STATISTICS IN TRANSITION-new series, August 2011
Vol. 12, No. 1, pp. 97—114

THE BETA PARETO DISTRIBUTION

AHMED HURAIRAH[1]

ABSTRACT

In this paper, we introduce a generalization—referred to as the beta Pareto distribution, generated from the logit of a beta random variable. We provide a comprehensive treatment of the mathematical properties of the beta Pareto distribution. We derive expressions for the *kth* moments of the distribution, variance, skewness, kurtosis, mean deviation about the mean, mean deviation about the median, Rényi entropy, Shannon entropy. We also discuss simulation issues, estimation of parameters by the methods of moments and maximum likelihood.

Key words: Beta Pareto distribution; Kurtosis; Rényi entropy

1. Introduction

The generalization is motivated by the following general class: If *G* denotes the cumulative distribution function (cdf) of a random variable then a generalized class of distributions can be defined by

$$F(x) = I_{G(x)}(a, b) \tag{1}$$

for $a > 0$ and $b > 0$, where

$$I_{y(a,b)} = \frac{B_y(a,b)}{B(a,b)},$$

denotes the incomplete beta function ratio, and

$$B_y(a, b) = \int_0^y x^{a-1}(1-x)^{b-1} \ dx$$

[1] Department of Statistics, Sana'a University, Yemen. E-mail: Hurairah69@yahoo.com.

denotes the incomplete beta function. This class of distributions came to prominence after the recent paper by Jones (2004). Eugene et al. (2002) introduced what is known as the beta normal distribution by taking *G* in (1) to be the cdf of the normal distribution with parameters μ and σ. The only properties of the beta normal distribution known are some first moments derived by Eugene et al. (2002) and some more general moment expressions derived by Gupta and Nadarajah (2004). More recently, Natarajahand and Kotz (2004) introduced what is known as the beta Gumbel distribution by taking *G* in (1) to be the cdf of the Gumbel distribution with parameters μ and σ. This distribution is a little more tractable than the beta normal distribution in that Natarajah and Kotz (2004) were able to provide closed-form expressions for the moments, the asymptotic distribution of the extreme order statistics and the estimation procedure. Another distribution that happens to belong to (1) class is the log *F* (or beta logistic) distribution. This distribution appears reasonably tractable partly because it has been around for over 20 years [5], but it did not originate directly from (1) idea. In this paper we will introduce the beta Pareto (BP) distribution by taking *G* in (1) to be the cdf of Pareto distribution with parameter λ. The cdf of the BP distribution becomes

$$F(x) = I_{\left(\frac{\lambda}{x}\right)^c}(a,b) \tag{2}$$

for $x \geq \lambda$, $\lambda > 0$, $c > 0$, $a > 0$ and $b > 0$. The corresponding probability density function (pdf) and the hazard rate function associated with (2) are:

$$f(x) = \frac{c\,x^{-1}}{B(a,b)}\left(\frac{\lambda}{x}\right)^{ac}\left(1-\left(\frac{\lambda}{x}\right)^{c}\right)^{b-1}, \quad x \geq \lambda \tag{3}$$

and

$$h(x) = \frac{c\,x^{-1}\left(\frac{\lambda}{x}\right)^{ac}\left(1-\left(\frac{\lambda}{x}\right)^{c}\right)^{b-1}}{B_{\left(\frac{\lambda}{x}\right)^c}(a,b)} \tag{4}$$

The shapes of the probability density function (3) and the hazard rate function (4) for the beta Pareto distributions are given in Figs. 1 and 2, respectively.

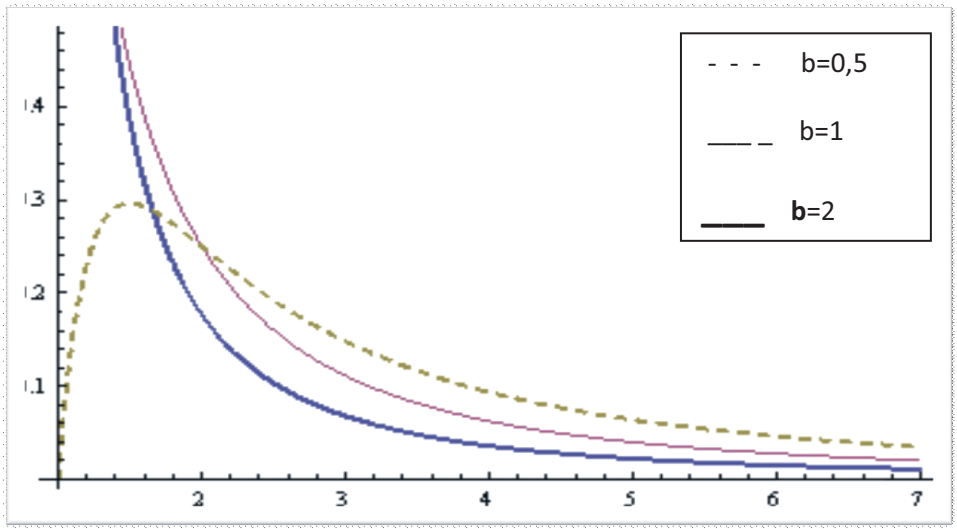

Figure. 1. The beta Pareto pdf (3) for $a = 1,\ \lambda = 1\ and\ c = 1$

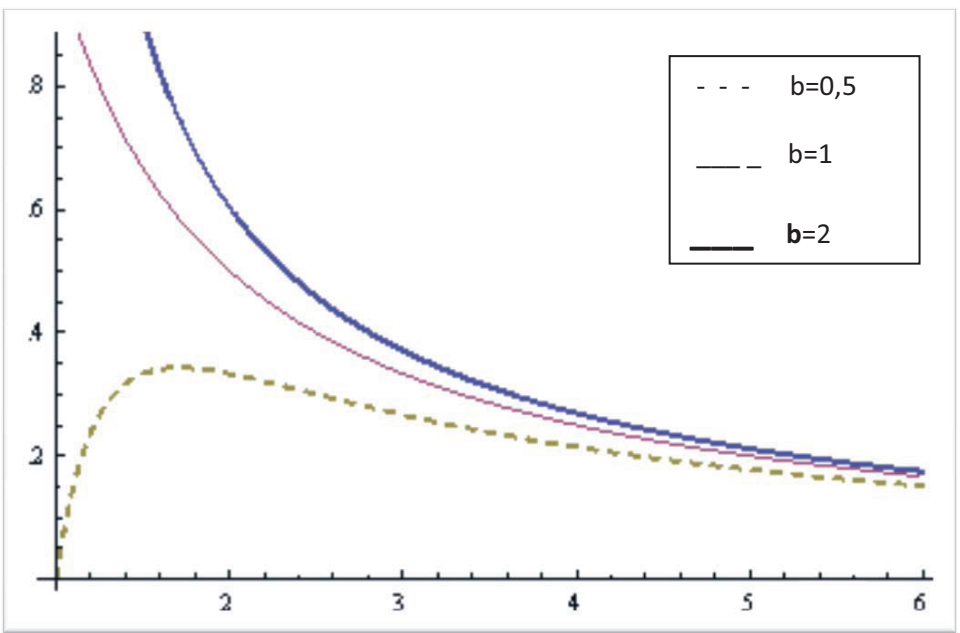

Figure. 2. The beta Pareto hazard rate function pdf (4)
for $a = 1,\ \lambda = 1\ and\ c = 1$

2. Moments

The *kth* moments of x can be written as

$$E\left(x^k\right) = \frac{\lambda^k \Gamma(a+b)\Gamma\left(a-\frac{k}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{k}{c}\right)} \tag{5}$$

In particular, the first four moments can be worked out as

$$E(x) = \frac{\lambda\, \Gamma(a+b)\, \Gamma\left(a-\frac{1}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{1}{c}\right)} \tag{6}$$

$$E(x^2) = \frac{\lambda^2 \Gamma(a+b)\Gamma\left(a-\frac{2}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{2}{c}\right)} \tag{7}$$

$$E(x^3) = \frac{\lambda^3 \Gamma(a+b)\Gamma\left(a-\frac{3}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{3}{c}\right)} \tag{8}$$

and

$$E(x^4) = \frac{\lambda^4 \Gamma(a+b)\Gamma\left(a-\frac{4}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{4}{c}\right)} \tag{9}$$

The first three entral moments, skewness and the kurtosis of x can be given by

$$Var(x) = E(x^2) - [E(x)]^2$$
(10)

$$Var(x) = \frac{\lambda^2 \Gamma(a+b)\Gamma\left(a-\frac{2}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{2}{c}\right)} - \left[\frac{\lambda\, \Gamma(a+b)\Gamma\left(a-\frac{1}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{1}{c}\right)}\right]^2 \tag{11}$$

$$E[x - E(x)]^3 = \frac{\lambda^3 \Gamma(a+b)\Gamma\left(a-\frac{3}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{3}{c}\right)} + \frac{2\lambda^3 \Gamma(a+b)^3 \Gamma\left(a-\frac{1}{c}\right)^3}{\Gamma(a)^3 \Gamma\left(a+b-\frac{1}{c}\right)^3} - \frac{3\lambda^3 \Gamma(a+b)^2 \Gamma\left(a-\frac{1}{c}\right)\Gamma\left(a-\frac{2}{c}\right)}{\Gamma(a)^2 \Gamma\left(a+b-\frac{1}{c}\right)\Gamma\left(a+b-\frac{2}{c}\right)}$$

(12)

$$E[x-E(x)]^4 = \frac{\lambda^4\Gamma(a+b)\Gamma\left(a-\frac{4}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{4}{c}\right)} - \frac{3\lambda^4\Gamma(a+b)^4\Gamma\left(a-\frac{1}{c}\right)^4}{\Gamma(a)^4\Gamma\left(a+b-\frac{1}{c}\right)^4} + \frac{6\lambda^4\Gamma(a+b)^3\Gamma\left(a-\frac{2}{c}\right)\Gamma\left(a-\frac{1}{c}\right)^2}{\Gamma(a)^3\Gamma\left(a+b-\frac{2}{c}\right)\Gamma\left(a+b-\frac{1}{c}\right)^2} - \frac{4\lambda^4\Gamma(a+b)^2\Gamma\left(a-\frac{3}{c}\right)\Gamma\left(a-\frac{1}{c}\right)}{\Gamma(a)^2\Gamma\left(a+b-\frac{3}{c}\right)\Gamma\left(a+b-\frac{1}{c}\right)} \tag{13}$$

$$Skewness(x) = \frac{\lambda^3\Gamma(a+b)}{\Gamma(a)^3}\left(\frac{\Gamma(a)^2\Gamma\left(a-\frac{3}{c}\right)}{\Gamma\left(a+b-\frac{3}{c}\right)} + \frac{2\Gamma(a+b)^2\Gamma\left(a-\frac{1}{c}\right)^3}{\Gamma\left(a+b-\frac{1}{c}\right)^3} - \left(3\Gamma(a)\Gamma(a+b)\Gamma\left(a-\frac{1}{c}\right)\Gamma\left(a-\frac{2}{c}\right)\right) \Big/ \left(\Gamma\left(a+b-\frac{2}{c}\right)\left(\frac{\lambda^2\Gamma(a+b)}{\Gamma(a)^2}\Gamma(a+b)\left(\frac{\Gamma(a)\Gamma\left(a-\frac{2}{c}\right)}{\Gamma\left(a+b-\frac{2}{c}\right)} - \frac{\Gamma(a+b)\Gamma\left(a-\frac{1}{c}\right)^2}{\Gamma\left(a+b-\frac{1}{c}\right)^2}\right)^{\frac{3}{2}}\Gamma\left(a+b-\frac{1}{c}\right)\right)\right) \tag{14}$$

$$Kurtosis(x) = \frac{1}{\Gamma(a)^4}\left(\frac{\left[\frac{\lambda^4\Gamma(a)^3\Gamma(a+b)\Gamma\left(a-\frac{4}{c}\right)}{\Gamma\left(a+b-\frac{4}{c}\right)} - \frac{3\lambda^4\Gamma(a+b)^4\Gamma\left(a-\frac{1}{c}\right)^4}{\Gamma\left(a+b-\frac{1}{c}\right)^4} + \frac{6\lambda^4\Gamma(a)\Gamma(a+b)^3\Gamma\left(a-\frac{2}{c}\right)\Gamma\left(a-\frac{1}{c}\right)^2}{\Gamma\left(a+b-\frac{2}{c}\right)\Gamma\left(a+b-\frac{1}{c}\right)^2} - \frac{4\lambda^4\Gamma(a)^2\Gamma(a+b)^2\Gamma\left(a-\frac{3}{c}\right)\Gamma\left(a-\frac{1}{c}\right)}{\Gamma\left(a+b-\frac{3}{c}\right)\Gamma\left(a+b-\frac{1}{c}\right)}\right]}{\left[\left(\frac{\lambda^2\Gamma(a)\Gamma(a+b)\Gamma\left(a-\frac{2}{c}\right)}{\Gamma\left(a+b-\frac{2}{c}\right)} - \frac{\lambda^2\Gamma(a+b)^2\Gamma\left(a-\frac{1}{c}\right)^2}{\Gamma\left(a+b-\frac{1}{c}\right)^2}\right)^2\right]}\right) \tag{15}$$

Note that the skewness and kurtosis measures depend on a, b, λ and c .

3. Mean deviations

The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean and median. These are known as the mean deviation about the mean and the mean deviation about the median defined by

$$\delta_1(x) = \int_{\lambda}^{\infty} |x-\mu| f(x)dx$$

and

$$\delta_2(x) = \int_{\lambda}^{\infty} |x - M| f(x) dx$$

where $\mu = E(x)$ and M denotes the median. These measures can be calculated using the relationships that

$$\delta_1(x) = \int_{\lambda}^{\mu} (\mu - x) f(x) dx + \int_{\mu}^{\infty} (x - \mu)\, f(x) dx$$

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

$$= 2\{\mu F(\mu) - \int_{\lambda}^{\mu} x f(x) dx\} \tag{16}$$

and

$$\delta_2(x) = \int_{\lambda}^{M} (M - x) f(x) dx + \int_{M}^{\infty} (x - M)\, f(x) dx$$

$$= 2MF(M) - M - \int_{\lambda}^{M} x f(x) dx + \int_{M}^{\infty} x\, f(x) dx$$

$$= E(x) + 2MF(M) - M - 2 \int_{\lambda}^{M} x f(x) dx \tag{17}$$

$$\int_{\lambda}^{m} x\, f(x) dx = \frac{c}{B(a,b)} \int_{\lambda}^{m} \left(\frac{\lambda}{X}\right)^{ac} \left(1 - \left(\frac{\lambda}{X}\right)^{c}\right)^{b-1} dx$$

$$= \frac{1}{B[a,b]} \lambda \left(B_1\left[a - \frac{1}{c}, b\right] - \lambda^{ac} \left(\left(\frac{1}{m}\right)^{c}\right)^{-a+\frac{1}{c}} \left(\frac{1}{\lambda\, m}\right)^{ac} m\, B_{\left(\frac{\lambda}{m}\right)^{c}}\left[a - \frac{1}{c}, b\right]\right)$$

(18)

Substituting (18) into (16) and (17), one obtains the following expressions for the mean deviations:

$$\delta_1(x) = 2\left\{\mu\, I_{\left(\frac{\lambda}{\mu}\right)^{c}}(a,b) - \frac{\lambda B_1\left[a - \frac{1}{c}, b\right]}{B[a,b]} + \frac{\lambda^{ca+1} \left(\left(\frac{1}{\mu}\right)^{c}\right)^{-a+\frac{1}{c}} \left(\frac{1}{\lambda\, \mu}\right)^{a\, c} \mu\, B_{\left(\frac{\lambda}{\mu}\right)^{c}}\left[a - \frac{1}{c}, b\right]}{B[a,b]}\right\}$$

and

$$\delta_2(x) = \frac{\lambda\,\Gamma(a+b)\,\Gamma\left(a-\frac{1}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{1}{c}\right)} + 2MI_{\left(\frac{\lambda}{M}\right)^c}(a,b) - M - 2\left\{\frac{\lambda B_1\left[a-\frac{1}{c},b\right]}{B[a,b]} - \frac{\lambda\left(\left(\frac{1}{\mu}\right)^c\right)^{-a+\frac{1}{c}}\left(\frac{1}{\lambda\,\mu}\right)^{a\,c}\mu\, B_{\left(\frac{\lambda}{\mu}\right)^{\alpha}}\left[a-\frac{1}{c},b\right]}{c}\right\}$$

It can be verified that in the particular case $a = 1$ and $b = 1$, $\delta_1(x)$ reduces to $2c\lambda(c-1)^{-1}\left(1-\frac{\lambda}{\mu}\right)^{c-1}$, which is the mean deviation for the Pareto distribution.

4. Rényi and Shannon entropies

An entropy of a random variable x is a measure of variation of the uncertainty. Rényi entropy is defined by

$$\mathcal{J}_R(\gamma) = \frac{1}{1-\gamma}\, log[\int f^{\gamma}(x)dx] \tag{19}$$

where $\gamma > 0$ and $\gamma \neq 1$ Rényi (1961). For the beta Pareto pdf given by (3)

$$\int_{\lambda}^{\infty} f^{\gamma}(x)dx = \frac{c^{\gamma}}{B^{\gamma}(a,b)}\int_{\lambda}^{\infty} x^{-\gamma}\left(\frac{\lambda}{x}\right)^{\gamma ac}\left(1-\left(\frac{\lambda}{x}\right)^c\right)^{\gamma b-\gamma} dx$$

$$= \frac{\left(\frac{c}{\lambda}\right)^{\gamma-1} B\left(\gamma b-\gamma+1, \frac{\gamma ac+\gamma-1}{c}\right)}{B^{\gamma}(a,b)}$$

The Rènyi entropy, (19) takes the expression

$$\mathcal{J}_R(\gamma) = -\log\left(\frac{c}{\lambda}\right) + \frac{1}{1-\gamma}\, log\left[\frac{\left(\frac{c}{\lambda}\right)^{\gamma-1} B\left(\gamma b-\gamma+1, \frac{\gamma ac+\gamma-1}{c}\right)}{B^{\gamma}(a,b)}\right] \tag{20}$$

Shannon entropy defined by $E[-logf(x)]$ is the particular case of (19) for $\gamma \uparrow 1$. Limiting $\gamma \uparrow 1$ in (20) and using L'Hospital's rule, one obtains

$$E[-logf(x)] = -logc + logB(a,b) + log\left(\frac{\lambda\,\Gamma(a+b)\,\Gamma\left(a-\frac{1}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{1}{c}\right)}\right)(c(a-b+1)+1) - c\, log\lambda(a-b+1)$$

Song (2001) observed that the gradient of the Rényi entropy $\mathcal{J}_R^{'}(\gamma) = \left(\frac{d}{d\gamma}\right)\mathcal{J}_R(\gamma)$ is related to the log likelihood by $\mathcal{J}_R^{'}(1) = -\left(\frac{1}{2}\right)Var\left[log\left(f(x)\right)\right]$. This equality and the fact that the quantity $-\acute{\mathcal{J}}_R(1)$ remains invariant under location and scale transformations motivated Song to propose $-2\,\acute{\mathcal{J}}_R(1)$ as a measure of the shape of a distribution. From (20), the first derivative is

$$\mathcal{J}_R^{'}(\gamma) = \Bigg(c\,B\left(1+(b-1)\gamma, \tfrac{\gamma ac+\gamma-1}{c}\right)\Bigg(B^{\gamma}(a,b)\Bigg((\gamma-1)log\left(\tfrac{c}{\lambda}\right) + log\left[\frac{\left(\frac{c}{\lambda}\right)^{\gamma-1}B\left(1+(b-1)\gamma, \frac{\gamma ac+\gamma-1}{c}\right)}{B^{\gamma}(a,b)}\right]\Bigg) + (\gamma-1)\left(B^{\gamma}\log\left(B\right)(a,b)\right) - (\gamma-1)B^{\gamma}(a,b)\Bigg((ac+1)B^{(0,1)}\left(1+(b-1)\gamma, \tfrac{\gamma ac+\gamma-1}{c}\right) + (b-1)cB^{(1,0)}\left(1+(b-1)\gamma, \tfrac{\gamma ac+\gamma-1}{c}\right)\Bigg)\Bigg)\Bigg) / \left(c(\gamma-1)^2 B\left(1+(b-1)\gamma, \tfrac{\gamma ac+\gamma-1}{c}\right)B^{\gamma}(a,b)\right)$$

Using L'Hospital's rule again, one gets the expression

$$-2\,\acute{\mathcal{J}}_R(1) = -log\left(\tfrac{B(b,a)}{B(a,b)}\right)\Big(c\,B(b,a)\big(B\;log(B)\big)(a,b) + B(a,b)\Big((1+ac)B^{(0,1)}(b,a) + (b-1)c\,B^{(1,0)}(b,a)\Big)\Big) \qquad (21)$$

for the measure proposed by Song (2001). When f(x) has a finite fourth moment, this measure plays a similar role as the kurtosis measure in comparing the shapes of various densities and measuring heaviness of tails. Song (2001) provided a detailed discussion including examples to show in general that his measure is better: "it measures more than what kurtosis measures. Kurtosis is often regarded as a measure of tail heaviness of a distribution relative to that of the normal distribution and a measure of deviation from normality depending on the relative frequency of values either near the mean or far from it to values an intermediate distance from the mean. However, one of the problems with the kurtosis measure is that in many situations of interest when the tails of a

distribution are heavy, the fourth moments may not exist or, even worse, the mean may not exist. This illustrates the limitation of the moments as indicators of distributional shape. However, our shape parameter is almost always applicable in virtually all situations encountered in practice. It does not require restrictions like symmetry assumption and may be applied to distributions with heavy tails such as Cauchy, Cramer, Levy distributions.

5. Estimation of the parameters

Here, we consider estimation by two methods: the method of moments and the method of maximum likelihood.

5.1. Method of moments

Let $x_1, x_2, \ldots, x_n$ be a random sample from (3). Equating $\mathrm{E}(x)$, $Var(x)$ and $E[x - E(x)]^3$ in (6), (10) and (11), respectively, with the corresponding sample estimates

$$S_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - S_1)^2$$

and

$$S_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - S_1)^3$$

respectively, one obtains the system of equations

$$S_1 = \frac{\lambda\,\Gamma(a+b)\,\Gamma\left(a-\frac{1}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{1}{c}\right)} \tag{22}$$

$$S_2 = \frac{\lambda^2\Gamma(a+b)\Gamma\left(a-\frac{2}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{2}{c}\right)} - \left[\frac{\lambda\,\Gamma(a+b)\Gamma\left(a-\frac{1}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{1}{c}\right)}\right]^2 \tag{23}$$

$$S_3 = \frac{\lambda^3\Gamma(a+b)\Gamma\left(a-\frac{3}{c}\right)}{\Gamma(a)\Gamma\left(a+b-\frac{3}{c}\right)} + \frac{2\lambda^3\Gamma(a+b)^3\Gamma\left(a-\frac{1}{c}\right)^3}{\Gamma(a)^3\Gamma\left(a+b-\frac{1}{c}\right)^3} - \frac{2\lambda^3\Gamma(a+b)^2\Gamma\left(a-\frac{1}{c}\right)\Gamma\left(a-\frac{2}{c}\right)}{\Gamma(a)^2\Gamma\left(a+b-\frac{1}{c}\right)\Gamma\left(a+b-\frac{2}{c}\right)} \tag{24}$$

Combining (22) with (23) and (22) with (24), one obtains the equations

$$\frac{s_2}{s_1^2} = -1 + \frac{\Gamma(a)\,\Gamma\left(a-\frac{2}{c}\right)\Gamma\left(a+b-\frac{1}{c}\right)^2}{\Gamma(a+b)\,\Gamma\left(a+b-\frac{2}{c}\right)\Gamma\left(a-\frac{1}{c}\right)^2}$$

$$\frac{s_3}{s_1^3} = 2 + \frac{\left(\Gamma(a)\,\Gamma\left(a+b-\frac{1}{c}\right)^2\left(-\frac{3\Gamma(a+b)\,\Gamma\left(a-\frac{2}{c}\right)\Gamma\left(a-\frac{1}{c}\right)}{\Gamma\left(a+b-\frac{2}{c}\right)}+\frac{3\Gamma(a)\,\Gamma\left(a-\frac{3}{c}\right)\Gamma\left(a+b-\frac{1}{c}\right)}{\Gamma\left(a+b-\frac{3}{c}\right)}\right)\right)}{\Gamma(a+b)^2\,\Gamma\left(a-\frac{1}{c}\right)^3}$$

which can be solved simultaneously to give estimates for a, b and c. The estimate for λ can then be obtained directly from (22). Note that if $b = 1$ then (22) gives $\hat{\lambda} = \frac{s_1\,(ac-1)}{a\,c}$, which is the usual estimator for λ under the Pareto model.

5.2. Maximum likelihood estimator

The logarithm of the likelihood function for a random sample $x_1, x_2, \ldots, x_n$ from (3) can be expressed as

$$\log L(a,b,c,\lambda) = n\log c - n\log B(a,b) - \sum_{i=1}^{n} log x_i + a\,c\sum_{i=1}^{n} log\left(\frac{\lambda}{x_i}\right) + (b-1)\sum_{i=1}^{n} log\left(1-\left(\frac{\lambda}{x_i}\right)^c\right) \quad (25)$$

To obtain the normal equations for the unknown parameters, we differentiates (25) partially with respect to the parameters a, b, λ and c and equating to zero. The resulting equations are given below in (26), (27), (28) and (29), respectively.

$$\frac{\partial \log L}{\partial a} = -n[\Psi(a) - \Psi(a+b)] + c\sum_{i=1}^{n} log\left(\frac{\lambda}{x_i}\right) \quad (26)$$

$$\frac{\partial \log L}{\partial b} = -n[\Psi(b) - \Psi(a+b)] + \sum_{i=1}^{n} log\left(1-\left(\frac{\lambda}{x_i}\right)^c\right) \quad (27)$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{anc}{\lambda} - (b-1)\sum_{i=1}^{n} \frac{c\left(\frac{\lambda}{x_i}\right)^{c-1}}{\left(1-\left(\frac{\lambda}{x_i}\right)^c\right)x_i} \quad (28)$$

$$\frac{\partial \log L}{\partial c} = \frac{n}{c} + a\sum_{i=1}^{n} log\left(\frac{\lambda}{x_i}\right) - (b-1)\sum_{i=1}^{n} \frac{\left(\frac{\lambda}{x_i}\right)^c \log\left(\frac{\lambda}{x_i}\right)}{1-\left(\frac{\lambda}{x_i}\right)^c} \quad (29)$$

The second order partial derivatives of the log-likelihood function $L(\theta/x)$ with respect to the parameters a, b, λ and c are given by:

$$\frac{\partial^2 \log L}{\partial a^2} = -n[\Psi(a) - \Psi(a+b)] \tag{30}$$

$$\frac{\partial^2 \log L}{\partial a \partial b} = n\,\Psi(a+b) \tag{31}$$

$$\frac{\partial^2 \log L}{\partial a \partial \lambda} = \frac{nc}{\lambda} \tag{32}$$

$$\frac{\partial^2 \log L}{\partial a \partial c} = \sum_{i=1}^{n} \log\left(\frac{\lambda}{x_i}\right) \tag{33}$$

$$\frac{\partial^2 \log L}{\partial b^2} = -n[\Psi(b) - \Psi(a+b)] \tag{34}$$

$$\frac{\partial^2 \log L}{\partial b \partial \lambda} = -\sum_{i=1}^{n} \frac{c\left(\frac{\lambda}{x_i}\right)^{c-1}}{x_i\left(1-\left(\frac{\lambda}{x_i}\right)^{c}\right)} \tag{35}$$

$$\frac{\partial^2 \log L}{\partial b \partial c} = -\sum_{i=1}^{n} \frac{\left(\frac{\lambda}{x_i}\right)^{c} \log\left(\frac{\lambda}{x_i}\right)}{1-\left(\frac{\lambda}{x_i}\right)^{c}} \tag{36}$$

$$\frac{\partial^2 \log L}{\partial \lambda^2} = -\frac{a\,nc}{\lambda^2} + (b-1)\sum_{i=1}^{n}\left(-\frac{c(c-1)\left(\frac{\lambda}{x_i}\right)^{c-2}}{x_i^2\left(1-\left(\frac{\lambda}{x_i}\right)^{c}\right)} - \frac{c^2\left(\frac{\lambda}{x_i}\right)^{2c-2}}{x_i^2\left(1-\left(\frac{\lambda}{x_i}\right)^{c}\right)^2}\right) \tag{37}$$

$$\frac{\partial^2 \log L}{\partial \lambda \partial c} = \frac{a\,n}{\lambda} + (b-1)\sum_{i=1}^{n}\left(-\frac{\left(\frac{\lambda}{x_i}\right)^{c-1}}{x_i\left(1-\left(\frac{\lambda}{x_i}\right)^{c}\right)} - \frac{c\left(\frac{\lambda}{x_i}\right)^{c-1}\log\left(\frac{\lambda}{x_i}\right)}{x_i\left(1-\left(\frac{\lambda}{x_i}\right)^{c}\right)} - \frac{c\left(\frac{\lambda}{x_i}\right)^{2c-1}\log\left(\frac{\lambda}{x_i}\right)}{x_i\left(1-\left(\frac{\lambda}{x_i}\right)^{c}\right)^2}\right) \tag{38}$$

$$\frac{\partial^2 \log L}{\partial c^2} = -\frac{n}{c^2} + (b-1)\sum_{i=1}^{n}\left(-\frac{\left(\frac{\lambda}{x_i}\right)^{c}\log\left(\frac{\lambda}{x_i}\right)^2}{\left(1-\left(\frac{\lambda}{x_i}\right)^{c}\right)} - \frac{\left(\frac{\lambda}{x_i}\right)^{2c}\log\left(\frac{\lambda}{x_i}\right)^2}{\left(1-\left(\frac{\lambda}{x_i}\right)^{c}\right)^2}\right) \tag{39}$$

where $\Psi(x)$ is the digamma function defined by $\Psi(x) = d\log\Gamma(x)/dx$, and $\Gamma(x)$ is the Gamma function. For interval estimation of $(a, b, \lambda$ and $c)$ and tests of hypothesis, one requires the Fisher information matrix. The elements of the Fisher information matrix can be easily derived using the equations (30-39) and results in (6) and (7). The equations (26 -29) cannot be solved analytically. An

iterative process such as Newton-Raphson method has to be adopted to solve this system of equations for $(a, b, \lambda \text{ and } c)$. To study the properties of the estimators of the beta Pareto distribution and their performances. There are many measures that can be used to get information about the performance of the estimators. The bias, the t value, mean square error (MSE), finite sample variance (FSV) and asymptotic variance (ASV) are such useful measures.

5.3. Simulation study

The outline of the simulation is as follows:

1. Note from (2) that if U be a uniform $\left(0, 1 \right)$ random number and by equating $F(x) = U$, then $x = \frac{\lambda}{U^{1/c}}$, which is the formula used for generating random numbers.
2. We took the sample size as $n = 10, 20, 30, 50, 80, 100,$ and the parameters a, b, λ and c as $1.2, 1.3, 1.4, 1.5$. Without loss of generality.
3. We generated 2000 samples from the beta Pareto distribution based on the cumulative distribution function described above. The sample observations $x_1, x_2, \cdots, x_n$ are generated from $x = \frac{\lambda}{U^{1/c}}$, where $u_1, u_2, \cdots, u_n$ random numbers are uniformly distributed on the interval $(0,1)$.
4. Then solve the likelihood equations (26 -29) using Newton-Raphson method. The solutions that we obtained are the maximum likelihood estimators of a, b, λ and c say $\hat{a}, \hat{b}, \hat{\lambda}$ and $\hat{c}$ for the samples obtained.
5. Then compute the following.

$$Bias\left(\widehat{\theta_\iota}\right) = E\left(\widehat{\theta_\iota} - \theta\right) = E\left(\widehat{\theta_\iota}\right) - \theta \ , \quad i = 1,2,3, \dots, n \tag{40}$$

where $\theta = a, b, \lambda$ and, c $\hat{\theta} = \hat{a}, \hat{b}, \hat{\lambda}$ and $\hat{c}$.

The t value

$$t = \sqrt{n}\frac{\left(\bar{\hat{\theta}} - \theta\right)}{\sqrt{var(\hat{\theta})}}, \tag{41}$$

$\bar{\hat{\theta}} = \frac{1}{r}\sum_{i=1}^{r} \widehat{\theta_\iota}$, r number of replications.

The mean square error (MSE)

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \tag{42}$$

The asymptotic Variance (ASV)

$$ASV(\hat{\theta}) = I^{-1}(\theta) \tag{43}$$

where the asymptotic variance (ASV), obtained from the inverse of the observed information matrix, and the finite sample variance (FSV),

$$FSV(\hat{\theta}) = E\left(\hat{\theta} - \bar{\hat{\theta}}\right)^2 \tag{44}$$

6. Repeat steps (2), (3), (4), and (5), 2000 times.

Where $\hat{\theta} = (\hat{a}, \hat{b}, \hat{\lambda}, \hat{c})$ is the maximum likelihood estimators of $\theta = (a, b, \lambda, \mathrm{c})$. The t value measures the significance of the bias. The bias is considered significant if $|t| > 1.96$. The asymptotic variance of $\theta = (a, b, \lambda, \mathrm{c})$ is calculated from the asymptotic approximation of the variance given by the inverse of the observed information matrix for the i -th simulated sample (Hinkley (1978) and Sprott (1980)). The finite sample variance is calculated from the values of the estimators of the distribution in the simulation study, and they represent the actual variance of the estimators. The approximation is valid because asymptotic variance decreases as sample size increases, also because the finite sample variance and asymptotic variance agrees closely with one another (Jennings, (1986)). In determining a good estimator it is also important to study the bias and variances of the estimators. Bias is much less important than the variance, and the main contribution to the mean squared error comes from the variance.

Table 1. Statistical properties of the MLE for the parameters of the beta Pareto distribution

Sample Size				
10	a	b	λ	c
Bias	1.1897473	1.1987473	0.7765018	-0.3039930
t	3.7685034	5.0027368	2.6693100	-0.0146470
ASV	0.2962686	0.0633647	-4.0702400	0.0100182
FSV	0.0154190	0.0028167	0.0341668	0.0080601
MSE	1.2036930	1.4349639	0.6759438	0.1004721
20				
Bias	1.0820847	1.1960311	0.7721289	-0.3051750
t	2.6329270	3.5852405	0.8101258	-0.0104680
ASV	0.1203950	0.0281434	-4.8003100	0.0029462
FSV	0.0130163	0.0007492	0.0267757	0.0052063
MSE	1.1839237	1.4312397	0.6681753	0.0983379
30				
Bias	1.0659877	1.1956768	0.7681117	-0.2907280
t	2.0869101	2.9306000	0.7573450	-0.0091850
ASV	0.0958127	0.0266895	-3.2867700	0.0028018
FSV	0.0146839	0.0008887	0.0283523	0.0057960
MSE	1.1783918	1.4292709	0.6371219	0.0960453
50				
Bias	1.0550919	1.1956061	0.7648447	-0.2856880
t	1.5906622	2.2851957	0.4007003	-0.0068300
ASV	0.0632573	0.0147032	-0.6622140	0.0016095
FSV	0.0116556	0.0007036	0.0215396	0.0045580
MSE	1.1650871	1.4281414	0.6229589	0.0949251
80				
Bias	1.0493616	1.1942572	0.7108231	-0.2015460
t	1.3102052	1.7892695	0.4022143	-0.0046030
ASV	0.0387269	0.0085875	-4.8407600	0.0009300
FSV	0.0133702	0.0004587	0.0185098	0.0051151
MSE	1.1510137	1.4267089	0.6183479	0.0903200
100				
Bias	1.0473837	1.1932963	0.7040942	-0.2003250
t	1.1624449	1.6043640	0.2860630	-0.0040780
ASV	0.0266321	0.0053432	-5.8658000	0.0005939
FSV	0.0119602	0.0005376	0.0216078	0.0047301
MSE	1.1248745	1.4183154	0.6065270	0.0861756

The results given in Table 1 indicate that.

1. The estimates of the c parameter have insignificant bias for all sample sizes. The estimates of the a parameter are significantly biased for small samples ($n < 50$), while it is insignificant for samples greater than 30. The estimates of the b parameter are significantly biased for small samples ($n \leq 50$), while it is insignificant for samples greater than 50. The estimates of the λ parameter are insignificantly biased for all sample sizes except ($n = 10$). These are indicated by the t values for the estimators. The bias is negative for the c parameter, while it is positive for the other parameters. c estimator has the smallest bias compared to the other estimators. In general, the bias tends to decrease as the values of the sample size increases.

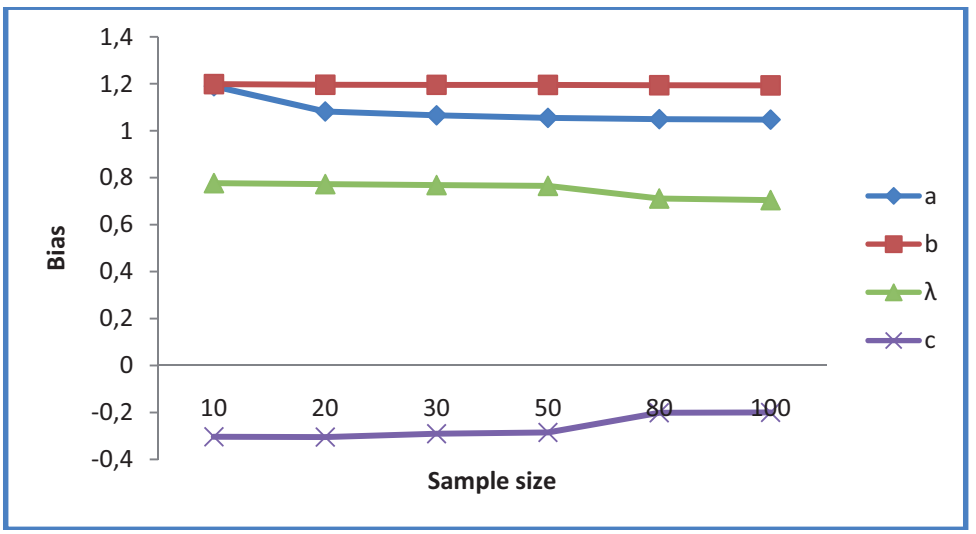

Figure 3. Relationship between the bias of the estimators and sample size.

2. The asymptotic variance of the a, b, λ and c estimators generally decreases as the sample size increases. λ estimator has the smallest ASV for all sample sizes. For the a estimator, the asymptotic variance is consistently higher.

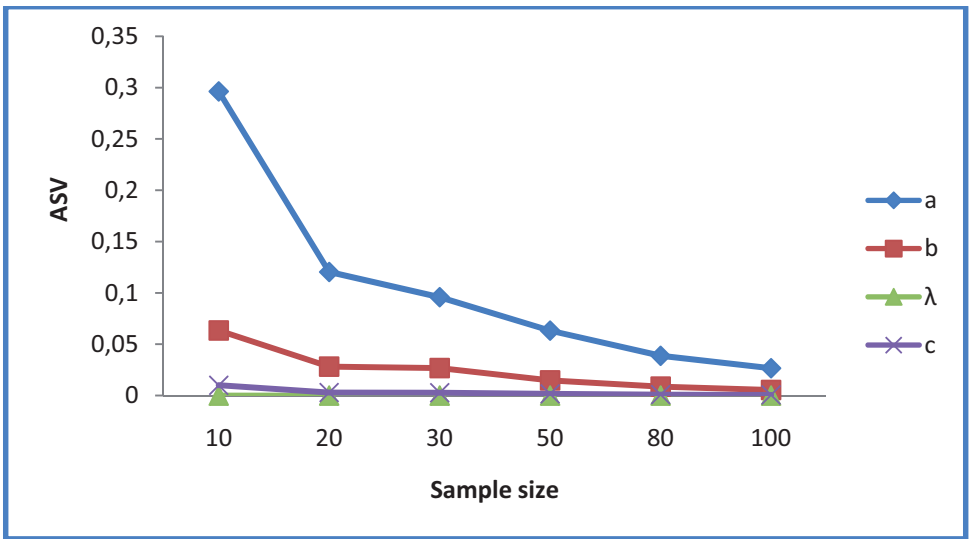

Figure 4. Relationship between the ASV of the estimators and sample size.

3. In general, the mean square error of the a, b, λ and c estimators decreases as the sample size increases as shown in Figure 5. c estimator has the smallest MSE values compared to the other estimators.

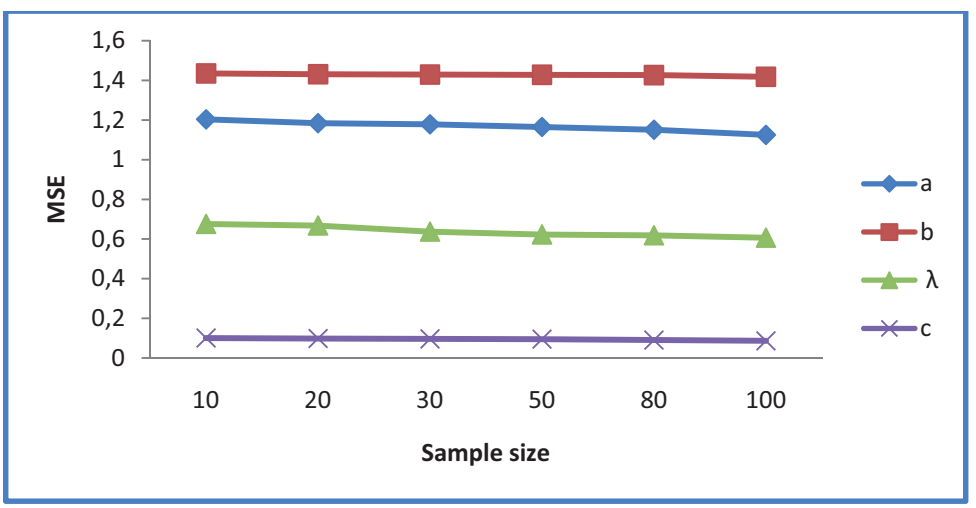

Figure 5. Relationship between the MSE of the estimators and sample size.

8. Conclusions

We have studied the beta Pareto distribution which arise by taking G in (1) to be Pareto. We have derived various properties of the beta Pareto distribution, including the mean, variance, skewness, kurtosis, the mean deviation about the mean, the mean deviation about the median, Rényi and Shannon entropies, the estimation procedures by the methods of moments and maximum likelihood. We have shown that beta Pareto distributions are the most tractable of all the known distributions out of (1). The results presented in this paper can be used as a reference to obtain the corresponding results for other distributions belonging to (1).

Acknowledgements

The author would like to thank the referee and the Editor-in-Chief for carefully reading the paper and for their great help in improving the paper.

REFERENCES

BROWN, B.W, SPEARS, F.M and LEVY, L.B. (2002): The log F: a distribution for all seasons. Comput. Statist., 17, 47–58.

EUGENE, N., LEE, C. and FAMOYE, F. (2002): Beta-normal distribution and its applications. Commun Statist—Theory Methods. 31, 497–512.

GUPTA, A.K. and NADARAJAH, S. (2004): On the moments of the beta normal distribution. Commun Statist—Theory Methods. 33, 1–13.

HINKLEY, D. (1978): Likelihood inference about location and scale parameters. Biometrika. 65(2), 253-261.

JENNINGS, D. (1986): Judging inference adequacy in logistic regression. JASA. 81(394), 471-476.

JONES, M. C. (2004): Families of distributions arising from distributions of order statistics. Test. 13, 1–43.

NADARAJAH, S. and KOTZ, S. (2004): The beta Gumbel distribution. Math Probab Eng. 10, 323–332.

RÉNYI, A. (1961): On measures of entropy and information. In: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability. vol. 1. Berkeley: University of California Press. p. 547–61.

SONG, K. S. (2001): RÉNYI information, log likelihood and an intrinsic distribution measure. J Statisst Plan Inference. 93, 51–69.

SPROTT, D. (1980): Maximum likelihood in small samples: Estimation in the presence of a nuisance parameters. Biometrija. 67(3), 515-523.