The Beta Truncated Pareto Distribution

Lourenzutti R.ᵃ, Duarte D.ᵃ*, Castellares F.ᵃ and Azevedo M.ᵃ

Universidade Federal de Minas Gerais,Belo Horizonte, MG, Brazil;

The Pareto distribution is widely used to modelling a diverse range of phenomena. Many transformations and generalization of the Pareto law distribution have been proposed in order to get more flexible models. In fact, these generalizations are very common in literature. Recently a new family of distribution, called Beta Generated distribution, was proposed. This family presents itself as a very flexible family, capable of modelling symmetric and skewed data. In this work we apply the beta transformation to the truncated Pareto distribution. We analyse some of its properties and apply it to real data. For estimation we the minimum distance method. This new distribution, called Beta truncated Pareto, proved to be a very flexible.

1. Introduction

In 1896, Vilfredo Pareto noticed that the income distribution "follows a power law behavior", i.e., the pdf of income has the form $f(x) \propto x^{-\alpha-1}$ [see 1]. Later, it was found that the power law distribution is present in a diverse range of phenomena, as for instance, the sizes of earthquakes [2], solar flares [3][also, see 4], the frequency of use of words in any human language [5], the frequency of occurrence of personal names in most cultures [6], the numbers of papers that scientists write [7], the number of citations received by papers [8] and hydrological data [9]. Also, some recent works about the power law distribution (also known as Pareto distribution) corroborate with the hypothesis made by Pareto [see 10–12].

Many extensions of the Pareto distribution were proposed in literature. For example, the transformed Pareto distribution [13, p. 193], the truncated Pareto distribution (TPD) [14, 15], the double Pareto distribution [16], the double Pareto-lognormal distribution [17] and the Generalized Pareto distribution (GPD) [see 18, 19]. The TPD and GPD are ones of the most important extensions of the Pareto distribution.

In [15], the Pareto and truncated Pareto distributions were applied to astrophysics data and they concluded that, in general, the truncated Pareto distribution performs better than the usual Pareto distribution. If the difference between the maximum and minimum values of the sample is high, then there is no real difference between the two distributions.

All of those generalizations mentioned above were not exclusive for the Pareto distribution. It also has been proposed some modifications for the Weibull distribution [20, 21], the gamma distribution [22], the exponencial distribution [23] and many others. These efforts to modify distributions are attempt to get more flexible models. When a new family of distributions has simpler families as special cases, [24] suggests, in order to define which model is appropriate, to adjust the gener-

alized family of distributions. Then, a statistical analysis of the parameters can indicate properly whether one of the simpler families is better fitted to the data or if the new proposed distribution provides a more appropriate solution.

Recently, a new family of distributions, generated from the Beta distribution has been proposed by [25]. In that proposal the Beta distribution was used to generalize the Normal family. Later, [26] proposed, in a general fashion, the Beta Generated family. The Pareto family was generalized through the same methodology by [27]. Many properties of the Beta Generated family are shown in [28]. It is becoming evident that the Beta Generated family is a very flexible family of distributions.

Motivated by the works of [14, 15] and [27], we propose the Beta truncated Pareto distribution. We explore some of its characteristics and we compare its performance to the Beta Pareto distribution and other distributions.

This paper is organized as follows. Section 2 introduces the Beta truncated Pareto distribution and presents some of its properties. Section 3 provides inferential results. Section 4 shows an application of BTPD to a real data and Section 5 presents concluding remarks.

2. The Beta Generated Family and the Beta truncated Pareto

Let Z and Y be two random variables, where $Z \sim Beta(\alpha, \beta)$ and Y has cdf F_Y. Define $X = F_Y^{-1}(Z)$. Then, X has probability density function (pdf) given by

$$f_X(x|\alpha, \beta, \mathbf{p}) = \frac{1}{B(\alpha, \beta)} f_Y(x) \left[F_Y(x)\right]^{\alpha-1} \left[1 - F_Y(x)\right]^{\beta-1} \tag{1}$$

and cumulative distribution function (cdf) given by

$$F_X(x|\alpha, \beta, \mathbf{p}) = \frac{B_{F_Y(x)}(\alpha, \beta)}{B(\alpha, \beta)}, \tag{2}$$

where $B(a, b)$ is the beta function, $\mathbf{p}$ is a parameter vector associated with F_Y, and $B_x(a, b) = \int_0^x u^{a-1}(1-u)^{b-1}du$, $0 < x < 1$, is the incomplete beta function.

In this case, F_X represents a beta generated distribution with *parent distribution* F_Y. Note that if α and β are integers then we have the distribution of order statistics. Furthermore, if $\alpha = 1$ and $\beta = 1$, the Equation 1 becomes the parent density.

Now make $Y \sim TP(\theta, \lambda, k)$, where TP means truncated Pareto distribution. Then Y has cdf given by

$$F_Y(y) = \frac{1 - (\frac{\theta}{y})^k}{1 - (\frac{\theta}{\lambda})^k}, \ 0 < \theta < \lambda, \ k > 0 \text{ and } \theta < y, < \lambda \tag{3}$$

and pdf given by

$$f_Y(y) = \frac{k\theta^k y^{-(k+1)}}{1 - (\frac{\theta}{\lambda})^k}, \ 0 < \theta < \lambda \text{ and } \theta < y < \lambda. \tag{4}$$

Define a new variable X as $X = F_Y^{-1}(Z) = \theta \left\{1 - Z\left[1 - (\theta/\lambda)^k\right]\right\}^{-1/k}$. Then, by Equation 1, X has pdf given by

$$f_X(x) = \frac{1}{B(\alpha,\beta)} \frac{k\theta^k x^{-(k+1)}}{1-(\frac{\theta}{\lambda})^k} \left[\frac{1-(\frac{\theta}{x})^k}{1-(\frac{\theta}{\lambda})^k} \right]^{\alpha-1} \left[1 - \frac{1-(\frac{\theta}{x})^k}{1-(\frac{\theta}{\lambda})^k} \right]^{\beta-1}, \quad \theta < x < \lambda. \quad (5)$$

In this case, it can be said that X follows a Beta truncated Pareto distribution (BTPD) with parameters α, β, θ, λ and k, hereafter defined as $X \sim BTP(\alpha,\beta,\theta,\lambda,k)$. As shown in Figure 1, the α and β parameters are related to the shape of the distribution whereas the parameter k is related to the scale of the distribution, but also has effect on the shape.

The cdf of X is obtained directly from Equation 2, shown as follows,

$$F_X(x) = \frac{B_{F_Y(x)}(\alpha,\beta)}{B(\alpha,\beta)}$$

$$= I(x)_{(\theta,\lambda)} \times \frac{1}{B(\alpha,\beta)} \sum_{i=0}^{\infty} \frac{(1-\beta)_i}{i!(\alpha+i)} \left[\frac{1-(\frac{\theta}{x})^k}{1-(\frac{\theta}{\lambda})^k} \right]^{\alpha+i} + I(x)_{[\lambda,\infty)}$$

where $(x)_n = x(x+1)(x+2)\ldots(x+n-1)$.

Now consider the Pareto distribution. The pdf and cdf are given, respectively, by

$$h(y) = \frac{k}{\theta} \left(\frac{\theta}{y} \right)^{k+1} \quad \text{and} \quad H(y) = 1 - \left(\frac{\theta}{y} \right)^k, \quad \theta < y < \infty.$$

Let $c(k,\theta,\lambda) := c(k) = \left[1 - \left(\frac{\theta}{\lambda} \right)^k \right]^{-1}$. Then we can rewrite (3) and (4) as

$$f_y(y) = c(k)\,h(y) \quad \text{and} \quad F_Y(y) = c(k)\,H(y), \quad \theta < y < \lambda.$$

By Equation 1, the BTPD density is

$$f_X(x) = \frac{1}{B(\alpha,\beta)} f_Y(x)\,[F_Y(x)]^{\alpha-1}\,[1-F_Y(x)]^{\beta-1}$$

$$= \frac{1}{B(\alpha,\beta)} c^\alpha(k)h(x)\,[H(x)]^{\alpha-1}\,[1-c(k)H(x)]^{\beta-1}.$$

As the term $|c(k)H(x)|$ is strictly less than 1 in the interval $\theta < x < \lambda$, then

$$f_X(x) = \sum_{i=0}^{\infty} \frac{(-1)^i \binom{\beta-1}{i}}{B(\alpha,\beta)} [c(k)]^{\alpha+i} h(x)\,[H(x)]^{\alpha-1+i}.$$

But $H(x)$ can be rewritten as $H(x) = \sum_{j=0}^{\infty}(-1)^j \binom{\alpha+i-1}{j}\left(\frac{\theta}{x}\right)^{kj}$. Therefrom

$$f_X(x) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^{i+j}\binom{\beta-1}{i}\binom{\alpha+i-1}{j}}{B(\alpha,\beta)}[c(k)]^{\alpha+i}\frac{k}{\theta}\left(\frac{\theta}{x}\right)^{k(j+1)+1}$$

$$= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}A(i,j)c(k(j+1))\frac{k(j+1)}{\theta}\left(\frac{\theta}{x}\right)^{k(j+1)+1}$$

where

$$A(i,j) = \frac{(-1)^{i+j}\binom{\beta-1}{i}\binom{\alpha+i-1}{j}}{(j+1)B(\alpha,\beta)}\frac{[c(k)]^{\alpha+i}}{c(k(j+1))}.$$

It shows that the Beta truncated Pareto distribution is an infinite linear combination of truncated Pareto distribution with parameters $k(j+1)$, θ and λ and with sum coefficients $A(i,j)$.

2.1. Special Cases

Three different families of distributions are identified as special cases from the BTPD. These families are already proposed in the literature and they are shown next.

(1) If $\alpha = 1$ and $\beta = 1$ then the density shown in Equation 5 becomes the density of a truncated Pareto random variable.
(2) The Beta Pareto family proposed by [27] is also a special case. In this case its density is given by

$$f(x) = \frac{k}{\theta \cdot B(\alpha,\beta)}\left[1-\left(\frac{\theta}{x}\right)^k\right]^{\alpha-1}\left(\frac{\theta}{x}\right)^{k\cdot\beta+1}, \text{ where } x > \theta,\ (\alpha,\beta,\theta,k) > 0.$$

(6)

This density represents a limiting density of BTPD when $\lambda \to \infty$.

(3) If $\alpha = 1$, $\beta = 1$ and $\lambda \to \infty$ then the Pareto family is a limiting family of BTPD family.

2.2. Limit Behavior

The behavior of BTPD when $x \to \theta$ and $x \to \lambda$ changes according to parameters α and β. The limits of BTPD in both cases are presented next.

Proposition 2.1:

$$\lim_{x\to\theta}f_X(x) = \begin{cases}\infty & ,\ 0 < \alpha < 1 \\ \frac{\beta k}{\theta\left[1-\left(\frac{\theta}{\lambda}\right)^k\right]} & ,\ \alpha = 1 \\ 0 & ,\ \alpha > 1\end{cases}$$

(7)

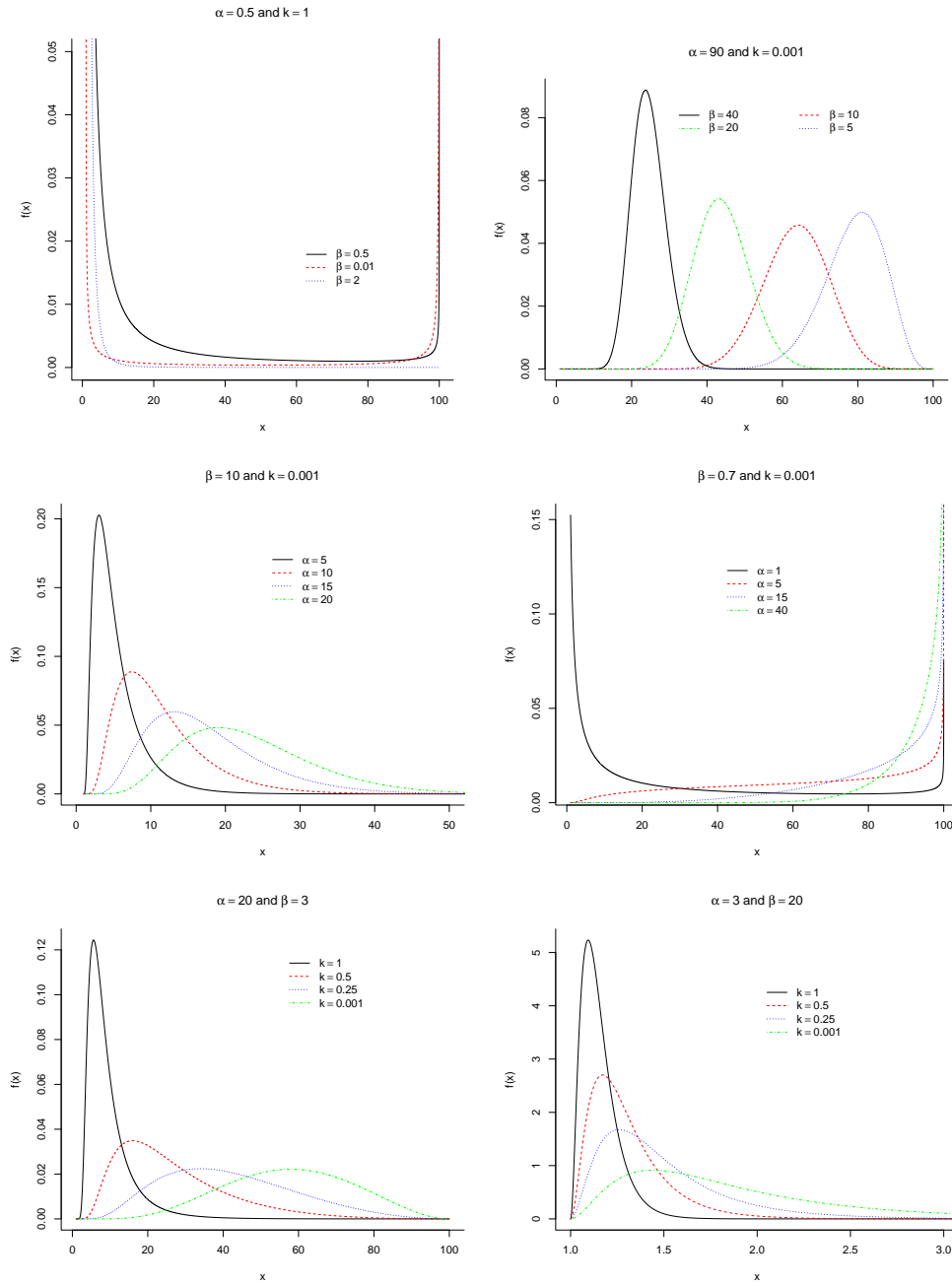Figure 1. Density function for $\theta = 1$, $\lambda = 100$ and some values de α, β and k.

$$\lim_{x \to \lambda} f_X(x) = \begin{cases} \infty & , \ 0 < \beta < 1 \\ \frac{\alpha k \theta^k}{\lambda^{k+1}\left[1-\left(\frac{\theta}{\lambda}\right)^k\right]} & , \ \beta = 1 \\ 0 & , \ \beta > 1 \end{cases} \tag{8}$$

Proof : Obtained from the pdf presented in Equation 5 $\qquad\qquad\square$

It is worth noting in Equation 7 that if $x \to \theta$ then the parameter α defines the limit behavior of the density. If $x \to \lambda$ then the parameter β controls the limit behavior, as shown in Equation 8.

2.3. *Moment Generating function of BTPD*

In this section the Moment Generating function for the BTPD is presented. We will make use of the following result.

Lemma 2.2: *The power serie $\sum_{i=0}^{\infty} \frac{(r/k)_i}{i!} u^i \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i$, is uniformly convergent in the interval $[0, 1]$.*

Proof: The radius of convergence, l, of a power serie, $\sum a_n x^n$, is given by $l^{-1} = \lim_{i\to\infty} \sup \sqrt[i]{a_i}$. In this case,

$$l^{-1} = \lim_{i\to\infty} \sup \sqrt[i]{\frac{(r/k)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i}$$

$$= \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right] \lim_{i\to\infty} \sup \sqrt[i]{\frac{(r/k)_i}{i!}}$$

But $\lim_{i\to\infty} \sup \sqrt[i]{\frac{(r/k)_i}{i!}}$ is the radius of convergence of the well known power serie $\sum_{i=0}^{\infty} (-1)^i \binom{r/k}{i} u^i = (1-u)^{r/k}$. Thus $\lim_{i\to\infty} \sup \sqrt[i]{\frac{(r/k)_i}{i!}} = 1$. Thereof $l = \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^{-1}$. So, since $l > 1$, the closed interval $[0, 1] \subset (-l, l)$. Therefore the serie is uniformly convergent in the interval $[0, 1]$. $\qquad\square$

The moments of BTPD are presented in the following proposition.

Proposition 2.3: *The moments of the BTPD are given by*

$$E[X^r] = \frac{\theta^r}{B(\alpha, \beta)} \sum_{i=0}^{\infty} \frac{(r/k)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i B(\alpha + i, \beta) \qquad (9)$$

Proof: : Using the transformation $u = F_Y(x)$, it can be shown that

$$E[X^r] = \int_{\theta}^{\lambda} \frac{1}{B(\alpha,\beta)} \left[\frac{1 - \left(\frac{\theta}{x}\right)^k}{1 - \left(\frac{\theta}{\lambda}\right)^k}\right]^{\alpha-1} \left[1 - \frac{1 - \left(\frac{\theta}{x}\right)^k}{1 - \left(\frac{\theta}{\lambda}\right)^k}\right]^{\beta-1} \frac{\frac{k\theta^k}{x^{k+1-r}}}{1 - \left(\frac{\theta}{\lambda}\right)^k} dx$$

$$= \frac{\theta^r}{B(\alpha, \beta)} \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} \left\{1 - u\left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]\right\}^{-r/k} du$$

$$= \frac{\theta^r}{B(\alpha, \beta)} \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} \sum_{i=0}^{\infty} \frac{(r/k)_i}{i!} u^i \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i du$$

$$= \frac{\theta^r}{B(\alpha, \beta)} \sum_{i=0}^{\infty} \frac{(r/k)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i B(\alpha + i, \beta)$$

$$\square$$

Note that $\int_{\theta}^{\lambda} x^r f(x) dx \leq \int_{\theta}^{\lambda} \lambda^r f(x) dx \leq \lambda^r$. So all the moments of the BTPD are finite.

Considering the above result we have the following.

Lemma 2.4: *The moment generating function of BTPD is*

$$M(t) = \frac{1}{B(\alpha,\beta)} \sum_{r=0}^{\infty} \sum_{j=0}^{\infty} \frac{(t\theta)^r}{r!} \frac{(r/k)_j}{j!} \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^j B(\alpha+j,\beta)$$

Proof: Since $\sum_{r=0}^{\infty} \frac{t^r}{r!} x^r f(x)$ converges and each term is integrable for all t close to 0, then we can rewrite the moment generating function as $M(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} E[X^r]$. By replacing $E[X^r]$ by the right side of Equation 9 the desired result is obtained. □

2.4. *Mean Deviations*

Consider the two following mean deviation:

i) Mean Deviation from the mean: $D(\mu) = E[|X - E(X)|]$.
ii) Mean Deviation from the median: $D(M) = E[|X - M|]$, where M is the median.

In general, the mean deviation from the mean is used for symmetric distributions while the mean deviation from the median is used for skewed distributions. For the BTPD density we have the following property.

Proposition 2.5: *The mean deviation from the mean and the mean deviation from the median for the Beta truncated Pareto density are given, respectively, by*

i) $D(\mu) = 2\mu F(\mu) - 2\frac{\theta}{B(\alpha,\beta)} \sum_{i=0}^{\infty} \frac{(1/k)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i B_{F_Y(E[X])}(\alpha+i,\beta)$.

ii) $D(M) = \mu - 2\frac{\theta}{B(\alpha,\beta)} \sum_{i=0}^{\infty} \frac{(1/k)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i B_{F_Y(M)}(\alpha+i,\beta)$.

Proof: Consider that

$$D(\mu) = \int_{\theta}^{\lambda} |x - \mu| f(x) dx$$

$$= 2\int_{\theta}^{\mu} (\mu - x) f(x) dx + \int_{\theta}^{\lambda} (x - \mu) f(x) dx$$

$$= 2\mu F(\mu) - 2\int_{\theta}^{\mu} x f(x) dx,$$

and, in a similar way,

$$D(M) = \mu - 2\int_{\theta}^{M} x f(x) dx.$$

Using the transformation $u = F_Y(x)$ and Lemma 2.2, then

$$\int_{\theta}^{c} x f(x) dx = \frac{\theta}{B(\alpha,\beta)} \int_{0}^{F_Y(c)} u^{\alpha-1} (1-u)^{\beta-1} \sum_{i=0}^{\infty} \frac{(1/k)_i}{i!} u^i \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i du$$

$$= \frac{\theta}{B(\alpha,\beta)} \sum_{i=0}^{\infty} \frac{(1/k)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i \int_{0}^{F_Y(c)} u^{\alpha+i-1} (1-u)^{\beta-1} du$$

$$= \frac{\theta}{B(\alpha,\beta)} \sum_{i=0}^{\infty} \frac{(1/k)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda}\right)^k\right]^i B_{F_Y(c)}(\alpha+i,\beta) \qquad (10)$$

Assuming that $c = \mu$ and $c = M$, the desired results are obtained. $\square$

2.5. Lorenz's Curve

The Lorenz's curve was proposed by [29] to study the distribution of income and wealth within the population. The Lorenz's curve was used to describe the proportion of population risk that falls below the p-th quantile of risk in [30].

The Lorenz's curve is defined as,

$$L(p) = \frac{1}{E[X]} \int_0^{F_X^{-1}(p)} t f(t) dt, \quad 0 \le p \le 1 \tag{11}$$

In the BTPD case we have the following.

Lemma 2.6: *The Lorenz's curve for the BTPD is given by*

$$L(p) = \frac{\theta}{E[X]B(\alpha,\beta)} \sum_{i=0}^{\infty} \frac{(1/k)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda} \right)^k \right]^i B_{F_Y(F_X^{-1}(p))}(\alpha + i, \beta) \tag{12}$$

Proof: Result obtained by making $c = F_X^{-1}(p)$ in Equation 10. $\square$

2.6. Entropies

In this section we present two entropy measures: the Renyi entropy and the Shannon's entropy.

2.6.1. Rényi's entropy

Consider the Rényi's entropy, defined as

$$I_R(\epsilon, f) = \frac{1}{1 - \epsilon} \log \left(\int_{\mathbb{R}} f^\epsilon(x) dx \right). \tag{13}$$

then, for the BTPD, we have the following proposition

Proposition 2.7 Rényi's entropy *The Rényi's entropy for the BTPD is given by*

$$I_R(\epsilon, f) = \log \left(\frac{\theta}{k} \right) - \frac{\epsilon}{1 - \epsilon} \log B(\alpha, \beta) +$$

$$+ \frac{1}{1 - \epsilon} \log \left\{ \sum_{i=0}^{\infty} \frac{\left(\frac{(\epsilon - 1)(k+1)}{k} \right)_i}{i!} \left[1 - \left(\frac{\theta}{\lambda} \right)^k \right]^{i+1-\epsilon} B(\alpha_1, \beta_1) \right\} \tag{14}$$

where $\alpha_1 = \epsilon(\alpha - 1) + i + 1$ and $\beta_1 = \epsilon(\beta - 1) + 1$.

Proof: Applying the transformation $u = F_Y(x)$, it can be shown that

$$\int_{\mathbb{R}} f^{\epsilon}(x)xdx = \int_{\theta}^{\lambda} \frac{1}{B^{\epsilon}(\alpha,\beta)} \left[\frac{1-\left(\frac{\theta}{x}\right)^k}{1-\left(\frac{\theta}{\lambda}\right)^k}\right]^{\epsilon(\alpha-1)} \left[1 - \frac{1-\left(\frac{\theta}{x}\right)^k}{1-\left(\frac{\theta}{\lambda}\right)^k}\right]^{\epsilon(\beta-1)} \left[\frac{\frac{k\theta^k}{x^{k+1}}}{1-\left(\frac{\theta}{\lambda}\right)^k}\right]^{\epsilon} dx$$

$$= \int_0^1 \frac{1}{B^{\epsilon}(\alpha,\beta)} u^{\epsilon(\alpha-1)} (1-u)^{\epsilon(\beta-1)} \left[\frac{\frac{k\theta^k}{x^{k+1}}}{1-\left(\frac{\theta}{\lambda}\right)^k}\right]^{\epsilon-1} du$$

$$= \frac{k^{\epsilon-1}\theta^{1-\epsilon}}{B^{\epsilon}(\alpha,\beta)\left[1-\left(\frac{\theta}{\lambda}\right)^k\right]^{\epsilon-1}} \times$$

$$\times \int_0^1 u^{\epsilon(\alpha-1)} (1-u)^{\epsilon(\beta-1)} \left\{1 - u\left[1-\left(\frac{\theta}{\lambda}\right)^k\right]\right\}^{(k+1)(\epsilon-1)/k} du$$

$$= \frac{k^{\epsilon-1}\theta^{1-\epsilon}}{B^{\epsilon}(\alpha,\beta)\left[1-\left(\frac{\theta}{\lambda}\right)^k\right]^{\epsilon-1}} \times$$

$$\times \sum_{i=0}^{\infty}(-1)^i\binom{\frac{(\epsilon-1)(k+1)}{k}}{i}\left[1-\left(\frac{\theta}{\lambda}\right)^k\right]^i B(\alpha_1,\beta_1)$$

then applying Equation 13, Equation 14 is obtained. Note that we made use of the Lemma 2.2 here. $\square$

2.6.2. Shannon's entropy

The Shannon's entropy [see 28, and references therein] is defined as

$$I_{Sh}(f) = -\int_{\mathbb{R}} f(x)\log f(x)dx.$$

Was proved by [28] that for the Beta generated family the Shannon's entropy is given by

$$I_{Sh}(f) = \log B(\alpha,\beta) - (\alpha-1)\left[\psi(\alpha)-\psi(\alpha+\beta)\right] - (\beta-1)\left[\psi(\beta)-\psi(\alpha+\beta)\right] - $$
$$-E\left[\log f_Y\left(F_Y^{-1}(Z)\right)\right]. \tag{15}$$

We present the following result for BTPD.

Lemma 2.8: *For the BTPD and following Equation 15, it can be shown that*

$$E\left[\log f_Y\left(F_Y^{-1}(Y)\right)\right] = \log\frac{k}{\theta} - \log\left[1-(\theta/\lambda)^k\right] - \sum_{i=1}^{\infty}\frac{\left[1-(\theta/\lambda)^k\right]^i}{iB(\alpha,\beta)}B(\alpha+i,\beta).$$

Proof:

$$E\left[\log f_Y\left(F_Y^{-1}(Y)\right)\right] = \log k + k\log\theta - \log\left[1-(\theta/\lambda)^k\right] - (k+1)\log\theta + $$
$$+\frac{k+1}{k}E\left[\log\left\{1-Y\left[1-(\theta/\lambda)^k\right]\right\}\right]$$

and

$$E\left[\log\left\{1-Y\left[1-(\theta/\lambda)^k\right]\right\}\right]=\int_0^1\log\left\{1-Y\left[1-(\theta/\lambda)^k\right]\right\}\frac{1}{B(\alpha,\beta)}y^{\alpha-1}(1-y)^{\beta-1}dy$$

$$=\int_0^1\sum_{i=1}^{\infty}\frac{y^i\left[1-(\theta/\lambda)^k\right]^i}{i}\frac{1}{B(\alpha,\beta)}y^{\alpha-1}(1-y)^{\beta-1}dy$$

$$=-\sum_{i=1}^{\infty}\frac{\left[1-(\theta/\lambda)^k\right]^i}{iB(\alpha,\beta)}B(\alpha+i,\beta)$$

□

2.7. *Hazard Rate Function*

It was showed in Figure 1 some of different forms that the density of BTPD is capable to assume. It is not different for the Hazard Rate function as it is shown here.

Lemma 2.9: *For BTPD, the Hazard Rate function has the form*

$$h(x)=\frac{f(x)}{1-F(x)}$$

$$=\frac{\left[\frac{1-\left(\frac{\theta}{x}\right)^k}{1-\left(\frac{\theta}{\lambda}\right)^k}\right]^{\alpha-1}\left[1-\frac{1-\left(\frac{\theta}{x}\right)^k}{1-\left(\frac{\theta}{\lambda}\right)^k}\right]^{\beta-1}\frac{k\theta^k x^{-k-1}}{1-\left(\frac{\theta}{\lambda}\right)^k}}{B(\alpha,\beta)-B_{F_Y(x)}(\alpha,\beta)}\tag{16}$$

The limit of the Hazard Rate function is given by

$$\lim_{x\to\theta}h(x)=\begin{cases}\infty & ,\ 0<\alpha<1\\\frac{\beta k}{\theta\left[1-\left(\frac{\theta}{\lambda}\right)^k\right]} & ,\ \alpha=1\\0 & ,\ \alpha>1\end{cases}\tag{17}$$

$$\lim_{x\to\lambda}h(x)=\infty\tag{18}$$

Some forms of Hazard Rate function of BTPD are shown in Figure 2.

3. **Inference Issues About the BTPD**

The maximum likelihood method for point estimation is the most used method in literature. Alternatively, estimates for the parameters can be found using the method of moments. In this case, estimates are obtained by solving the following non-linear equations

$$\frac{1}{n}\sum_{i=1}^{n}X^r=\frac{\theta^r}{B(\alpha,\beta)}\sum_{i=0}^{\infty}\frac{(r/k)_i}{i!}\left[1-\left(\frac{\theta}{\lambda}\right)^k\right]^i B(\alpha+i,\beta),\quad r=1,\ldots,5\tag{19}$$
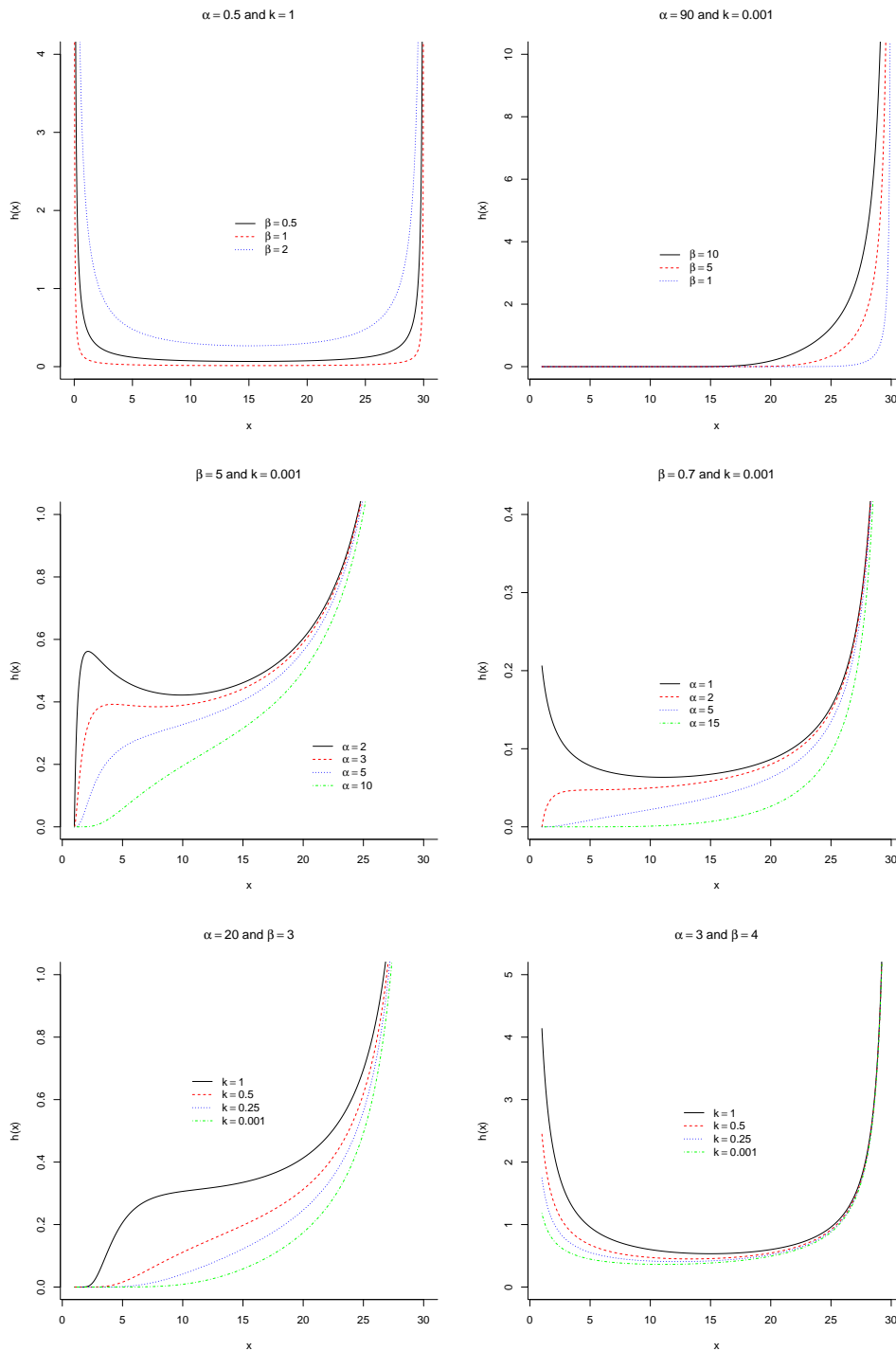
Figure 2. Hazard Hate function for $\theta = 0.01$, $\lambda = 30$ and some values de α, β and k.

A modified method of moments was proposed by [28]. Nevertheless, the method of moments is not a very reliable approach and it is usually used as initial estimative for the maximum likelihood approach. The main advantage of maximum likelihood estimators is that, under some regularity conditions, they have desirable properties. However, these conditions are not satisfied here.

The empirical cdf is generally used in the literature to test whether a parametric distribution fits the data. But also, there are some works in literature which use a empirical distribution function in order to estimate the parameters. Because the idea

behind the method is minimizing the distance between the empirical cdf and the theoritical cdf, this method was called estimation by the minimum distance method (MDM) [see 31, and references therein]. Also, [32] shows the consistency and found the asymptotic distribution of the estimator of minimum distance (MDE). Through a simulation study of many different MDM statistics, [33] analysed the perform of MDE for generalized Pareto distribution and concluded that MDE had a very good performance. Thus, we also apply this method to estimate the parameters of the BTPD.

The BTPD has $f_X(x) > 0$, only if $\theta < x < \lambda$. Estimates for θ and λ are: $\hat{\theta} = X_{(1)}$ and $\hat{\lambda} = X_{(n)}$, where $X_{(1)}$ is the first order statistic and $X_{(n)}$ is the last one. The estimates for α, β and k are chosen so that the maximum distance between the empirical cdf and the BTPD cdf (called by Kolmogorov distance in [33]) is minimized, i.e.,

$$(\hat{\alpha}, \hat{\beta}, \hat{k}) = \arg \min_{\alpha, \beta, k} \max_{1 \leq i \leq n} \left| F_n(x_i) - F_X\left(x_i \Big| \hat{\theta}, \hat{\lambda}\right) \right| \tag{20}$$

where $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(X_i \leq x)}$.

Proposition 3.1: *The estimators $\hat{\theta} = X_{(1)}$ and $\hat{\lambda} = X_{(n)}$ are consistent.*

Proof: For any $\epsilon > 0$ we have

$$P\left[\left|\hat{\theta} - \theta\right| < \epsilon\right] = P\left[\theta - \epsilon < X_{(1)} < \theta + \epsilon\right]$$
$$= F_{X_{(1)}}(\theta + \epsilon) - F_{X_{(1)}}(\theta - \epsilon)$$
$$= [1 - F_X(\theta - \epsilon)]^n - [1 - F_X(\theta + \epsilon)]^n$$

Since the support of f_X is in the interval (θ, λ), we have that $F_X(\theta - \epsilon) = 0$ and $F_X(\theta + \epsilon) > 0$. Thus $\lim_{n \to \infty} P\left[\left|\hat{\theta} - \theta\right| < \epsilon\right] = 1$.

In an analogous way we have,

$$\lim_{n \to \infty} P\left[\left|\hat{\lambda} - \lambda\right| < \epsilon\right] = \lim_{n \to \infty} \left\{[F_X(\lambda + \epsilon)]^n - [F_X(\lambda - \epsilon)]^n\right\}$$
$$= 1$$

$\square$

4. Applications

In order to illustrate the use and the performance of the BTPD proposal, we applied the proposed distribution to real data and we compared the results with the following distributions:

   (i) The truncated Pareto distribution (TPD).
       The density of this distribution is presented in Equation 4.
   (ii) The Generalized Pareto distribution (GPD).

The GP distribution has the following pdf

$$f(x) = \frac{1}{a}\left(1 - \frac{kx}{a}\right)^{(1-k)/k} \qquad 0 < x \text{ if } k \leq 0$$

$$0 < x < a/k \text{ if } k > 0$$

(iii) The Weibull distribution with three parameters (TPWD). The density of this distribution is given by

$$f(x) = \gamma b^{-\gamma}(x-a)^{-\gamma-1}\exp\left\{-\left(\frac{x-a}{b}\right)^{\gamma}\right\}, x > a \quad \gamma, b > 0.$$

(iv) The Beta Pareto distribution (BPD). The density of this distribution is presented in Equation 6.

In order to compare the models we will use the K-S statistic given by

$$\text{KS}\left(F_{est}\right) = \max_{1 \leq i \leq n} |F_n(x_i) - F_{est}\left(x_i\right)|, \tag{21}$$

where F_{est} is the model's c.d.f using the estimated parameters. We will also provide the absolute mean distance statistic given by

$$D\left(F_{est}\right) = \frac{1}{n}\sum_{i=1}^{n}|F_n(x_i) - F_{est}\left(x_i\right)| \tag{22}$$

4.1. *Castelo River*

The data consist of 684 maximum values of monthly flood rates of the Castelo River, in Brazil, and it is available online (hidroweb.ana.gov.br). Table 1 shows summary statistics of the data. The Figure 3 shows the histogram and box plot of the data. It is worth noticing the strong asymmetry of the distribution of the data to the right. This behavior is quite common in hydrology data.

Table 1. Descriptive statistics of maximum monthly flood rate of Castelo River.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.
0.740	4.602	6.555	7.704	9.678	33.4	4.6178

The Table 2 and Figure 4 present the results obtained by the MDE. The K-S and D statistics present very high values for TPD and GPD, suggesting a quite poor adjust to the data. Despite this improvement, the KS-statistic remained very high. The K-S test rejected both, TPD and GPD distributions, as appropriate models. For the TPWD, the D and K-S statistics are better than for the TPD and GPD. The K-S test now does not reject the TPWD as an appropriate model at 5% of significance.

For the Beta Pareto distribution and BTPD the K-S test does not reject either models, even at 10% of significance. The Figure 4 shows the empirical cdf and fitted cdf for all models.
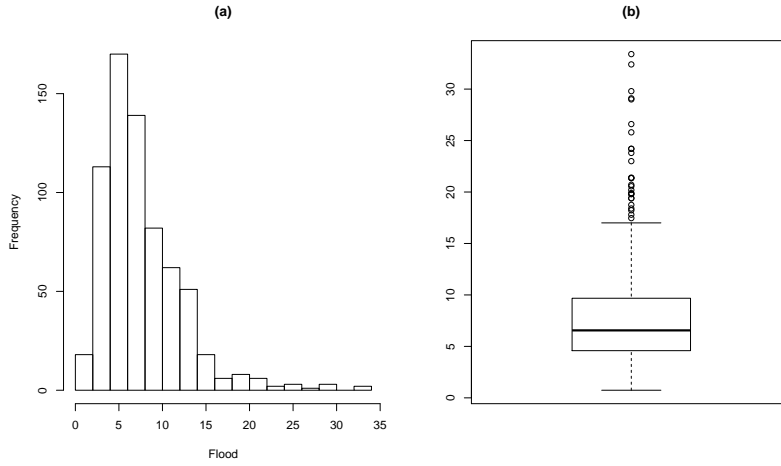
Figure 3. Castelo River flood: (a) Histogram (b) Box-Plot.

Table 2.   KS Estimative of parameters and KS and D statistics.

	TPD	GPD	TPWD	BPD	BTPD
Estimatives	$\hat{k} = \frac{3}{10^{12}}$;	$\hat{a} = 12.0612$; $\hat{k} = 0.3397$;	$\hat{a} = 0.74$; $\hat{b} = 7.4360$; $\hat{\gamma} = 1.8658$;	$\hat{\alpha} = 15.79$; $\hat{\beta} = 64.14$; $\hat{k} = 0.1$;	$\hat{\alpha} = 9.8228$; $\hat{\beta} = 3.9035$; $\hat{k} = 0.347$;
KS	0.3005	0.1586	0.0399	0.0276	0.0182
D	0.1438	0.0995	0.0207	0.0116	0.0082

5.   Conclusion

In this work a new family of distribution called Beta truncated Pareto distribution was introduced and some of its properties were analysed. The BTPD proved to be a very flexible model with many different forms for its density (as it can be seen in Figure 1). Also, the BTPD includes some other families that were already proposed in literature. For the parameters estimation we used the minimum distance method.

In order to evaluate the BTPD performance we applied the BTPD to real data. The TPD and GPD a very poor fit. The TPWD had a good fit was not rejected by the KS test. The BTPD and BPD have the best fit to the data with a small advantage for BTPD. They both followed the empirical distribution very closely when the MDE were applied.

Due to the fact that the BTPD has two more paramaters than the TPWD, one could be favorable to TPWD arguing that the difference between the adjust obtained by the TPWD and the adjust obtained by the BTPD is small. That is not so clear. Despite the KS test does not reject either, we have a sample size of 684 which makes the *cost* of the two additional parameters not so high. Beyond that the K-S statistic for the BTPD is less than a half of the KS statistic for the TPWD.

References

[1] B. C. Arnold, *Pareto Distribution*, Fairland: International Co-operative Publishing House (1983).
[2] B. Gutenberg and R. F. Richter, *Frequency of earthquakes in california.*, Bulletin of the Seismological Society of America (1944).
[3] E. T. Lu and R. J. Hamilton, *Avalanches of the distribution of solar flares.*, Astrophysical Journal (1991).
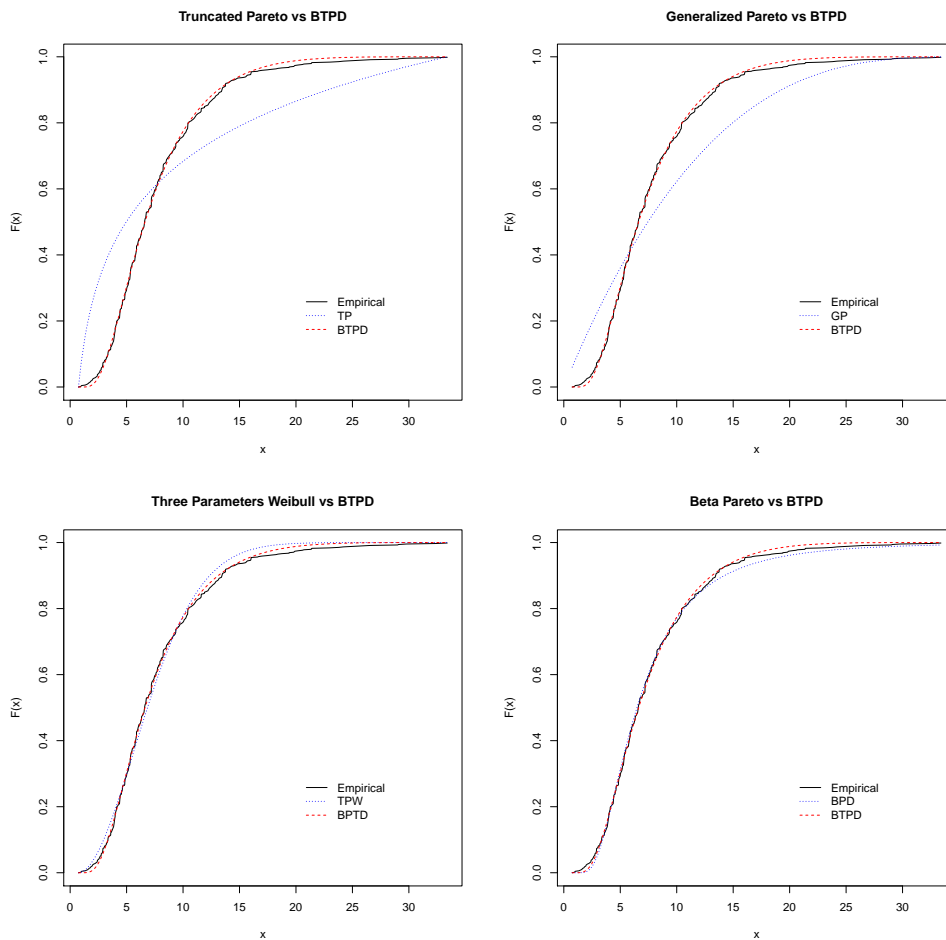
Figure 4. Empirical cdf and fitted cdf for the truncated Pareto, Three Parameter Weibull, Generalized Pareto, Beta Pareto and BTPD distributions, for the Castelo River Data, using MDM

[4] M. S. Wheatland and P. A. Sturrock, *Avalanche models of solar flares and the distribution of active regions.*, Astrophysical Journal (1991).

[5] G. K. Zipf, *Human Behaviour and the Principle of Least Effort.*, Addison-Wesley, Cambridge (1949).

[6] D. H. Zanette and S. C. Manrubia, *Vertical transmission of culture and the distribution of family names.*, Physic A (2001).

[7] A. J. Lotka, *The frequency distribution of scientific production.*, J. Wash. Acad. Sci. (1926).

[8] D. J. de S. Price, *Networks of scientific papers.*, Science (1965).

[9] B. D. Malamud and D. L. Turcotte, *The applicability of power-law frequency statistics to floods*, Journal of Hydrology (2006).

[10] O. S. Klass, O. Biham, M. Levy, O. Malcai and S. Solomon, *The Forbes 400 and the Pareto wealth distribution*, Economics Letters (2006).

[11] A. Jayadev, *A power law tail in India's wealth distribution: Evidence from survey data*, Physic A (2008).

[12] S. Sinha, *Evidence for power-law tail of the wealth distribution in India*, Physic A (2006).

[13] R. V. Hogg, J. W. Mckean and A. T. Craig *Introduction to Mathematical Statistics*, 6th ed. Pearson Prentice-Hall, New Jersey (2005).

[14] I. B. Aban, M. M. Meerschaert and A. K. Panorska *Paramete Estimation for the truncated Pareto Distribution*, Jounal of the American Statistical Association (2006).

[15] L. Zaninetti and M. Ferraro *On the truncated Pareto distribution with application*, Central European Journal of Physics (2008).

[16] W. J. Reed, *The Pareto, Zipf and other power laws*, Economics Letter (2001).

[17] W. J. Reed, *The Pareto law of incomes - an explanation and an extension*, Physic A (2003).

[18] J. R. M. Hosking and J. R. Wallis, *Parameter and quantile estimation for the generalized Pareto distribution.* Technometrics (1987).

[19] S. D. Grimshaw, *Computing maximum likelihood estimates for the generalized Pareto distribution.*, Technometrics (1993).

[20] R. A. Lockhart and M. A. Stephens *Estimation and Tests of Fit for the Three-parameter Weibull distribution*, Journal of the Royal Statistical Society. B (1994).

[21] J. M. F. Carrasco, E. M. M. Ortega and G. M. Cordeiro *A generalized modified Weibull distribution for lifetime modeling*, Computational Statistics and Data Analysis (2008).

[22] A. C. Cohen and B. J. Whitten *Modified moment and maximum likelihood estimators for parameters*

of the three-parameter gamma distribution, Communications in Statistics - Simulation and Computation (1982).

[23] R. D. Gupta and D. Kundu *Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions*, Biometrical Journal (2001).

[24] W. Nelson, *Applied Life Data Analysis*, John Wiley and Sons Inc. (1982), New York.

[25] N. Eugene, C. Lee and F. Famoye, *B*eta-Normal Distributions and Its Applications, Communications in Statistics - Theory and Methods (2002), pp. 497–512

[26] M. C. Jones, *Families of distributions arising from distributions of orders statistics*, TEST vol. 13 (2004), pp. 1–43.

[27] A. Akinsete, F. Famoye and C. Lee, *The beta-Pareto distribution*, Statistcs (2008), pp. 547–563

[28] K. Zografos and N. Balakrishnan, *On families of beta- and generalized gamma-generated distributions and associated inference*, Statistical Methodology (2009).

[29] M. O. Lorenz *Methods for measuring concentration of wealth.*, Journal of American Statistical Association (1905).

[30] M. H. Gail and R. M. Pfeiffer *On criteria for evaluating models of absolute risk*, Biostatistics (2005).

[31] J. Wolfowitz *Estimation by the minimum distance method*, Ann. Inst. Statist. Math (1953).

[32] D. Pollard *The minimum distance method of testing*, Metrika (1980).

[33] A. Luceño *Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators*, Computational Statistics & Data Analysis (2006).