Nik
Department of Statistics, University of Mazandaran, Babolsar, Iran
Akbar Asgharzadeh [1]
Department of Statistics, University of Mazandaran, Babolsar, Iran
Saralees Nadarajah
Department of Mathematics, University of Manchester, Manchester, UK

1. Introduction

The Pareto distribution was first proposed by Pareto (1964) as a model for the distribution of income. This distribution is used to describe the allocation of wealth among individuals in many societies. This distribution is now applied in different fields such us insurance, business, economics, engineering, physics, hydrology, geology and reliability. In hydrology, the Pareto distribution is applied to extreme events such as annual maximum one-day rainfalls and river discharges. Some authors discussed the applications of the Pareto distribution in physics. Newman (2005) provided many quantities measured in physical systems where the Pareto distribution has applications. Zaninetti and Ferraro (2008) provided an application of the Pareto distribution to astrophysics and more precisely to the statistical analysis of masses of stars and of diameters of asteroids. For various applications of the Pareto distribution, one could refer to Arnold (1983), Johnson *et al.* (1994) and Dagum (2006).

The new Pareto-type (NP) distribution was recently proposed by Bourguignon *et al.* (2016) to model income and reliability data. This distribution is a generalization of the well-known Pareto distribution. The two-parameter NP distribution (denoted by NP(α,β)) has the probability density function (PDF)

$$f(x;\alpha,\beta)=\frac{2\alpha\,(\beta/x)^{\alpha+1}}{\beta[1+(\beta/x)^{\alpha}]^{2}},\quad x\geq\beta,\ \alpha>0,\ \beta>0, \tag{1}$$

[1] Corresponding Author. E-mail: a.asgharzadeh@umz.ac.ir

where α and β are shape and scale parameters, respectively. The cumulative distribution function (CDF) of the NP distribution is

$$F(x;\alpha,\beta)=1-\frac{2\beta^{\alpha}}{x^{\alpha}+\beta^{\alpha}}, \quad x\geq\beta,\ \alpha>0,\ \beta>0. \tag{2}$$

As mentioned by Bourguignon *et al.* (2016), for high incomes, the NP CDF closely approximates the form

$$F(x;\alpha,\beta)=1-Ax^{-\alpha},$$

which is the form predicted by Pareto's law. Therefore, the NP distribution converges in distribution to the Pareto distribution for x sufficiently large.

The NP PDF is decreasing. So, similarly to the Pareto distribution, the NP distribution can be used as a model for the distribution of income. The income distributions with decreasing PDFs show that the "probability" or fraction of the population that owns a small amount of wealth per person is rather high, and then decreases steadily as wealth increases, see e.g., Sankaran *et al.* (2014). The hazard rate of the NP distribution can be upside-down bathtub (unimodal) shaped or decreasing depending on the values of its parameters. Decreasing and unimodal hazard rates have many applications in reliability and survival analysis. A decreasing failure rate describes a phenomenon where the probability of an event in a fixed time interval in the future decreases over time. A practical example is infant mortality where earlier failures are eliminated or corrected. A unimodal hazard rate function is used to model a failure rate that has a relatively high rate of failure in the middle of expected life time. When failures of products are caused by fatigue and corrosion, the corresponding failure rates exhibit unimodal shapes (Lai and Xie, 2006). Also, in some medical situations, for example breast cancer and infection with some new viruses, the hazard rate is unimodal shaped, e.g., Demicheli *et al.* (2004) and Abdi *et al.* (2019). Bourguignon *et al.* (2016) studied mathematical properties of the NP distribution and showed the usefulness of this distribution for modeling income and reliability data by analyzing seven real data sets.

The aim of this paper is to consider different estimation methods for estimating the unknown parameters of the NP distribution from both frequentist and Bayesian points of view. We first compute the maximum likelihood estimates (MLEs) which are the most natural frequentist estimates. We then discuss the existence and uniqueness of the MLEs. We also consider other frequentist estimates including the method of moment estimates (MMEs), percentile estimates, least square estimates (LSEs), weighted least square estimates (WLSEs) and maximum product of spacing (MPS) estimates. We further consider the Bayes estimates of the unknown parameters under the squared error loss (SEL) function. Since the Bayes estimates can not be obtained in closed forms, an importance sampling method is used to compute the Bayes estimates and the associate credible intervals. Finally, we compare the performance of the different estimates using extensive computer simulations. Comparisons of estimation methods for other statistical distributions have been discussed in the literature, see e.g., Gupta and Kundu

(2001), Kundu and Raqab (2005), Alkasabeh and Raqab (2009), Asgharzadeh *et al.* (2011) and Dey *et al.* (2014).

The paper is organized as follows. In Section 2, we provide the MLEs. We also discuss in this section the conditions for existence and uniqueness of the MLEs. Other estimation methods are presented in Sections 3-7. The interval estimates of α for known β are described in Section 8. A Monte Carlo simulation study is used to compare the performance of the different estimates in Section 9. Some numerical examples are given in Section 10 to illustrate different methods of estimation discussed in this paper. The paper is concluded in Section 11.

2. Maximum likelihood estimates

In this section, the MLEs of α and β of the NP(α,β) distribution are considered. If $x_1,\ldots,x_n$ is an observed random sample from NP(α,β), then the likelihood function is

$$L(\alpha,\beta)=\prod_{i=1}^{n} f(x_i,\alpha,\beta)=(\frac{2\alpha}{\beta})^n\prod_{i=1}^{n}\frac{(\beta/x_i)^{\alpha+1}}{(1+(\beta/x_i)^{\alpha})^2},\ \beta\le x_{(1)}, \tag{3}$$

where $x_{(1)}=\min(x_1,\ldots,x_n)$. The log-likelihood function for $\beta\le x_{(1)}$ is

$$\begin{aligned} l(\alpha,\beta)=\log L \;&=\; n\log(2\alpha)-n\log(\beta)+(\alpha+1)\sum_{i=1}^{n}\log(\frac{\beta}{x_i}) \\ &\quad -2\sum_{i=1}^{n}\log\left[1+(\frac{\beta}{x_i})^{\alpha}\right]. \end{aligned} \tag{4}$$

The MLEs of the unknown parameters are obtained by maximizing the log-likelihood function in (4) with respect to α and β. It can be seen that $l(\alpha,\beta)$ is monotonically increasing with β, so the MLE of β is $\widehat{\beta}=x_{(1)}$. Substituting $\widehat{\beta}$ in (4), we obtain the profile log-likelihood function of α without the additive constant as

$$\begin{aligned} g(\alpha)=l(\alpha,x_{(1)}) \;&=\; n\log(2\alpha)-n\log(x_{(1)})+(\alpha+1)\sum_{i=1}^{n}\log(\frac{x_{(1)}}{x_i}) \\ &\quad -2\sum_{i=1}^{n}\log\left[1+(\frac{x_{(1)}}{x_i})^{\alpha}\right]. \end{aligned} \tag{5}$$

Therefore, the MLE of α, say $\widehat{\alpha}$, can be obtained by maximizing (5) with respect to α. Consequently, the MLE $\widehat{\alpha}$ of α is obtained as the solution to the following equation

$$h(\alpha)=\frac{\partial l(\alpha,x_{(1)})}{\partial\alpha}=-2\sum_{i=1}^{n}\frac{(\frac{x_{(1)}}{x_i})^{\alpha}\ \log\left(\frac{x_{(1)}}{x_i}\right)}{1+(\frac{x_{(1)}}{x_i})^{\alpha}}+\sum_{i=1}^{n}\log\left(\frac{x_{(1)}}{x_i}\right)+\frac{n}{\alpha}=0. \tag{6}$$

There is no closed-form expression for the MLE $\widehat{\alpha}$ and its computation has to be performed numerically using a nonlinear optimization algorithm. Some iterative methods can be applied to solve the likelihood equation and compute the estimate $\widehat{\alpha}$. It can be shown that the MLE $\widehat{\alpha}$ can be derived as a fixed-point solution of the equation $H(\alpha) = \alpha$, where

$$H(\alpha) = \frac{n}{2\sum_{i=1}^{n} \frac{(\frac{x_{(1)}}{x_i})^{\alpha} \log\left(\frac{x_{(1)}}{x_i}\right)}{1+(\frac{x_{(1)}}{x_i})^{\alpha}} - \sum_{i=1}^{n} \log\left(\frac{x_{(1)}}{x_i}\right)}.$$

Note that

$$\lim_{\alpha\to 0} h(\alpha) = \infty, \lim_{\alpha\to\infty} h(\alpha) = \sum_{i=1}^{n} \log\left(\frac{x_{(1)}}{x_i}\right) < 0$$

and

$$h'(\alpha) = \frac{\partial^2 l(\alpha, x_{(1)})}{\partial^2 \alpha} = -2\sum_{i=1}^{n} \frac{(\frac{x_{(1)}}{x_i})^{\alpha} \log^2\left(\frac{x_{(1)}}{x_i}\right)}{\left(1+(\frac{x_{(1)}}{x_i})^{\alpha}\right)^2} - \frac{n}{\alpha^2} < 0.$$

Therefore, $h(\alpha)$ is a continuous function on $(0,\infty)$ which decreases monotonically from $+\infty$ to negative values. Therefore, the MLE of α which is a solution to $h(\alpha) = 0$, exists and is unique.

Let us now consider the MLE of α when the scale parameter β is known. Without loss of generality, we can assume that $\beta = 1$. With $\beta = 1$, the log-likelihood function becomes

$$l(\alpha) = n\log(2\alpha) - (\alpha+1)\sum_{i=1}^{n} \log(x_i) - 2\sum_{i=1}^{n} \log\left[1 + (\frac{1}{x_i})^{\alpha}\right]. \tag{7}$$

The MLE of α can be obtained directly by maximizing the log-likelihood function in (7) with respect to α, or can be obtained as the solution to the following equation

$$h(\alpha) = \frac{\partial l(\alpha)}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \log(x_i) + 2\sum_{i=1}^{n} \frac{\log(x_i)}{(1+x_i^{\alpha})^2} = 0. \tag{8}$$

It can be shown that the MLE of α can be obtained as a fixed solution of $\alpha = H(\alpha)$, where

$$H(\alpha) = \frac{n}{\sum_{i=1}^{n} \log(x_i) - 2\sum_{i=1}^{n} \frac{\log(x_i)}{(1+x_i^{\alpha})^2}}.$$

Again, note that since $\lim_{\alpha\to 0} h(\alpha) = \infty$, $\lim_{\alpha\to\infty} h(\alpha) = -\sum_{i=1}^{n} \log x_i < 0$ and

$$h'(\alpha) = \frac{\partial^2 l(\alpha)}{\partial^2 \alpha} = -2\sum_{i=1}^{n} \frac{x_i^\alpha \log^2(x_i)}{(1+x_i^\alpha)^2} - \frac{n}{\alpha^2} < 0,$$

the MLE of α which is a solution to $h(\alpha) = 0$, exists and is unique.

3. METHOD OF MOMENT ESTIMATES

Here we obtain the MMEs of α and β of the NP(α, β) distribution. If the random variable X has the NP(α, β) distribution, then the r-th moment of X is given by

$$\begin{aligned} E(X^r) &= 2\alpha\, \beta^\alpha \int_\beta^\infty \frac{x^{r+\alpha-1}}{(x^\alpha + \beta^\alpha)^2} \mathrm{d}x, \\ &= 2\alpha\, \beta^r \int_0^1 y^{\alpha-r-1}\, (1+y^\alpha)^{-2} \mathrm{d}y \\ &= 2\alpha\beta^r\, J_r(\alpha), \qquad r < \alpha, \end{aligned}$$

where

$$J_r(\alpha) = \int_0^1 y^{\alpha-r-1}\, (1+y^\alpha)^{-2} \mathrm{d}y.$$

The above integral can be computed numerically in software such as MAPLE, MATHEMATICA and R.

Note that the moments of X can be obtained as a series too. By using the negative binomial expansion,

$$(1+x)^{-2} = \sum_{j=1}^{\infty} j\, (-1)^{j-1}\, x^{j-1}, \qquad |x| < 1,$$

we can write

$$\begin{aligned} E(X^r) &= 2\alpha\, \beta^r \int_0^1 y^{\alpha-r-1}\, (1+y^\alpha)^{-2} \mathrm{d}y \\ &= 2\alpha\, \beta^r \sum_{j=1}^{\infty} j\, (-1)^{j-1} \left(\int_0^1 y^{\alpha\, j-r-1} \mathrm{d}y \right) \\ &= 2\alpha\, \beta^r \sum_{j=1}^{\infty} \frac{j\, (-1)^{j-1}}{\alpha\, j - r}, \qquad r < \alpha\, . \end{aligned}$$

Therefore, the first and second moments of X are

$$E(X) = 2\alpha\,\beta \sum_{j=1}^{\infty} \frac{j\,(-1)^{j-1}}{\alpha\,j-1}, \qquad \alpha > 1,$$

$$E(X^2) = 2\alpha\,\beta^2 \sum_{j=1}^{\infty} \frac{j\,(-1)^{j-1}}{\alpha\,j-2}, \qquad \alpha > 2.$$

In Table 1, we have presented the first and second moments of the NP distribution for some selected values of the shape parameter α, when $\beta = 1$. These values have been computed numerically using R.

TABLE 1
First and second moments of the standard NP distribution for different shape parameter α.

α	0.5	1	1.5	2	2.5	3	3.5	4
E(X)	-	-	4.342	2.570	2.015	1.747	1.590	1.487
E(X^2)	-	-	-	-	8.106	4.342	3.147	2.570

Now, to obtain the MMEs of the unknown parameters α and β, we need to equate the sample moments with the population moments and solve the following equations:

$$2\alpha\,\beta\,A_1(\alpha) = \overline{x}, \tag{9}$$

and

$$2\alpha\,\beta^2\,A_2(\alpha) = \overline{x^2}, \tag{10}$$

where $A_r(\alpha) = \sum_{j=1}^{\infty} \frac{j\,(-1)^{j-1}}{\alpha\,j-r}$, $\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ and $\overline{x^2} = \frac{1}{n}\sum_{j=1}^{n} x_j^2$. Therefore, the MMEs of α and β are the simultaneous solutions of the two equations (9) and (10). From (9) and (10), we obtain

$$\beta = \frac{\overline{x}}{2\alpha\,A_1(\alpha)}$$

and

$$\alpha = \frac{\overline{x}^2}{\overline{x^2}}\,\frac{A_2(\alpha)}{A_1(\alpha)}.$$

Therefore, the MME of α, say $\widehat{\alpha}_{MME}$, can be obtained by solving the equation $\alpha = \frac{\overline{x}^2}{\overline{x^2}}\frac{A_2(\alpha)}{A_1(\alpha)}$ with respect to α, numerically. Once $\widehat{\alpha}_{MME}$ is obtained, the MME of β can be obtained easily as

$$\widehat{\beta}_{MME} = \frac{\overline{x}}{2\alpha\,A_1(\widehat{\alpha}_{MME})}.$$

Note that the MMEs exist only, when $\alpha > 2$.

If the scale parameter β is known, we take $\beta = 1$. In this case, the MME of α can be obtained by solving the non-linear equation $2\alpha\,A_1(\alpha) = \overline{x}$ by some iterative methods.

4. ESTIMATES BASED ON PERCENTILES

In this section, we estimate the unknown parameters by the percentile method. In the percentile method, the unknown parameters are estimated by equating the sample percentile points with the population percentile points. Kao (1959a,b) proposed this method when the CDF is in a closed form. Some authors have used this method of estimation, see for example Mann *et al.* (1974), Gupta and Kundu (2001), Kundu and Raqab (2005) and Alkasabeh and Raqab (2009).

Here, we apply this method of estimation for the NP distribution. Since the CDF of the NP distribution can be written in the closed form

$$F(x;\alpha,\beta)=1-\frac{2\,\beta^{\alpha}}{x^{\alpha}+\beta^{\alpha}},$$

we obtain

$$x=\beta\left(\frac{1+F(x;\alpha,\beta)}{1-F(x;\alpha,\beta)}\right)^{1/\alpha}. \tag{11}$$

If $x_{1:n}<\cdots<x_{n:n}$ is the sample order statistics and p_i denotes some estimate of $F(x_{i:n};\alpha,\beta)$, then the Euclidean distance between the sample percentile and population percentile is

$$E(\alpha,\beta)=\sum_{i=1}^{n}\left[x_{i:n}-\beta\left(\frac{1+p_i}{1-p_i}\right)^{1/\alpha}\right]^2. \tag{12}$$

The PCEs of α and β are obtained by minimizing the Euclidean distance $E(\alpha,\beta)$ with respect to α and β. In this paper, we have used $p_i=\frac{i}{n+1}$ which is the unbiased estimate of $F(x_{i:n};\alpha,\beta)$. Therefore, the PCEs of α and β, say $\widehat{\alpha}_{PCE}$ and $\widehat{\beta}_{PCE}$, can be obtained as the solution of the following equations

$$\frac{\partial E(\alpha,\beta)}{\partial\alpha}=\frac{2\beta}{\alpha^2}\sum_{i=1}^{n}\ln\left(\frac{1+p_i}{1-p_i}\right)\left(\frac{1+p_i}{1-p_i}\right)^{1/\alpha}\left[x_{i:n}-\beta\left(\frac{1+p_i}{1-p_i}\right)^{1/\alpha}\right]=0 \tag{13}$$

and

$$\frac{\partial E(\alpha,\beta)}{\partial\beta}=-2\sum_{i=1}^{n}\left(\frac{1+p_i}{1-p_i}\right)^{1/\alpha}\left[x_{i:n}-\beta\left(\frac{1+p_i}{1-p_i}\right)^{1/\alpha}\right]=0. \tag{14}$$

From (14), we obtain the PCE of β as a function of α as

$$\widehat{\beta}(\alpha)=\frac{\sum_{i=1}^{n}\left(\frac{1+p_i}{1-p_i}\right)^{1/\alpha}x_{i:n}}{\sum_{i=1}^{n}\left(\frac{1+p_i}{1-p_i}\right)^{2/\alpha}}.$$

Putting the value of $\widehat{\beta}(\alpha)$ in (13), $\widehat{\alpha}$ can be obtained as a solution of the following equation

$$\begin{aligned} h(\alpha) &= \sum_{i=1}^{n} \ln(\frac{1+p_i}{1-p_i})\,(\frac{1+p_i}{1-p_i})^{1/\alpha}\, x_{i:n} \\ &\quad - \frac{\left[\sum_{i=1}^{n}(\frac{1+p_i}{1-p_i})^{1/\alpha}\, x_{i:n}\right]\left[\sum_{i=1}^{n}\ln(\frac{1+p_i}{1-p_i})\,(\frac{1+p_i}{1-p_i})^{2/\alpha}\right]}{\sum_{i=1}^{n}(\frac{1+p_i}{1-p_i})^{2/\alpha}} = 0. \end{aligned}$$

Therefore, the PCE of α, say $\widehat{\alpha}_{PCE}$, is derived by solving the equation $h(\alpha)=0$. Once, $\widehat{\alpha}_{PCE}$ is derived, the PCE of β can be obtained as $\widehat{\beta}_{PCE}=\widehat{\beta}(\widehat{\alpha}_{PCE})$.

For known β, we assume $\beta=1$. The PCE of α is obtained by minimizing

$$E(\alpha)=\sum_{i=1}^{n}\left[x_{i:n}-\left(\frac{1+p_i}{1-p_i}\right)^{1/\alpha}\right]^2 \tag{15}$$

with respect to α, where $p_i=\frac{i}{n+1}$. Alternatively, the PCE of α can be obtained by solving the equation

$$h(\alpha)=\frac{\partial E(\alpha)}{\partial \alpha}=\frac{2}{\alpha^2}\sum_{i=1}^{n}\ln(\frac{1+p_i}{1-p_i})\,(\frac{1+p_i}{1-p_i})^{1/\alpha}\left[x_{i:n}-(\frac{1+p_i}{1-p_i})^{1/\alpha}\right]=0. \tag{16}$$

5. Least squares and weighted least squares estimates

The LSEs and WLSEs are used generally for estimation of parameters in linear models. These estimates were used by Swain *et al.* (1988) to estimate the parameters of a beta distribution. Recently, some authors have used the method of estimation in their work. See, for example, Gupta and Kundu (2001), Kundu and Raqab (2005), Alkasabeh and Raqab (2009) and Bakouch *et al.* (2017).

Let $x_{1:n}\leq\cdots\leq x_{n:n}$ be order statistics from a random sample of size n from a CDF $G(\cdot)$. Since $G(X_{j:n})$ behaves like the jth order statistic of a sample of size n from $U(0,1)$, we have

$$E[G(X_{j:n})]=\frac{j}{n+1}, \qquad \mathrm{Var}[G(X_{j:n})]=\frac{j(n-j+1)}{(n+1)^2(n+2)}.$$

The LSEs are obtained by minimizing

$$\sum_{j=1}^{n}\left[G(X_{j:n})-\frac{j}{n+1}\right]^2 \tag{17}$$

with respect to the unknown parameters of $G(\cdot)$. In case of the NP distribution, the

LSEs of α and β, say $\widehat{\alpha}_{LSE}$ and $\widehat{\beta}_{LSE}$, can be obtained by minimizing

$$\sum_{j=1}^{n}\left[1-\frac{2\,\beta^{\alpha}}{x_{j:n}^{\alpha}+\beta^{\alpha}}-\frac{j}{n+1}\right]^{2} \tag{18}$$

with respect to α and β.

The WLSEs of the unknown parameters can be derived by minimizing

$$\sum_{j=1}^{n} w_j\left[G(X_{j:n})-\frac{j}{n+1}\right]^{2} \tag{19}$$

with respect to the unknown parameters, where

$$w_j=\frac{1}{\mathrm{Var}[G(X_{j:n})]}=\frac{(n+1)^2(n+2)}{j(n-j+1)}.$$

In case of the NP distribution, the WLSEs of α and β, say $\widehat{\alpha}_{WLSE}$ and $\widehat{\beta}_{WLSE}$, can be obtained by minimizing

$$\sum_{j=1}^{n} w_j\left[1-\frac{2\,\beta^{\alpha}}{x_{j:n}^{\alpha}+\beta^{\alpha}}-\frac{j}{n+1}\right]^{2} \tag{20}$$

with respect to α and β.

For known β, let us fix $\beta=1$. The LSE of α can be obtained by minimizing

$$\sum_{j=1}^{n}\left[1-\frac{2}{x_{j:n}^{\alpha}+1}-\frac{j}{n+1}\right]^{2} \tag{21}$$

with respect to α. On the other hand, the WLSE of α can be obtained by minimizing

$$\sum_{j=1}^{n} w_j\left[1-\frac{2}{x_{j:n}^{\alpha}+1}-\frac{j}{n+1}\right]^{2} \tag{22}$$

with respect to α.

6. Method of maximum product of spacings

Cheng and Amin (1979, 1983) introduced the MPS method as an alternative to MLE for estimating the parameters of continuous distributions. They showed that the MPS method provides consistent and asymptotically efficient estimators in both the situations whether MLE exists or not. The MPS method of estimation was also developed by Ranneby (1984) using the Kullback-Leibler measure of information.

Let $x_{1:n} \leq \cdots \leq x_{n:n}$ be order statistics from a random sample of size n from a CDF $G(x;\alpha,\beta)$, where α and β are unknown parameters. The ith spacing $D_i(\alpha,\beta)$ is defined as

$$D_i(\alpha,\beta) = G(x_{i:n};\alpha,\beta) - G(x_{i-1:n};\alpha,\beta),\ i = 1,\ldots,n+1,$$

where $G(x_{0:n};\alpha,\beta) = 0$ and $G(x_{n+1:n};\alpha,\beta) = 1$. Clearly $\sum_{i=1}^{n+1} D_i(x;\alpha,\beta) = 1$.

The MPS estimators $\widehat{\alpha}_{MPS}$ and $\widehat{\beta}_{MPS}$ of the parameters α and β are obtained by maximizing the geometric mean of the spacings, i.e.,

$$G(\alpha,\beta) = \left[\prod_{i=1}^{n+1} D_i(\alpha,\beta)\right]^{\frac{1}{n+1}} \tag{23}$$

with respect to α and β or, equivalently, by maximizing the function

$$H(\alpha,\beta) = \frac{1}{n+1}\sum_{i=1}^{n+1} \log D_i(\alpha,\beta) \tag{24}$$

with respect to α and β.

In case of the NP distribution, the MPSs of α and β can be obtained by maximizing

$$H(\alpha,\beta) = \frac{1}{n+1}\sum_{i=1}^{n+1} \log\left[\frac{2\,\beta^{\alpha}}{x_{i-1:n}^{\alpha} + \beta^{\alpha}} - \frac{2\,\beta^{\alpha}}{x_{i:n}^{\alpha} + \beta^{\alpha}}\right] \tag{25}$$

with respect to α and β. If the scale parameter β is known and $\beta = 1$, then the MPS of α can be obtained by maximizing

$$H(\alpha) = \frac{1}{n+1}\sum_{i=1}^{n+1} \log\left[\frac{2}{x_{i-1:n}^{\alpha} + 1} - \frac{2}{x_{i:n}^{\alpha} + 1}\right] \tag{26}$$

with respect to α.

7. Bayes estimates and credible intervals

In this section, Bayesian inference of the unknown parameters of the NP(α,β) distribution is considered when both the parameters α and β are unknown. We obtain the Bayes estimates and the associated credible intervals. We consider the following joint prior PDF

$$\pi(\alpha,\beta) \propto \alpha^{\gamma}\, \beta^{\alpha b-1}\, c^{-\alpha}, \quad \alpha > 0,\ 0 < \beta < d \tag{27}$$

for α and β, where γ, b, c, d are positive constants and $d^b < c$. This prior was first proposed by Lwin (1972) and later generalized by Arnold and Press (1983, 1989). Such a prior specifies $\pi(\alpha)$ as a gamma distribution with parameters γ and $\log c - b \log d$ and $\pi(\beta|\alpha)$ as a power function distribution of the form

$$\pi(\beta|\alpha) \propto b \; \alpha \; \beta^{b\alpha-1} \, d^{-b\alpha}, \quad 0 < \beta < d.$$

Note that the noninformative prior

$$\pi(\alpha,\beta) \propto \frac{1}{\alpha\beta}, \quad \alpha > 0, \; \beta > 0$$

is specified by letting $\gamma = -1$, $c = 1$, $b = 0$ and $d \to \infty$.

Based on the observed sample, the joint posterior PDF of α and β becomes

$$\pi(\alpha,\beta|\underline{x}) = \frac{1}{R(\underline{x})} \alpha^{n+\gamma} \; \beta^{\alpha(n+b)-1} \; c^{-\alpha} \prod_{i=1}^{n} \frac{x_i^{\alpha-1}}{(x_i^\alpha + \beta^\alpha)^2}, \quad \alpha > 0, \; 0 < \beta < M, \tag{28}$$

where $M = \min(d, x_{(1)})$ and

$$R(\underline{x}) = \int_0^\infty \int_0^M \alpha^{n+\gamma} \; \beta^{\alpha(n+b)-1} \; c^{-\alpha} \prod_{i=1}^{n} \frac{x_i^{\alpha-1}}{(x_i^\alpha + \beta^\alpha)^2} \mathrm{d}\beta \; \mathrm{d}\alpha.$$

Therefore, the Bayes estimate of any function of α and β, say $\theta(\alpha,\beta)$ under the SEL function is

$$\widehat{\theta}_{\text{Bayes}} = E[\theta(\alpha,\beta)|\underline{x}] = \frac{1}{R(\underline{x})} \int_0^\infty \int_0^M \theta(\alpha,\beta) \; \alpha^{n+\gamma} \; \beta^{\alpha(n+b)-1} \; c^{-\alpha} \prod_{i=1}^{n} \frac{x_i^{\alpha-1}}{(x_i^\alpha + \beta^\alpha)^2} \mathrm{d}\beta \; \mathrm{d}\alpha. \tag{29}$$

Clearly, the Bayes estimates of α and β can not be obtained in explicit forms. Here, we use an importance sampling method to compute the Bayes estimate and also to compute the associated credible interval. To implement the importance sampling method, we rewrite the posterior distribution (28) as

$$\begin{aligned} \pi(\alpha,\beta|\underline{x}) \quad & \propto \; \alpha^{n+\gamma-1} \; e^{-\alpha \sum_{i=1}^{n} \log x_i} \; \alpha(n+b) \; \beta^{\alpha(n+b)-1} \; M^{-\alpha(n+b)} \; c^{-\alpha} M^{\alpha(n+b)} \\ & \times \prod_{i=1}^{n} \frac{x_i^{2\alpha-1}}{(x_i^\alpha + \beta^\alpha)^2}. \end{aligned}$$

Therefore, we have

$$\pi(\alpha,\beta|\underline{x}) \propto G_\alpha(n+\gamma, \sum_{i=1}^{n} \log x_i) \; PF_{\beta|\alpha}\Big(\alpha(n+b), M\Big) \; h(\alpha,\beta,\underline{x}),$$

where $G(n+\gamma, \sum_{i=1}^{n} \log x_i)$ is a gamma PDF with parameters $n+\gamma$ and $\sum_{i=1}^{n} \log x_i$ and $PF(\beta, \alpha(n+b), M)$ is the power function distribution with the CDF

$$F(\beta)=\left(\frac{\beta}{M}\right)^{\alpha(n+b)}, \quad 0<\beta<M,$$

also

$$h(\alpha,\beta,\underline{x})=c^{-\alpha}\ M^{\alpha(n+b)}\prod_{i=1}^{n}\frac{x_i^{2\alpha-1}}{(x_i^{\alpha}+\beta^{\alpha})^2}.$$

Now, we use the following algorithm to generate the samples from the posterior distribution $\pi(\alpha,\beta|\underline{x})$ and also to compute the Bayes estimates:

Step 1. Generate (α_1,β_1) as: $\alpha_1\sim G(n+\gamma,\sum_{i=1}^{n}\log x_i)$ and $\beta_1|\alpha_1\sim PF(\alpha_1(n+b),M)$.

Step 2. Repeat Step 1 N times and obtain $(\alpha_2,\beta_2),\ldots,(\alpha_N,\beta_N)$.

Step 3. Compute $h(\alpha_i,\ \beta_i,\underline{x})$; $i=1,\ldots,N$.

Step 4. Obtain the approximate Bayes estimates of α and β under the SEL function as

$$\widehat{\alpha}_{BS}\approx\frac{\sum_{i=1}^{N}\alpha_i\ h(\alpha_i,\beta_i,\underline{x})}{\sum_{i=1}^{N}h(\alpha_i,\beta_i,\underline{x})} \tag{30}$$

and

$$\widehat{\beta}_{BS}\approx\frac{\sum_{i=1}^{N}\beta_i\ h(\alpha_i,\beta_i,\underline{x})}{\sum_{i=1}^{N}h(\alpha_i,\beta_i,\underline{x})}, \tag{31}$$

respectively.

Next, we obtain the credible intervals of α and β using the results in Chen and Shao (1999). Let $\pi(\alpha,\beta|x)$ and $\Pi(\alpha,\beta|x)$ be the posterior PDF and posterior CDF of (α,β), respectively, and let $\alpha^{(\mu)}$, be the μth quantile of α, i.e,

$$\alpha^{(\mu)}=\inf\{\alpha:\Pi(\alpha,\beta|x)\geq\mu\}, \quad 0<\mu<1. \tag{32}$$

For a given α^*, we have $\Pi(\alpha^*,\beta|x)=E\{I_{\alpha\leq\alpha^*}(\alpha,\beta)|x\}$, where I_A denotes the indicator function such that $I_A(\alpha)=1$ if A is true and $I_A(\alpha)=0$ otherwise. Therefore, a simulation consistent of $\Pi(\alpha^*,\beta|x)$ is

$$\widehat{\Pi}(\alpha^*,\beta|\underline{x})=\frac{\frac{1}{N}\sum_{i=1}^{n}I_{\alpha_i\leq\alpha^*}(\alpha,\beta)\ h(\alpha_i,\beta_i,\underline{x})}{\frac{1}{N}\sum_{i=1}^{N}h(\alpha_i,\beta_i,\underline{x})}. \tag{33}$$

Let $(\alpha_{(i)},\beta_{(i)})$ for $i=1,\ldots,N$ be the ordered values of (α_i,β_i), and

$$w_i=\frac{h(\alpha_{(i)},\beta_{(i)},\underline{x})}{\sum_{i=1}^{N}h(\alpha_{(i)},\beta_{(i)},\underline{x})}$$

be the associated weight, then we have

$$\hat{\Pi}(\alpha^*, \beta|\underline{x}) = \begin{cases} 0, & if\ \alpha^* < \alpha_{(1)}, \\ \sum_{j=1}^{i} w_j, & if\ \ \alpha_{(i)} \leq \alpha^* < \alpha_{(i+1)} \\ 1, & if\ \ \alpha^* \geq \alpha_{(N)}. \end{cases} \tag{34}$$

Therefore, we can approximate $\alpha^{(\mu)}$ as

$$\hat{\alpha}^{(\mu)} = \begin{cases} \alpha_{(1)}, & if\ \ \mu = 0, \\ \alpha_{(i)}, & if\ \ \sum_{j=1}^{i-1} w_j < \mu \leq \sum_{j=1}^{i} w_j. \end{cases} \tag{35}$$

To obtain a $100(1-\mu)\%$ highest posterior density (HPD) credible interval for α, consider intervals of the form

$$R_j = \left[\widehat{\alpha}^{(\frac{j}{N})}, \widehat{\alpha}^{(\frac{j+[(1-\mu)N]}{N})}\right] \tag{36}$$

for $j = 1, 2, \ldots, N - [(1-\mu)N]$, where $[(1-\mu)N]$ denotes the largest integer less than or equal to $[(1-\mu)N]$. Among all R_j, $j = 1, \ldots, N - [(1-\mu)N]$, choose the interval which has the smallest length. The same procedure can be applied to calculate the HPD interval for β.

Now we consider the Bayes estimate of α, when the scale parameter β is known. Without generality, we take $\beta = 1$. We assume that α has the gamma prior distribution with PDF

$$g(\alpha) \propto \alpha^{c-1}\, e^{-d\alpha}, \quad \alpha > 0,\ c, d > 0,$$

where the hyper parameters c and d are known and non-negative. The posterior PDF of α given the data is

$$\pi(\alpha|\underline{x}) \propto\ \alpha^{n+c-1}\ e^{-\alpha(d+\sum_{i=1}^{n}\log x_i)} \prod_{i=1}^{n} \frac{x_i^{2\alpha-1}}{(x_i^{\alpha}+1)^2},$$

which can be rewritten as

$$\pi(\alpha|\underline{x}) \propto G_{\alpha}(n+c, d+\sum_{i=1}^{n}\log x_i)\ h(\alpha, \underline{x}),$$

where

$$h(\alpha, \underline{x}) = \prod_{i=1}^{n} \frac{x_i^{2\alpha-1}}{(x_i^{\alpha}+1)^2}.$$

Again, we can apply the importance sampling scheme to generate samples from the posterior distribution $\pi(\alpha|\underline{x})$ using the following algorithm:

Step 1. Generate $\alpha_1,\ldots,\alpha_N$ from $G(n+c,d+\sum_{i=1}^{n}\log x_i)$.

Step 2. Obtain $h(\alpha_i,\ x)$; $i=1,\ldots,N$.

Step 3. Obtain the approximate Bayes estimate of α under SEL as

$$\widehat{\alpha}_{BS}\approx\frac{\sum_{i=1}^{N}\alpha_i\ h(\alpha_i,\underline{x})}{\sum_{i=1}^{N}h(\alpha_i,\underline{x})}. \tag{37}$$

The credible interval of α can be obtained as described before.

8. Interval estimates of α for known β

Since the two-parameter NP distribution does not satisfy the standard regularity conditions, it is not easy to obtain asymptotic confidence intervals of α and β. However, when the scale parameter β is known, exact and asymptotic confidence intervals for α can be constructed. Without loss of generality, we assume $\beta=1$.

If $x_1,\ldots,x_n$ is a random sample from the NP$(\alpha,1)$ distribution with the CDF

$$F(x;\alpha)=1-\frac{2}{x^{\alpha}+1},\quad x\geq 1,\ \alpha>0,$$

then the pivotal quantity

$$Q(\alpha)=-2\sum_{i=1}^{n}\ln[1-F(x_i;\alpha)]=-2\sum_{i=1}^{n}\ln\left(\frac{2}{x_i^{\alpha}+1}\right)$$

has the chi-square distribution with $2n$ degrees of freedom. So, a $100(1-\gamma)\%$ confidence interval for α can be constructed from the relation

$$P(\chi^2_{(2n,\gamma/2)}<Q(\alpha)<\chi^2_{(2n,1-\gamma/2)})=\gamma, \tag{38}$$

where $\chi^2_{(2n,\gamma/2)}$ and $\chi^2_{(2n,1-\gamma/2)}$ are the lower and upper $\gamma/2$ percentage points of a chi-square distribution with $2n$ degrees of freedom. Note that

$$\frac{dQ(\alpha)}{d\alpha}=2\sum_{i=1}^{n}\frac{x_i^{\alpha}\ \ln x_i}{x_i^{\alpha}+1}>0.$$

This implies that $Q(\alpha)$ is an increasing function in α. Therefore, an exact $100(1-\gamma)\%$ confidence interval for α based on the pivotal quantity $Q(\alpha)$ can be computed as

$$\left(\varphi(x_1,\ldots,x_n,\chi^2_{(2n,\alpha/2)})<\alpha<\varphi(x_1,\ldots,x_n,\chi^2_{(2n,1-\alpha/2)})\right),$$

where $\varphi(x_1,\ldots,x_n,t)$ is the solution of α for the equation $Q(\alpha)=t$.

From asymptotic normality of the MLE, an asymptotic confidence interval for α can be constructed. If $\widehat{\alpha}$ is the MLE of α, then according to Equation (7), the observed Fisher information can be computed as

$$I(\widehat{\alpha}) = -\frac{d^2 l(\alpha)}{d\alpha^2}|_{\alpha=\widehat{\alpha}} = 2\sum_{i=1}^{n} \frac{x_i^{\widehat{\alpha}} \log^2 x_i}{\left(1 + x_i^{\widehat{\alpha}}\right)^2} + \frac{n}{\widehat{\alpha}}. \tag{39}$$

The variance of $\widehat{\alpha}$ can be approximated by the inverse of the observed Fisher information, i.e.,

$$\widehat{\mathrm{Var}}(\widehat{\alpha}) = I^{-1}(\widehat{\alpha}).$$

Therefore, an asymptotic $100(1-\gamma)\%$ confidence interval for α is

$$\widehat{\alpha} \pm z_{1-\gamma/2}\sqrt{\widehat{\mathrm{Var}}(\widehat{\alpha})},$$

where z_q is the q-th upper percentile of the standard normal distribution.

9. SIMULATION RESULTS

To evaluate the performance of different estimation procedures developed in this paper, a Monte Carlo simulation study is presented in this section. We compare the performances of the different estimators in terms of their biases and mean squared errors (MSEs) for different sample sizes and different parameter values. Since β is the scale parameter, we take $\beta = 1$ in all cases considered. We consider $\alpha = 0.5,\ 1.0,\ 1.5,\ 2.0,\ 2.5$ and $n = 10, 30, 50, 100$. For computing Bayes estimates, we use two priors. The first is the non-informative prior: $\gamma = -1$, $b = 0$, $c = 1$ and $d \to \infty$ and the second prior is the informative prior $\gamma = 0.001$, $b = 2$, $c = 5$ and $d = 2$. We call the Bayes estimators under the non-informative and informative priors as “BAYES I” and “BAYES II”, respectively.

9.1. *Estimation of α and β when both are unknown*

Let us consider estimation of α and β when both of them are unknown. In this case, the MLE of β is $\beta = x_{(1)}$. The MLE of α can be obtained by maximizing (5) or equivalently computing the fixed point solution of (6). The MMEs can be computed by solving the non-linear equations (9) and (10). The PCEs can be computed by minimizing (12) with respect to α and β, or equivalently solving the non-linear equations (13) and (14). The LSEs and WLSEs can be obtained by minimizing (18) and (20), respectively, with respect to α and β. The MPS estimates can be obtained by minimizing (25) with respect to α and β. The Bayes estimates can be obtained directly from (30) and (31). In this study, the `optim` function in the `R` software was used for minimization problems. Also, the function `uniroot` in `R` was used to solve the nonlinear equations.

For given n, (α,β) and (γ,b,c,d), we generated the random sample $x_1,\ldots,x_n$ from the NP(α,β) distribution and then computed the estimates of α and β based on different methods. Tables 2 and 3 present the average biases and MSEs based on 1000 replications. The average biases and the MSEs decrease as sample size increases. This shows that all estimates are asymptotically unbiased and consistent. The MPS and Bayes II estimates provide the smallest MSEs. The MMEs have the largest biases whereas the PCEs have the largest MSEs. The MSEs of the WLSEs are smaller than those of the LSEs. Also, the Bayes estimates based on the informative prior perform better than the Bayes estimates based on the non-informative prior, in terms of both biases and MSEs.

TABLE 2
MSEs and average biases (values in parentheses) of different estimates of α.

n	Method	$\alpha=0.5,\ \beta=1.0$	$\alpha=1.0,\ \beta=1.0$	$\alpha=1.5,\ \beta=1.0$	$\alpha=2.0,\ \beta=1.0$	$\alpha=2.5,\ \beta=1.0$
	MLE	0.053 (0.121)	0.212 (0.229)	0.482 (0.327)	0.788 (0.434)	1.289 (0.522)
	MME	- (-)	- (-)	- (-)	- (-)	3.573 (1.417)
	LSE	0.041 (0.018)	0.170 (0.032)	0.436 (0.069)	0.611 (0.052)	1.093 (0.087)
	WLSE	0.038 (0.022)	0.169 (0.043)	0.364 (0.069)	0.566 (0.069)	1.038 (0.112)
10	PCE	0.082 (0.058)	0.441 (0.109)	0.672 (0.162)	1.715 (0.331)	2.588 (0.301)
	MPS	0.025 (-0.002)	0.101 (-0.015)	0.242 (-0.038)	0.381 (-0.054)	0.661 (-0.080)
	BAYES I	0.051 (0.106)	0.233 (0.248)	0.497 (0.362)	0.783 (0.455)	0.530 (-0.159)
	BAYES II	0.015 (-0.014)	0.059 (-0.098)	0.126 (-0.215)	0.283 (-0.402)	0.524 (-0.616)
	MLE	0.008 (0.036)	0.038 (0.076)	0.069 (0.086)	0.144 (0.131)	0.199 (0.152)
	MME	- (-)	- (-)	- (-)	- (-)	0.953 (0.751)
	LSE	0.009 (0.007)	0.043 (0.021)	0.083 (0.014)	0.160 (0.012)	0.226 (0.008)
	WLSE	0.008 (0.012)	0.039 (0.034)	0.070 (0.026)	0.146 (0.041)	0.202 (0.040)
30	PCE	0.014 (0.013)	0.057 (0.028)	0.138 (0.047)	0.246 (0.074)	0.360 (0.067)
	MPS	0.006 (-0.007)	0.027 (-0.012)	0.054 (-0.045)	0.108 (-0.043)	0.152 (-0.067)
	BAYES I	0.008 (0.030)	0.035 (0.071)	0.077 (0.096)	0.139 (0.120)	0.165 (-0.049)
	BAYES II	0.005 (-0.010)	0.022 (-0.036)	0.051 (-0.083)	0.095 (-0.157)	0.165 (-0.242)
	MLE	0.005 (0.025)	0.018 (0.046)	0.040 (0.060)	0.072 (0.083)	0.116 (0.120)
	MME	- (-)	- (-)	- (-)	- (-)	0.696 (0.650)
	LSE	0.005 (0.006)	0.022 (0.015)	0.048 (0.018)	0.086 (0.018)	0.133 (0.029)
	WLSE	0.005 (0.012)	0.019 (0.024)	0.043 (0.033)	0.077 (0.039)	0.116 (0.059)
50	PCE	0.008 (0.008)	0.032 (0.010)	0.068 (0.013)	0.129 (0.036)	0.210 (0.052)
	MPS	0.004 (-0.003)	0.014 (-0.009)	0.032 (-0.023)	0.059 (-0.028)	0.091 (-0.019)
	BAYES I	0.004 (0.023)	0.017 (0.044)	0.041 (0.066)	0.070 (0.079)	0.100 (-0.070)
	BAYES II	0.003 (-0.009)	0.015 (-0.024)	0.034 (-0.057)	0.063 (-0.093)	0.098 (-0.164)
	MLE	0.001 (0.009)	0.007 (0.018)	0.017 (0.024)	0.033 (0.046)	0.048 (0.045)
	MME	- (-)	- (-)	- (-)	- (-)	0.364 (0.458)
	LSE	0.002 (0.000)	0.010 (0.003)	0.022 (0.000)	0.043 (0.019)	0.060 (0.007)
	WLSE	0.002 (0.005)	0.008 (0.009)	0.019 (0.010)	0.038 (0.034)	0.052 (0.025)
100	PCE	0.004 (-0.000)	0.017 (0.013)	0.037 (0.006)	0.067 (0.013)	0.110 (0.007)
	MPS	0.001 (-0.005)	0.007 (-0.011)	0.016 (-0.020)	0.029 (-0.013)	0.043 (-0.029)
	BAYES I	0.002 (0.012)	0.011 (0.023)	0.029 (0.023)	0.044 (0.043)	0.076 (-0.158)
	BAYES II	0.002 (-0.027)	0.010 (-0.058)	0.025 (-0.082)	0.042 (-0.110)	0.067 (-0.165)

Sometime models can be misspecified. Figures 1 to 4 in Appendix A show the effect of misspecification of the NP(α,β) distribution. We simulated 1000 random samples each of size n from the NP$(\alpha+\epsilon,\beta+\delta)$ distribution. We then computed the bias and MSE of each estimator for each n assuming that the true distribution was NP(α,β). The relative changes in the bias and MSE with respect to the not misspecified case are shown in Figures 1 to 4. The biases and the MSEs for α do not appear to change a lot with respect to misspecification. However, the biases and the MSEs for β appear to change a lot with respect to misspecification. The figures correspond to BAYES II estimators and $n=10,30,50,100$. But the results were similar for other estimators and other values of n.

TABLE 3
MSEs and average biases (values in parentheses) of different estimates of β.

n	Method	$\alpha = 0.5$, $\beta = 1.0$	$\alpha = 1.0$, $\beta = 1.0$	$\alpha = 1.5$, $\beta = 1.0$	$\alpha = 2.0$, $\beta = 1.0$	$\alpha = 2.5$, $\beta = 1.0$
	MLE	0.871 (0.538)	0.106 (0.213)	0.037 (0.137)	0.018 (0.098)	0.012 (0.077)
	MME	- (-)	- (-)	- (-)	- (-)	0.108 (0.267)
	LSE	0.475 (0.026)	0.108 (-0.014)	0.048 (0.000)	0.024 (-0.014)	0.018 (-0.005)
	WLSE	0.418 (0.034)	0.090 (-0.001)	0.040 (0.008)	0.019 (-0.004)	0.015 (0.001)
10	PCE	0.462 (0.016)	0.052 (-0.017)	0.022 (-0.004)	0.011 (-0.007)	0.007 (-0.011)
	MPS	0.288 (0.029)	0.046 (-0.011)	0.017 (-0.010)	0.009 (-0.011)	0.006 (-0.010)
	BAYES I	0.706 (0.532)	0.121 (0.242)	0.047 (0.155)	0.025 (0.118)	0.008 (0.049)
	BAYES II	0.181 (0.283)	0.063 (0.137)	0.019 (0.066)	0.012 (0.045)	0.007 (0.028)
	MLE	0.048 (0.146)	0.010 (0.066)	0.004 (0.045)	0.002 (0.033)	0.001 (0.025)
	MME	- (-)	- (-)	- (-)	- (-)	0.042 (0.175)
	LSE	0.104 (0.020)	0.023 (-0.004)	0.010 (0.001)	0.005 (-0.005)	0.003 (-0.006)
	WLSE	0.075 (0.035)	0.016 (0.008)	0.007 (0.009)	0.003 (0.003)	0.001 (0.000)
30	PCE	0.024 (-0.021)	0.005 (-0.010)	0.002 (-0.008)	0.001 (-0.005)	0.000 (-0.006)
	MPS	0.021 (0.000)	0.005 (-0.003)	0.002 (-0.001)	0.001 (-0.001)	0.000 (-0.002)
	BAYES I	0.051 (0.159)	0.013 (0.088)	0.005 (0.061)	0.003 (0.048)	0.001 (0.024)
	BAYES II	0.039 (0.128)	0.008 (0.062)	0.003 (0.039)	0.001 (0.025)	0.001 (0.022)
	MLE	0.015 (0.085)	0.003 (0.040)	0.001 (0.027)	0.000 (0.019)	0.000 (0.016)
	MME	- (-)	- (-)	- (-)	- (-)	0.029 (0.148)
	LSE	0.045 (-0.003)	0.012 (0.001)	0.005 (0.000)	0.002 (-0.003)	0.001 (-0.001)
	WLSE	0.027 (0.015)	0.007 (0.010)	0.003 (0.007)	0.001 (0.002)	0.001 (0.003)
50	PCE	0.007 (-0.015)	0.001 (-0.010)	0.000 (-0.005)	0.000 (-0.005)	0.000 (-0.003)
	MPS	0.007 (0.000)	0.001 (-0.000)	0.000 (-0.000)	0.000 (-0.001)	0.000 (-0.000)
	BAYES I	0.019 (0.101)	0.005 (0.058)	0.002 (0.041)	0.001 (0.036)	0.000 (0.015)
	BAYES II	0.014 (0.080)	0.003 (0.038)	0.001 (0.024)	0.000 (0.018)	0.000 (0.014)
	MLE	0.003 (0.040)	0.000 (0.020)	0.000 (0.013)	0.000 (0.009)	0.000 (0.007)
	MME	- (-)	- (-)	- (-)	- (-)	0.019 (0.120)
	LSE	0.023 (0.004)	0.005 (-0.000)	0.002 (-0.000)	0.001 (0.000)	0.000 (0.000)
	WLSE	0.012 (0.017)	0.002 (0.006)	0.001 (0.004)	0.000 (0.004)	0.000 (0.003)
100	PCE	0.001 (-0.012)	0.000 (-0.005)	0.000 (-0.002)	0.000 (-0.003)	0.000 (-0.002)
	MPS	0.001 (-0.000)	0.000 (-0.000)	0.000 (-0.000)	0.000 (-0.000)	0.000 (-0.000)
	BAYES I	0.004 (0.056)	0.001 (0.037)	0.001 (0.029)	0.000 (0.026)	0.000 (0.007)
	BAYES II	0.003 (0.041)	0.000 (0.020)	0.000 (0.013)	0.000 (0.010)	0.000 (0.007)

9.2. *Estimation of α when β is known*

In this section, we consider estimation of α when β is known. The MLE of α can be obtained by maximizing (7) with respect to α or equivalently by solving the solution (8). The MME of α can be computed by solving the equation $2\alpha A_1(\alpha) = \overline{x}$. The PCE can be obtained by minimizing (15) with respect to α or equivalently by solving (16). The LSE and WLSE of α can be obtained by minimizing (21) and (22), respectively, with respect to α only. The MPS estimate can be obtained by minimizing (26) with respect to α. The Bayes estimate can be obtained directly from (37). For Bayesian estimation, we used the two priors $d = c = 0$ (non-informative) and $c = 1$, $d = 4$ (informative prior). Table 4 presents the average biases and MSEs based on 1000 replications.

From Table 4, again as sample size increases, the average biases and the MSEs decrease. The PCE and Bayes II have the smallest MSEs. The MME provides the largest biases and MSEs.

TABLE 4
MSEs and average biases (values in parentheses) of different estimates of α.

n	Method	$\alpha = 0.5$	$\alpha = 1.0$	$\alpha = 1.5$	$\alpha = 2.0$	$\alpha = 2.5$
	MLE	0.026 (0.046)	0.094 (0.086)	0.237 (0.133)	0.476 (0.179)	0.729 (0.203)
	MME	- (-)	- (-)	0.281 (0.283)	0.469 (0.289)	0.680 (0.326)
	LSE	0.045 (0.029)	0.128 (0.058)	0.324 (0.101)	0.601 (0.118)	0.871 (0.132)
	WLSE	0.042 (0.026)	0.119 (0.052)	0.308 (0.090)	0.572 (0.108)	0.862 (0.126)
10	PCE	0.019 (-0.037)	0.076 (-0.076)	0.174 (-0.107)	0.292 (-0.139)	0.672 (-0.132)
	MPS	0.020 (-0.007)	0.071 (-0.019)	0.182 (-0.023)	0.361 (-0.033)	0.566 (-0.060)
	BAYES I	0.025 (0.042)	0.108 (0.072)	0.256 (0.101)	0.397 (0.147)	0.610 (0.168)
	BAYES II	0.021 (0.033)	0.060 (-0.007)	0.102 (-0.101)	0.193 (-0.250)	0.316 (-0.092)
	MLE	0.007 (0.014)	0.027 (0.030)	0.061 (0.042)	0.102 (0.035)	0.175 (0.089)
	MME	- (-)	- (-)	0.086 (0.147)	0.129 (0.125)	0.180 (0.118)
	LSE	0.009 (0.009)	0.033 (0.021)	0.075 (0.031)	0.130 (0.024)	0.235 (0.058)
	WLSE	0.008 (0.009)	0.031 (0.019)	0.069 (0.029)	0.120 (0.019)	0.211 (0.057)
30	PCE	0.006 (-0.030)	0.027 (-0.051)	0.063 (-0.079)	0.116 (-0.087)	0.240 (-0.124)
	MPS	0.006 (-0.009)	0.024 (-0.017)	0.055 (-0.028)	0.096 (-0.058)	0.152 (-0.029)
	BAYES I	0.007 (0.011)	0.025 (0.008)	0.066 (0.035)	0.119 (0.035)	0.165 (0.056)
	BAYES II	0.006 (0.007)	0.023 (-0.007)	0.054 (-0.043)	0.087 (-0.103)	0.134 (-0.025)
	MLE	0.003 (0.006)	0.015 (0.015)	0.036 (0.041)	0.066 (0.043)	0.098 (0.050)
	MME	- (-)	- (-)	0.060 (0.110)	0.085 (0.099)	0.115 (0.097)
	LSE	0.004 (0.003)	0.019 (0.013)	0.046 (0.031)	0.087 (0.027)	0.133 (0.035)
	WLSE	0.004 (0.003)	0.018 (0.011)	0.042 (0.030)	0.079 (0.028)	0.120 (0.035)
50	PCE	0.004 (-0.021)	0.016 (-0.039)	0.039 (-0.056)	0.066 (-0.074)	1.488 (-0.089)
	MPS	0.003 (-0.009)	0.014 (-0.017)	0.032 (-0.007)	0.061 (-0.020)	0.090 (-0.030)
	BAYES I	0.004 (-0.000)	0.017 (-0.009)	0.037 (-0.008)	0.064 (0.002)	0.101 (-0.008)
	BAYES II	0.003 (-0.006)	0.014 (-0.018)	0.034 (-0.049)	0.059 (-0.099)	0.069 (-0.078)
	MLE	0.001 (0.004)	0.007 (0.013)	0.016 (0.009)	0.031 (0.030)	0.046 (0.014)
	MME	- (-)	- (-)	0.031 (0.063)	0.045 (0.061)	0.053 (0.049)
	LSE	0.002 (0.003)	0.010 (0.010)	0.021 (0.003)	0.040 (0.021)	0.059 (0.007)
	WLSE	0.002 (0.003)	0.009 (0.010)	0.019 (0.003)	0.036 (0.023)	0.053 (0.008)
100	PCE	0.002 (-0.014)	0.008 (-0.024)	0.019 (-0.036)	0.032 (-0.049)	0.121 (-0.138)
	MPS	0.001 (-0.004)	0.007 (-0.005)	0.015 (-0.018)	0.029 (-0.006)	0.045 (-0.031)
	BAYES I	0.002 (-0.027)	0.010 (-0.050)	0.023 (-0.073)	0.043 (-0.097)	0.064 (-0.132)
	BAYES II	0.002 (-0.029)	0.010 (-0.052)	0.023 (-0.082)	0.042 (-0.116)	0.047 (-0.161)

10. NUMERICAL EXAMPLES

In this section, we use some real data sets to illustrate the proposed estimation methods discussed in the previous sections.

10.1. Example 1 (both parameters are unknown)

The real data set (see Table 5) represents the times to breakdown of a type of electronic insulating material subjected to a constant-voltage stress. These data are taken from Nelson (1970) and has been used earlier by Tiku and Akkaya (2004).

TABLE 5
Data set in Example 1.

0.35	0.59	0.96	0.99	1.69	1.97	2.07	2.58
2.71	2.90	3.67	3.99	5.35	13.77	25.50	

An important characteristic to decide whether a particular distribution is suitable or not for a data set is the empirical hazard rate function. Here, we use the scaled total time on test (TTT) function to detect the type of the hazard rate function that the data have and then choose a suitable distribution; see Aarset (1987). The TTT plot is obtained by plotting

$$T(\frac{r}{n}) = \frac{\sum_{i=1}^{r} x_{i:n} + (n-r)\, x_{r:n}}{\sum_{i=1}^{n} x_{i:n}}$$

against r/n, where $r = 1, \ldots, n$. It is a straight diagonal for constant hazard rates, convex for decreasing hazard rates and concave for increasing hazard rates. It is first convex and then concave if the hazard rate is bathtub-shaped. It is first concave and then convex if the hazard rate is upside-down bathtub (unimodal) shaped. The TTT plot for the above data set is presented in Figure 5. This plot indicates that the empirical hazard rate function of the data set is decreasing. Therefore, the NP distribution is appropriate to fit the data set since this distribution can present decreasing hazard rate functions.

First we compute the MLEs of the unknown parameters. The MLE of β is $\widehat{\beta}_{ML} = 0.35$. The MLE of α can be computed by maximizing the profile log-likelihood $g(\alpha)$ in (5). The MLE of α is obtained as $\widehat{\alpha}_{ML} = 0.7403$. The profile log-likelihood $g(\alpha)$ is plotted in Figure 5. As we can see, it is a unimodal function. The Kolmogorov-Smirnov (K-S) distances between the fitted and empirical CDFs was 0.25, and the corresponding p-value was 0.22. Therefore, based on the MLEs, we can not reject the assumption that the data set are coming from the NP distribution.

We also computed K-S distance based on MMEs, LSEs, WLSEs, PCEs, MPSs and Bayes I estimates. Their estimates, K-S distances and the corresponding p-values are given in Table 6. From this table, all considered estimates provide a satisfactory to the data set. Table 7 presents 95% credible intervals.

Figure 6 plots the empirical CDF and histogram. Superimposed are the fitted CDFs and PDFs of the parameter estimates under consideration. These plots confirm the results in Table 6.

TABLE 6
Estimates, K-S distances and corresponding p-values based on different estimates for Example 1.

Estimation method	MLEs	MMEs	LSEs	WLSEs	PCEs	MPSs	BAYES I
$\widehat{\alpha}$	0.740	2.399	0.935	0.935	0.478	0.646	0.649
$\widehat{\beta}$	0.350	1.699	0.682	0.671	0.279	0.293	0.275
K-S	0.258	0.333	0.133	0.140	0.258	0.245	0.262
p-value	0.227	0.054	0.918	0.889	0.224	0.277	0.212

TABLE 7
95% *credible intervals of the parameters.*

Parameter	α	β
Credible interval	[1.329, 2.507]	[0.985, 1.012]

10.2. *Example 2 (β is known)*

Dyer (1981) reported annual wage data (in multiplies of 100 US dollars) of a random sample of 30 production-line workers in a large industrial firm, as presented in Table 8. He showed that the Pareto distribution provided an adequate fit for this data set. Here we fit the NP distribution to this data set. We observed that the NP distribution with $\alpha = 8.1261$ and $\beta = 101$ fits to above data set. We checked the validity of the NP distribution based on the K-S test. The K-S distance was 0.09 and the corresponding p-value was 0.95. Let us transform this data set to the standard NP distribution with the scale parameter $\beta = 1$. We know that if the random variable X follows NP(α, β), then the random variable $Z = X/\beta$ has the standard NP$(\alpha, 1)$ distribution. Therefore, we transform the above data to NP$(\alpha, 1)$ by dividing by β. The transformed data set are reported in Table 8.

We fitted the standard NP$(\alpha, 1)$ distribution to the transformed data set. We computed the MLE, MME, LSE, WLSE, PCE, MPS and Bayes I estimates of α as described in Section 2. We also computed K-S distance based on these estimates. The estimates, K-S distances and the corresponding p-values are presented in Table 9. The results in Table 9 show that the standard NP$(\alpha, 1)$ model is fitted reasonably well to the transformed data set and all the estimates provide satisfactory fits. Table 10 presents 95% exact and asymptotic confidence intervals and also the 95% credible interval for α.

Figure 7 plots the empirical CDF and histogram. Superimposed are the fitted CDFs and PDFs of the parameter estimates under consideration. This figure supports the results in Table 9.

TABLE 8
The annual wage data.

Data Set	112	154	119	108	112	156	123	103
	115	107	125	119	128	132	107	151
	103	104	116	140	108	105	158	104
	119	111	101	157	112	115		
Transformed data	1.108	1.524	1.178	1.069	1.108	1.544	1.217	1.019
	1.138	1.059	1.237	1.178	1.267	1.306	1.059	1.495
	1.019	1.029	1.148	1.386	1.069	1.039	1.564	1.029
	1.178	1.099	1.000	1.554	1.108	1.138		

TABLE 9
Estimates, K-S distances and corresponding p-values based on different estimates for Example 2.

Estimation method	MLEs	MMEs	LSEs	WLSEs	PCEs	MPSs	BAYES I
$\widehat{\alpha}$	8.126	8.244	8.328	8.213	7.745	6.491	7.683
K-S	0.093	0.096	0.098	0.095	0.105	0.179	0.108
p-value	0.956	0.942	0.931	0.946	0.894	0.288	0.871

TABLE 10
95% intervals of α based on the annual wage data.

Exact confidence interval	Asymptotic confidence interval	Credible interval
$[5.887, 10.773]$	$[5.971, 10.280]$	$[6.117, 9.844]$

11. CONCLUSIONS

We have compared eight methods to estimate parameters of a new Pareto distribution due to Bourguignon *et al.* (2016). Six of these are frequentist methods: maximum likelihood estimators, method of moment estimators, percentile estimators, least square and weighted least square estimators and maximum product of spacing estimators. The remaining two are Bayes estimators based on informative and non-informative priors. The performance of the estimators was assessed by a simulation study and two real data applications. The maximum product of spacing estimators and Bayes estimators based on informative priors were shown to provide the best performance when both parameters of the distribution are unknown. The percentile estimators and Bayes estimators based on informative priors were shown to provide the best performance when the scale parameter of the distribution is known. A future work is to derive multivariate, matrix variate and complex variate extensions of the distribution and to study their estimation issues.

ACKNOWLEDGEMENTS

The authors would like to thank the Editor and the Referee for careful reading and comments which greatly improved the paper.

APPENDIX

A. FIGURES

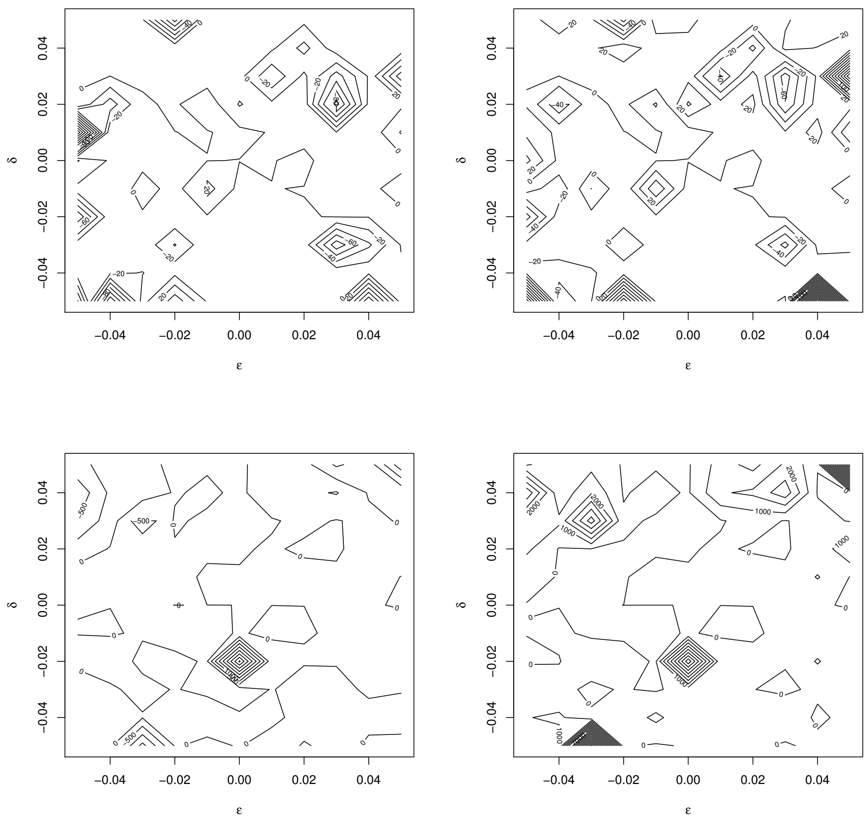

Figure 1 – Contours of the relative percentage changes in the bias of the estimator of α (top left), the MSE of the estimator of α (top right), the bias of the estimator of β (bottom left) and the MSE of the estimator of β (bottom right) when $n = 10$.

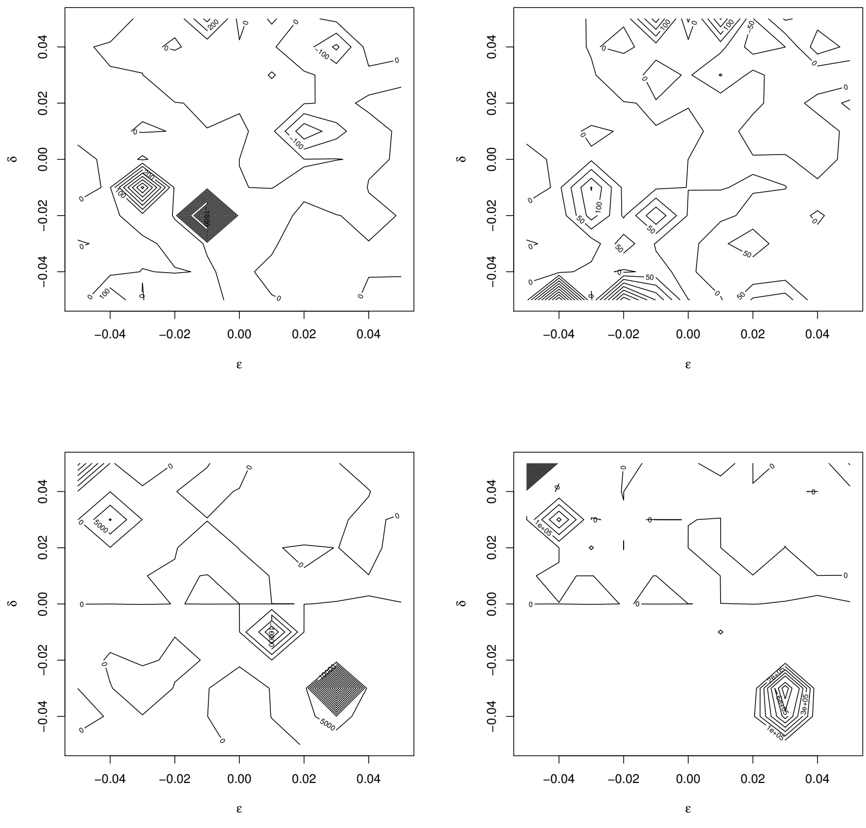

Figure 2 – Contours of the relative percentage changes in the bias of the estimator of α (top left), the MSE of the estimator of α (top right), the bias of the estimator of β (bottom left) and the MSE of the estimator of β (bottom right) when $n = 30$.

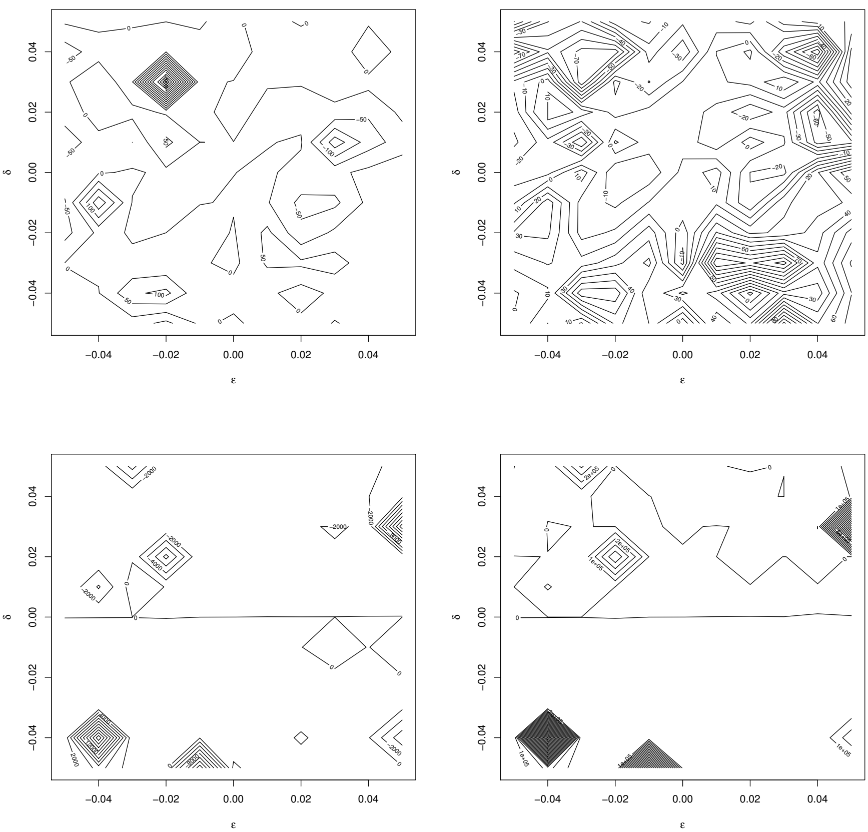

Figure 3 – Contours of the relative percentage changes in the bias of the estimator of α (top left), the MSE of the estimator of α (top right), the bias of the estimator of β (bottom left) and the MSE of the estimator of β (bottom right) when $n = 50$.

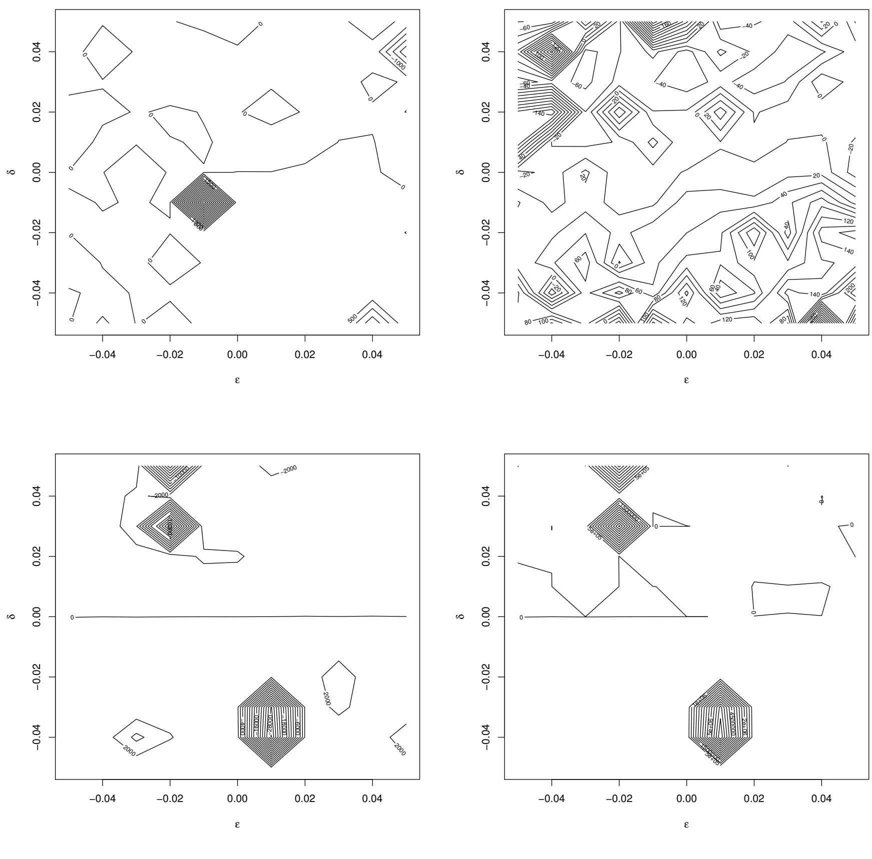

Figure 4 – Contours of the relative percentage changes in the bias of the estimator of α (top left), the MSE of the estimator of α (top right), the bias of the estimator of β (bottom left) and the MSE of the estimator of β (bottom right) when $n = 100$.

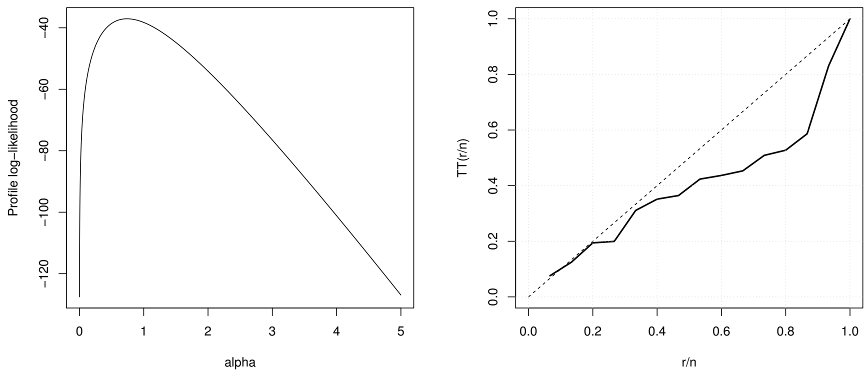

Figure 5 – Profile log-likelihood function of α (left) and TTT plot for the data set (right).

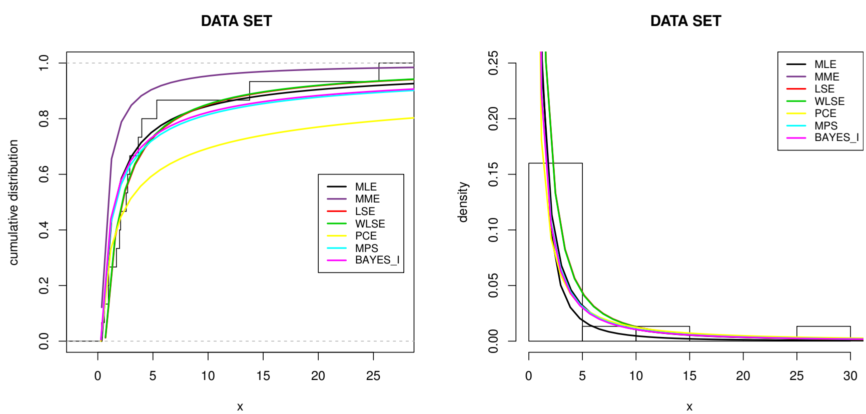

Figure 6 – The empirical CDF and histogram with the fitted CDFs and PDFs.

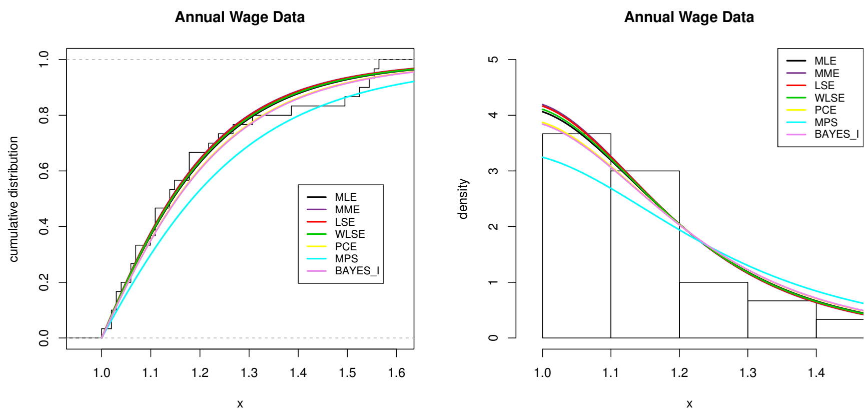

Figure 7 – The empirical CDF and histogram with the fitted CDFs and PDFs.

REFERENCES

M. V. AARSET (1987). *How to identify bathtub hazard rate*. IEEE Transactions on Reliability, 36, pp. 106–108.

M. ABDI, A. ASGHARZADEH, H. S. BAKOUCH, Z. ALIPOUR (2019). *A new compound gamma and Lindley distribution with application to failure data*. Austrian Journal of Statistics, 48, pp. 54–75.

M. R. ALKASABEH, M. Z. RAQAB (2009). *Estimation of the generalized logistic distribution parameters: Comparative study*. Statistical Methodology, 6, pp. 262–279.

B. C. ARNOLD (1983). *Pareto Distributions*. International Cooperative Publishing House, Fairland, Maryland, USA.

B. C. ARNOLD, S. J. PRESS (1983). *Bayesian inference for Pareto populations*. Journal of Econometrics, 21, pp. 287–306.

B. C. ARNOLD, S. J. PRESS (1989). *Bayesian estimation and prediction for Pareto data*. Journal of the American Statistical Association, 84, pp. 1079–1084.

A. ASGHARZADEH, R. REZAIE, M. ABDI (2011). *Comparisons of methods of estimation for the half-logistic distribution*. Selcuk Journal of Applied Mathematics, Special Issue, pp. 93–108.

H. S. BAKOUCH, S. DEY, P. L. RAMOS, F. LOUUZADA (2017). *Binomial-exponential 2 distribution: Different estimation methods with weather applications*. Trends in Applied and Computational Mathematics, 18, no. 2, pp. 233–251.

M. Bourguignon, H. Saulo, R. N. Fernandez (2016). *A new Pareto-type distribution with applications in reliability and income data*. Physica A, 457, pp. 166–175.

M. H. Chen, Q. M. Shao (1999). *Monte Carlo estimation of Bayesian credible and HPD intervals*. Journal of Computational and Graphical Statistics, 8, pp. 69–92.

R. C. H. Cheng, N. A. K. Amin (1979). *Maximum product of spacings estimation with applications to the lognormal distribution*. Tech. rep., Department of Mathematics, University of Wales.

R. C. H. Cheng, N. A. K. Amin (1983). *Estimating parameters in continuous univariate distributions with a shifted origin*. Journal of the Royal Statistical Society, Series B, , no. 3, pp. 394–403.

C. Dagum (2006). *Wealth distribution models: Analysis and applications*. Statistica, 66, no. 3, pp. 235–268.

R. Demicheli, G. Bonadonna, W. J. Hrushesky, M. W. Retsky, P. Valagussa (2004). *Menopausal status dependence of the timing of breast cancer recurrence after surgical removal of the primary tumour*. Breast Cancer Research, 6, no. 6, pp. 689–696.

S. Dey, T. Dey, D. Kundu (2014). *Two-parameter Rayleigh distribution: Different methods of estimation*. American Journal of Mathematical and Management Sciences, 33, no. 1, pp. 55–74.

D. Dyer (1981). *Structural probability bounds for the strong Pareto laws*. Canadian Journal of Statistics, 9, pp. 71–77.

R. D. Gupta, D. Kundu (2001). *Generalized exponential distribution: Different methods of estimation*. Journal of Statistical Computation and Simulation, 69, pp. 315–338.

N. L. Johnson, S. Kotz, N. Balakrishnan (1994). *Continuous Univariate Distributions*, vol. 1. Wiley, New York.

J. Kao (1959a). *Computer methods for estimating Weibull parameters in reliability studies*. IRE Transactions on Reliability and Quality Control, 13, pp. 15–22.

J. Kao (1959b). *A graphical estimation of mixed Weibull parameters in life testing electron tubes*. Technometrics, 1, pp. 389–407.

D. Kundu, M. Z. Raqab (2005). *Generalized Rayleigh distribution: Different methods of estimation*. Computational Statistics and Data Analysis, 49, no. 1, pp. 187–200.

C. Lai, M. Xie (2006). *Stochastic Ageing and Dependence for Reliability*. Springer, New York.

T. Lwin (1972). *Estimation of the tail of the Paretian law*. Skand Aktuarietidskr, 55, pp. 170–178.

N. R. MANN, R. E. SCHAFER, N. D. SINGPURWALLA (1974). *Methods for Statistical Analysis of Reliability and Life Data*. Wiley, New York.

W. B. NELSON (1970). *Statistical methods for accelerated life test data the inverse power law model*. Technical report 71-c011, General Electric Company.

M. E. J. NEWMAN (2005). *Pareto distributions and Zipfs law*. Contemporary Physics, 46, no. 5, pp. 323–351.

V. PARETO (1964). *Cours d'économie politique*, vol. 1. Librairie Droz, Lausanne.

B. RANNEBY (1984). *The maximum spacing method, an estimation method related to the maximum likelihood method*. Scandinavian Journal of Statistics, 11, pp. 93–112.

P. G. SANKARAN, N. U. NAI, P. JOHN (2014). *A family of bivariate Pareto distributions*. Statistica, 74, no. 2, pp. 199–215.

J. SWAIN, S. VENKATRAMAN, J. WILSON (1988). *Least squares estimation of distribution function in Johnson's translation system*. Journal of Statistical Computation and Simulation, 29, pp. 271–297.

M. L. TIKU, A. D. AKKAYA (2004). *Robust Estimation and Hypothesis Testing*. New Age International (P) Limited, New Delhi.

L. ZANINETTI, M. FERRARO (2008). *On the truncated Pareto distribution with applications*. Central European Journal of Physics, 6, no. 1, pp. 1–6.

SUMMARY

Bourguignon *et al.* (2016) introduced a new Pareto-type distribution to model income and reliability data. The aim of this paper is to estimate the parameters of this distribution from both frequentist and Bayesian view points. The maximum likelihood estimates, method of moment estimates, percentile estimates, least square and weighted least square estimates and maximum product of spacing estimates are considered as frequentist estimates. We have also considered the Bayes estimates of the unknown parameters and the associated credible intervals. The Bayes estimates are computed using an importance sampling method. To evaluate the performance of the different estimates, a Monte Carlo simulation study is carried out. Some real life data sets have been analyzed for illustrative purposes.

Keywords: Bayesian estimates; Least squares estimates; Maximum likelihood estimates; Method of moment estimates; Monte Carlo simulation; Percentiles estimates; Weighted least squares estimates.