Accepted Manuscript

On a new Pareto-type distribution with applications in the study of income inequality and risk analysis

José María Sarabia, Vanesa Jordá, Faustino Prieto

PII: S0378-4371(19)30743-5
DOI: <https://doi.org/10.1016/j.physa.2019.121277>
Article number: 121277
Reference: PHYSA 121277

To appear in: *Physica A*

Received date: 19 March 2019

Revised date: 15 April 2019

Please cite this article as: J.M. Sarabia, V. Jordá and F. Prieto, On a new Pareto-type distribution with applications in the study of income inequality and risk analysis, *Physica A* (2019), <https://doi.org/10.1016/j.physa.2019.121277>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Highlights

- We obtain important properties of a new Pareto-type distribution.
- We derive simple expressions for inequality indices and risk measures.
- We provide the R code for computing the measures presented in this paper.
- Empirical applications with real data are included.

On a new Pareto-type distribution with applications in the study of income inequality and risk analysis

José María Sarabia¹, Vanesa Jordá, Faustino Prieto

Department of Economics, University of Cantabria, 39005-Santander, Spain

Abstract

A recent paper published in this journal [Bourguignon et al, 2016, [1]] introduces and studies a new Pareto-type distribution, deriving some of its probabilistic and inferential properties. In this paper, we obtain additional and important properties of this distribution and derive simpler expressions for relevant inequality indices and risk measures. One important feature of this distribution is that many of these expressions are simple functions of the incomplete beta function. Using this result, we obtain closed expressions for the moments and for several inequality measures including the Gini index and the Lorenz curve, the Donaldson-Weymark-Kakwani index, the Pietra index and the generalized entropy measures. Basic risk measures are also obtained. Finally, empirical applications with real income data are given.

Key Words: Incomplete beta function; Lorenz curve; Gini index; Donaldson-Weymark-Kakwani index; Pietra index; generalized entropy measures; value at risk; tail value at risk.

1 Introduction

The Pareto distribution has long been used as a suitable model for many economic variables, including income and wealth as well as other variables in risk analysis, actuarial science and econophysics. A relevant and important book about Pareto distributions is [2]. Additional properties about Pareto

Email addresses: sarabiaj@unican.es (J.M. Sarabia) Corresponding author, vanesa.jorda@unican.es (V. Jordá), faustino.prieto@unican.es (F. Prieto)

distributions related with majorization and Lorenz ordering can also be found in [3].

In [1], the authors propose a new Pareto distribution with cumulative distribution function (cdf) given by,

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

where $\alpha > 0$ is a shape parameter and $\beta > 0$ a scale parameter. The corresponding probability density function is given by,

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

The quantile function is given by,

$$Q(p) = \beta \left(\frac{1+p}{1-p} \right)^{1/\alpha}, \quad 0 < p < 1. \quad (3)$$

A random variable with cdf (1) is represented by $X \sim NP(\alpha, \beta)$.

The new Pareto-type distribution is connected with other Pareto-type of distributions, in special with the Pareto III distribution (see Appendix 1). The Type-I Pareto distribution is more convenient than the new Pareto distribution when we need to test the power-law behavior (including the Zipf law) at the upper tail. On the contrary, the new Pareto distribution is a better candidate when we work with a wider range of the data. In addition, the new Pareto distribution presents the property of the weak Pareto law, that is, converges in distribution to the Pareto model for x sufficiently large (see [1]).

The advantages and the motivation to use the new Pareto distribution instead of the Type-I Pareto distribution have been highlighted in [1]. The new Pareto distribution provides better fit than the Type-I Pareto in some practical situations, including reliability data and some kind of income data. Moreover, the new distribution presents an upside-down bathtub or a decreasing hazard rate function, whereas the Type-I Pareto is only able to represent decreasing hazard rates. As well, many of the features of the new Pareto distribution can be easily obtained in a closed form.

In this paper, we obtain additional and important properties of the distribution defined in (1) and derive simpler expressions for many relevant economic inequality and risk indices than those provided in [1]. One important

feature of this distribution is that many of these expressions are functions of the incomplete beta function.

The contents of this paper are following. Section 2 provides a simple formula for the moments of the new Pareto distribution. Expressions for several important inequality measures are given in Section 3 and basic risk measures are discussed in Section 4. Simple initial estimates are given in Section 5. Finally, empirical applications with real data are given in Section 6. The paper ends with some conclusions.

2 Moments

Bourguignon et al. [1] obtained the moments of the new Pareto distribution in terms of the ${}_2F_1$ hypergeometric function. However, many of the features of this model can be defined in terms of the incomplete beta function defined by,

$$B(x; p, q) = \int_0^x t^{p-1}(1-t)^{q-1}dt, \quad (4)$$

where $p, q > 0$ and $0 < x < 1$. In this theorem, we get a simpler expressions for the moments of this model in terms of the incomplete beta function defined in (4).

Theorem 1 *Let $X \sim NF(\alpha, \beta)$ be the new Pareto-type distribution with cdf (1). Then, the r -th moment of X is given by,*

$$E(X^r) = 2\beta^r B\left(\frac{1}{2}; 1 - \frac{r}{\alpha}, 1 + \frac{r}{\alpha}\right), \quad \text{if } \alpha > r, \quad (5)$$

where $B(x; p, q)$ denotes the incomplete beta function defined in (4).

Proof: See Appendix 2. ■

Equation (5) will be very useful later for obtaining of many features of the new Pareto-type distribution. The first two moments are given by,

$$\mu = E(X) = 2\beta B\left(\frac{1}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right), \quad \alpha > 1, \quad (6)$$

$$\sigma^2 + \mu^2 = E(X^2) = 2\beta^2 B\left(\frac{1}{2}; 1 - \frac{2}{\alpha}, 1 + \frac{2}{\alpha}\right), \quad \alpha > 2, \quad (7)$$

where $\sigma^2 = \text{var}(X)$.

3 Inequality measures

In this section we obtain closed expressions for the most popular inequality measures.

3.1 Lorenz curve

Let $\mathcal{L}$ be the class of all non-negative random variables with positive finite expectations. Let X be a random variable in $\mathcal{L}$ with cdf F_X and expectation μ_X , its inverse distribution function is defined by,

$$F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}.$$

According to the Gastwirth definition [4], the Lorenz curve associated with X is defined as (see [5], [6]),

$$L_X(u) = \frac{\int_0^u F_X^{-1}(y) dy}{\int_0^1 F_X^{-1}(y) dy} = \frac{1}{\mu_X} \int_0^u F_X^{-1}(y) dy, \quad 0 \leq u \leq 1. \quad (8)$$

In [1], the authors provide an expression of the Lorenz curve in terms of the quotient of an integral and the ${}_2F_1$ hypergeometric function. In this lemma we obtain a simple expression for the Lorenz curve.

Lemma 1 *Let $X \sim NP(\alpha, \beta)$ be the new Pareto-type distribution with cdf (1). If $B(x; p, q)$ denotes the incomplete beta function defined in (4), the Lorenz curve is given by,*

$$L(u; \alpha) = 1 - \frac{B(\frac{1-u}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha})}{B(\frac{1}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha})}, \quad 0 \leq u \leq 1, \quad (9)$$

with $\alpha > 1$.

Proof: See Appendix 2. ■

Note that expression (9) can also be written in terms of the incomplete beta function ratio. Figure 1 shows the Lorenz curve for the new-Pareto type distribution for selected values of α . The nested nature of the Lorenz curve gives place to the Lorenz ordering. If $X_i \sim NP(\alpha_i, \beta_i)$, $i = 1, 2$, with $1 < \alpha_1 \leq \alpha_2$ and $\beta_i > 0$, $i = 1, 2$ we have,

$$L_{X_1}(u; \alpha_1) \geq L_{X_2}(u; \alpha_2), \quad 0 \leq u \leq 1,$$

and then X_1 exhibits less inequality than X_2 in the Lorenz sense.

The Lorenz curve (Eq. (9)) can be interpreted as a linear combination of an infinite set of potential Lorenz curves $L_k(u) = \frac{u^k}{k}$, $k = 1, 2, \dots$. If we consider the four first terms of the expansion of (9) we have

$$L(u, \alpha) \approx \frac{1}{2k_\alpha}u + \frac{1}{2\alpha k_\alpha}u^2 + \frac{1}{3\alpha^2 k_\alpha}u^3 + \frac{2 + \alpha^2}{12\alpha^3 k_\alpha}u^4, \quad 0 \leq u \leq 1,$$

where

$$k_\alpha = B\left(\frac{1}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right). \quad (10)$$

The general series expansion of the Lorenz curve (9) can be found in Appendix 2.

Figure 1: Lorenz curves for the new Pareto-type distribution for some selected values of α .

3.2 Gini index

In [1], the authors provide an expression of the Gini index in terms of a double integral. The next result provides a simpler expression for the Gini index.

Lemma 2 Let $X \sim NP(\alpha, \beta)$ be the new Pareto-type distribution with cdf (1). The Gini index is given by,

$$G(\alpha) = 1 - \frac{4B(\frac{1}{2}; 2 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha})}{B(\frac{1}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha})}, \quad (11)$$

with $\alpha > 1$, being $B(x; p, q)$ the incomplete beta function defined in (4).

Proof: See Appendix 2. ■

Figure 2: Gini index for the new Pareto-type distribution.

Figure 2 shows the Gini index, which is decreasing in α . The Gini index covers all the range in $[0, 1]$ and satisfies that $\lim_{\alpha \rightarrow \infty} G(\alpha) = 0$ and $\lim_{\alpha \rightarrow 1} G(\alpha) = 1$.

3.3 The Donaldson-Weymark-Kakwani index

A relevant generalization of the Gini index was proposed by Donaldson and Weymark [7] and Kakwani [8] and studied by Yitzhaki [9]. These authors proposed the following generalized Gini index defined as,

$$D_X(\nu) = 1 - \nu(\nu + 1) \int_0^1 (1 - p)^{\nu-1} L_X(p) dp, \quad (12)$$

where the parameter $\nu > 1$ and $L_X(\cdot)$ is the Lorenz curve. If we set $\nu = 1$ in Eq. (12) we obtain the Gini index. When ν increases, the index becomes more sensitive to changes at the lower tail of the distribution. In the limit,

when ν goes to infinity, this inequality measure depends only on the minimum value, expressing the judgement introduced by Rawls, that the assessment of social welfare should be based solely on the poorest society member. The proof is similar to the proof of Lemma 2 and will be omitted.

Lemma 3 *If $X \sim NP(\alpha, \beta)$, the Donaldson-Weymark-Kakwani index is,*

$$D_X(\nu, \alpha) = 1 - (\nu + 1)2^\nu \frac{B(\frac{1}{2}; 1 - \nu - \frac{1}{\alpha}, 1 + \frac{1}{\alpha})}{B(\frac{1}{2}, 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha})}, \quad (13)$$

where $\nu \geq 1$, $\alpha > 1$.

If we set $\nu = 1$ in (13) we get (11). Figure 3 shows the Donaldson-Weymark-Kakwani index as a function of α for selected values of ν .

Figure 3: Donaldson-Weymark-Kakwani index for the new Pareto-type distribution index as a function of α for selected values of ν .

3.4 The Pietra index

The Pietra index is defined as the maximal vertical deviation between the Lorenz curve and the egalitarian line, that is,

$$P_X = \max_{0 \leq p \leq 1} \{p - L_X(p)\}.$$

This index was proposed by Pietra [10]. Properties and applications of this index can be found in [11, 12, 3].

If we assume that F_X is strictly increasing on its support, the function $p - L_X(p)$ will be differentiable everywhere on $(0, 1)$ and its maximum will be reached when $1 - F_X^{-1}(x)/\mu$ is zero, that is, when $x = F_X(\mu)$. The value of $p - L_X(p)$ in this point is given by

$$P_X = F_X(\mu) - L_X(F_X(\mu)). \quad (14)$$

Using (14), we obtain the following expression for the Pietra index,

$$P_X(\alpha) = \frac{-2}{1 + (2k_\alpha)^\alpha} + \frac{1}{k_\alpha} B\left(\frac{1}{1 + (2k_\alpha)^\alpha}, 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right), \quad (15)$$

where $\alpha > 1$ and k_α is defined in (10).

3.5 The Generalized Entropy indices

The generalized entropy (GE) indices of a non-negative random variable with $E(X^\theta) < \infty$ are defined as (see [13]),

$$GE_X(\theta) = \frac{E(X/\mu)^\theta - 1}{\theta(\theta - 1)}, \quad \theta \neq 0, 1.$$

For the new Pareto distribution we have,

$$GE_X(\theta) = \frac{1}{\theta(\theta - 1)} \left\{ \frac{B(1/2; 1 - \theta/\alpha, 1 + \theta/\alpha)}{2^{\theta-1} \{B(1/2; 1 - 1/\alpha, 1 + 1/\alpha)\}^\theta} - 1 \right\}, \quad (16)$$

if $\theta < \alpha$. The limit case $\theta = 0$ is known as the Theil index, which is given by,

$$GE_X(0) = -E\left(\log \frac{X}{\mu}\right) = \log(2) - \frac{\log(4)}{\alpha} + \log B\left(\frac{1}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right), \quad (17)$$

with $\alpha > 1$.

4 Risk measures

In this section we provide some basic risk measures for the new Pareto-type distribution.

The value at risk (VaR) at level p , with $0 < p < 1$ of a random variable X with cdf $F_X(x)$ is defined by

$$\text{VaR}[X; p] = \inf\{x \in \mathbb{R}, F_X(x) \geq p\}.$$

Therefore, the VaR is obtained directly from the quantile function as follows,

$$\text{VaR}[X; p] = \beta \left(\frac{1+p}{1-p} \right)^{1/\alpha}, \quad 0 < p < 1 \quad (18)$$

The VaR has the potential limitation of not being a coherent risk measure because it does not satisfy the sub-additivity property. On the contrary, the tail value at risk (TVaR) is a coherent risk measure (see [14]), which is given by

$$\text{TVaR}[X; p] = \frac{1}{1-p} \int_p^1 \text{VaR}[X; p] dp,$$

If X is continuous, $\Pr(X \leq \text{VaR}[X; p]) = p$, so the TVaR is the conditional tail expectation $\text{TVaR}[X; p] = E[X | X > \text{VaR}[X; p]]$. The TVaR for the new Pareto distribution is given in the next lemma.

Lemma 4 *Let $X \sim NP(\alpha, \beta)$ be the new Pareto-type distribution with cdf (1). The Tail Value at Risk is given by,*

$$\text{TVaR}[X; p] = \frac{2\beta}{1-p} B\left(\frac{1-p}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right) \quad (19)$$

with $\alpha > 1$, being $B(x; p, q)$ the incomplete beta function defined in (4).

Proof: See Appendix 2. ■

Figure 4 shows the Value at Risk (Equation (18)) and the Tail Value at Risk (Equation (19)) for new-Pareto type distribution as a function of p , for $\alpha=1.5, 2$ and 5 and unit scale.

5 Simple initial estimates

In [1], the authors provide maximum likelihood estimates for the α and β parameters. In this section we include simple initial estimates for both parameters based on quantiles and moments. These estimates can be used as initial values in the maximum likelihood estimation.

Let $Q(p_1)$ and $Q(p_2)$ two quantiles of the new-Pareto distribution, with $0 < p_1 < p_2 < 1$. Taking the quotient and solving for α , we get the following estimate,

$$\hat{\alpha} = \frac{\log h(p_1) - \log h(p_2)}{\log Q(p_1) - \log Q(p_2)},$$

and

$$\hat{\beta} = Q(p_1)h(p_1)^{-1/\hat{\alpha}},$$

where $h(u) = (1+u)/(1-u)$.

We obtain now closed formulas for the moment estimators of the parameters α and β using the first two moments. If we consider the first and the second moments of X defined in (5), we have

$$\frac{E(X^2)}{\{E(X)\}^2} = \frac{B(\frac{1}{2}; 1 - \frac{2}{\alpha}, 1 + \frac{2}{\alpha})}{2[B(\frac{1}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha})]^2}.$$

if $\alpha > 2$.

Now, let us consider a sample of size n of $x_1, x_2, \dots, x_n$ from the new Pareto distribution. To obtain moment estimates we proceed in two steps. In a first step, the α estimate is the unique solution of the equation

$$\frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{(\frac{1}{n} \sum_{i=1}^n x_i)^2} = \frac{B(\frac{1}{2}; 1 - \frac{2}{\alpha}, 1 + \frac{2}{\alpha})}{2[B(\frac{1}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha})]^2}. \quad (20)$$

Because the right side of (20) is a decreasing function in α , we have a unique solution. **The unique solution can be derived numerically using standard methods.** We denote as $\hat{\alpha}$ this estimate. In a second step, we use the first moment to estimate β . The moment estimate is,

$$\hat{\beta} = \frac{\frac{1}{n} \sum_{i=1}^n x_i}{2B(\frac{1}{2}; 1 - \frac{1}{\hat{\alpha}}, 1 + \frac{1}{\hat{\alpha}})}. \quad (21)$$

To illustrate we take the Data Set II considered in Bourguignon et al. [1]. Using (20) and (21), we get the estimators $\alpha = 2.836$ and $\beta = 0.0671$, which are very close to the maximum likelihood estimators obtained in [1].

6 Applications

In this section we provide some applications of the results presented above. In doing so, we draw on some of the examples considered in [1].

6.1 Personal Income Data

We focus first on income variables. In particular, we consider the median annual personal data (by state) in the United States in 2012 and 2013 reported in [1]. Table 1 includes the different inequality measures obtained in this paper, according to the new Pareto model fitted in [1]. Figure 5 shows the Lorenz curves corresponding to 2012 and 2013, which suggests Lorenz dominance between both years.

Table 1: Inequality indices corresponding to the personal income data by state in USA for 2012 and 2013.

Indices	USA 2012	USA 2013
Mean	54434.4	58186.7
Median	48708.6	47739.8
CV	0.3621	0.6754
Gini	0.1502	0.2444
D(2)	0.2671	0.3105
D(3)	0.2310	0.3420
Pietra	0.1157	0.1789
GE(0)	0.0431	0.1001
GE(2)	0.0656	0.2281
GE(2)	0.0374	0.0888

6.2 Failure times

Now we consider the Data Set II used in [1] (taken from [15]), which consists of failure times of 20 mechanical components. The maximum likelihood estimators are $\hat{\alpha} = 2.871$ and $\hat{\beta} = 0.067$. Table 2 includes the mean, the median, the coefficient of variation and the risk measures VaR and TVaR for $p = 0.95, 0.99$ and 0.999 .

7 Conclusions

In this paper we have obtained important and new properties of a new generalization of the well-known Pareto distribution, introduced by Bourguignon

Table 2: Risk measures for the failure times of 20 mechanical components

Indices	Set II
Mean	0.1207
Median	0.0982
CV	0.7107
$\text{VaR}[X;0.95]$	0.2400
$\text{VaR}[X;0.99]$	0.4234
$\text{VaR}[X;0.999]$	0.9458
$\text{TVaR}[X;0.95]$	0.3703
$\text{TVaR}[X;0.99]$	0.6334
$\text{TVaR}[X;0.999]$	1.4714

et al. [1]. We have derived simple expressions for many relevant economic inequality and risk indices using the incomplete beta function. We have also obtained closed expressions for the moments and for several inequality measures including the Gini index and the Lorenz curve, the Donaldson-Weymark-Kakwani index, the Pietra index and the generalized entropy measures. Moreover, we have derived expressions for basic risk measures. These properties have been illustrated with two sets of real data. In Appendix 3, we include the R functions for computing the different inequality and risk measures obtained in the paper.

Acknowledgements

The authors thank Ministerio de Economía y Competitividad (project ECO2016-476203-C2-1-P) for partial support of this work. The authors are grateful for the constructive suggestions provided by the reviewers, which improved the paper.

Appendix 1: Relationships with other Pareto-type distributions

Let X be a Pareto III distribution (see [2]) with cdf,

$$F(x) = 1 - \frac{1}{1 + x^\alpha}, \quad x > 0, \quad (22)$$

and $F(x) = 0$ if $x < 0$, with $\alpha > 0$. We have the following lemma (the proof is straightforward and will be omitted).

Lemma 5 *Let X be a Pareto III distribution with cdf (22). Then, if $\beta > 0$, the random variable*

$$Y = \beta(X|X > 1),$$

is distributed as a $NP(\alpha, \beta)$.

Appendix 2: Proofs

Series expansion of the Lorenz curve (9). First, we have,

$$B(x; a, b) = x^a \sum_{k=0}^{\infty} \frac{(1-b)_k}{k!(a+k)} x^k,$$

with $|x| < 1$ and $-a \notin \mathbb{N}$. Then, the Lorenz curve can be written as,

$$\begin{aligned} L(u; \alpha) &= 1 - \frac{u^{1-1/\alpha}}{k_\alpha} \sum_{k=0}^{\infty} \frac{(-1/\alpha)_k}{k!(1+k-1/\alpha)2^k} (1-u)^k \\ &= 1 - \sum_{k=0}^{\infty} \frac{(-1/\alpha)_k}{k!(1+k-1/\alpha)2^k k_\alpha} \sum_{j=0}^k (-1)^j \binom{k}{j} u^{j+1-1/\alpha}, \end{aligned}$$

where k_α is defined in (10).

Proof of Theorem 1. If $r < \alpha$

$$\begin{aligned}
 E(X^r) &= \int_0^1 \{Q_X(u)\}^r du \\
 &= \beta^r \int_0^1 \left(\frac{1+u}{1-u} \right)^{r/\alpha} du \\
 &= 2\beta^r \int_1^\infty \frac{t^{r/\alpha}}{(1+t)^2} dt \\
 &= 2\beta^r \int_0^{1/2} z^{-1/\alpha} (1-z)^{r/\alpha} dz \\
 &= 2\beta^r B\left(\frac{1}{2}, 1 - \frac{r}{\alpha}, 1 + \frac{r}{\alpha}\right),
 \end{aligned}$$

where we have used the changes of variable $t = (1+u)/(1-u)$ and $z = (1+t)^{-1}$.

Proof of Lemma 1. Using the definition of Lorenz curve given in (8), we have

$$\begin{aligned}
 &\int_0^u \beta \left(\frac{1+y}{1-y} \right)^{1/\alpha} dy \\
 &= 2\beta \int_1^{(1+u)/(1-u)} \frac{t^{1/\alpha}}{(1+t)^2} dt \\
 &= 2\beta \int_{(1-u)/2}^{1/2} z^{-1/\alpha} (1-z)^{1/\alpha} dz \\
 &= 2\beta \left[k_\alpha - B\left(\frac{1-u}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right) \right],
 \end{aligned}$$

which combined with (6) gives the result.

Proof of Lemma 2. The Gini index can be written as $G = 1 - 2 \int_0^1 L(u) du$.

If $k_\alpha = B(1/2, 1 - 1/\alpha, 1 + 1/\alpha)$ and $\alpha > 1$, we have,

$$\begin{aligned} \int_0^1 L(u) du &= 1 - k_\alpha^{-1} \int_0^1 B\left(\frac{1-u}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right) du \\ &= 1 - k_\alpha^{-1} \int_0^1 \left[\int_0^{(1-u)/2} z^{-1/\alpha} (1-z)^{1/\alpha} dz \right] du \\ &= 1 - k_\alpha^{-1} \int_0^{1/2} \left[\int_0^{1-2z} z^{-1/\alpha} (1-z)^{1/\alpha} du \right] dz \\ &= 1 - k_\alpha^{-1} \left[k_\alpha - 2B\left(\frac{1}{2}, 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right) \right], \end{aligned}$$

and we obtain formula (11).

Proof of Lemma 4. We have,

$$\int_p^1 \beta\left(\frac{1+u}{1-u}\right) du = 2\beta\left(\frac{1-p}{2}; 1 - \frac{1}{\alpha}, 1 + \frac{1}{\alpha}\right),$$

and we get (19).

Appendix 3: R functions of inequality indices and risk measures for the new Pareto-type model

```
# -----
# Inequality indices and risk measures for the new Pareto type distribution
# -----
#
#####
# Definition of the Lower Incomplete Beta Function,
# where  $0 \leq x \leq 1$  and  $p, q > 0$ ; otherwise, it does not exist.
#
ibeta<-function(x,p,q){
  if (x>=0&&x<=1&&p>0&&q>0){pbeta(x,p,q)*beta(p,q)}
  else{message("Error ibeta function: please check that  $0 \leq x \leq 1$  &  $p, q > 0$ ")}}
}
ibeta(0.5,2,3) # => 0.05729167
#
#####
# Moments,
# where  $a > r$  and  $a, b > 0$ ; otherwise it does not exist.
#
moment<-function(r,a,b){
  if (a>r&&a>0&&b>0){2*b^r*ibeta(1/2,1-r/a,1+r/a)}
  else{message("Error moment function: please check that  $a > r$  &  $a, b > 0$ ")}}
}
moment(1,2,3) # => 7.712389
#
#####
# Lorenz curve,
# where  $0 \leq x \leq 1$  and  $a > 1$ ; otherwise, it does not exist.
#
LorenzCurve<-function(u,a){
  if (u>=0&&u<=1&&a>1){1-1/a*((1-u)/2,1-1/a,1+1/a)/ibeta(1/2,1-1/a,1+1/a)}
  else{message("Error LorenzCurve function: check that  $0 \leq u \leq 1$  &  $a > 1$ ")}}
}
LorenzCurve(0.5,2) # => 0.2557859
#
#####
# Gini Index,
# where  $a > 1$ ; otherwise, it does not exist.
#
GiniIndex<-function(a){
  if (a>1){1-4*1/a*((1/2,2-1/a,1+1/a)/ibeta(1/2,1-1/a,1+1/a))}
  else{message("Error GiniIndex function: please check that  $a > 1$ ")}}
}
GiniIndex(2) # => 0.3889845
#
#####
# Donaldson-Weymark-Kakwani index,
# where  $nu \geq 1$  &  $a > 1$ ; otherwise, it does not exist.
#
DWIndex<-function(nu,a){
  if (nu>=1&&a>1){
    1-(nu+1)*2^nu*ibeta(1/2,1+nu-1/a,1+1/a)/ibeta(1/2,1-1/a,1+1/a)}
  else{message("Error DWIndex function: please check that  $nu \geq 1$  &  $a > 1$ ")}}
```

```

}
DWKIndex(3,2) # => 0.5097534
#
#####
# Pietra index
# where a>1; otherwise, it does not exist.
#
PietraIndex<-function(a){
  if (a>1){ka<-ibeta(1/2,1-1/a,1+1/a);
    -2/(1+(2*ka)^a)+ibeta(1/(1+(2*ka)^a),1-1/a,1+1/a)/ka}
  else{message("Error PietraIndex function: please check that a>1")}
}
PietraIndex(2) # => 0.2886066
#
#####
# Generalized Entropy Indices
# th<a (limiting case, th=0) & th!=0,1 & a>1; otherwise, it doesn't exist.
#
GEntropyIndices<-function(th,a){ # general case, when th!=0,1
  if (th>0&&th<a&&th!=1&&a>1){
    1/(th*(th-1))*(ibeta(1/2,1-th/a,1+1/a)/(2^(th-1))*
      ibeta(1/2,1-1/a,1+1/a)^th)-1)}
  else{message("Error GEntropyIndices function: 0<th<a & th!=1 & a>1")}
}
GEntropyIndices(0.6,2) # => 0.3012451
#
GEntropyIndex0<-function(a){ # limiting case, when th=0 (Theil index)
  if (a>1){log(2)-(log(4)/a)+log(ibeta(1/2,1-1/a,1+1/a))}
  else{message("Error GEntropyIndex0 function: please check that a>1")}
}
GEntropyIndex0(3) # => 0.00785998
#
#####
# Value at Risk VaR
# 0<p<1& a,b>0; otherwise it does not exist.
#
vatrisk<-function(p,a,b){
  if (p>0&&p<1&&a>0&&b>0){b*((1+p)/(1-p))^(1/a)}
  else{message("Error vatrisk function: please check that 0<p<1 & a,b>0")}
}
vatrisk(0.5,2,3) # => 5.196152
#
#####
# Tail Value at Risk TVaR
# 0<p<1& a>1, b>0; otherwise it does not exist.
#
tvatrisk<-function(p,a,b){
  if (p>0 &&p<1&&a>1&&b>0){2*b*ibeta((1-p)/2,1-1/a,1+1/a)/(1-p)}
  else{message("Error tvatrisk function: check that 0<p<1 & a>1 & b>0")}
}
tvatrisk(0.5,2,3) # => 11.47934
#
# -----
# End
# -----

```

References

- [1] Bourguignon, M., Saulo, H., Fernandez, R.N. (2016). A new Pareto-type distribution with applications in reliability and income data *Physica A: Statistical Mechanics and its Applications*, 457, 166-175.
- [2] Arnold, B.C. (2015). *Pareto Distributions*. Second Edition. Chapman and Hall/CRC.
- [3] Arnold, B.C., Sarabia, J.M. (2018). *Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics*. Springer, Statistics for Social and Behavioral Sciences Series.
- [4] Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, 39, 1037-1039.
- [5] Sarabia, J.M., Castillo, E., and Sjötte, D.J. (1999). An ordered family of Lorenz curves. *Journal of Econometrics*, 91, 43-60.
- [6] Sarabia, J.M. (2008). Parametric Lorenz Curves: Models and Applications. In: *Modeling Income Distributions and Lorenz Curves*. Series: Economic Studies in Inequality, Social Exclusion and Well-Being 4, Chotikapanich, D. (Ed.), pp. 167-190, Springer-Verlag.
- [7] Donaldson, D., Weymark, J.A. (1980). A single-parameter generalization of the Gini index of inequality. *Journal of Economic Theory* 22, 67-86.
- [8] Kakwani, N. (1980). On a class of poverty measures. *Econometrica* 48, 437-446.
- [9] Yitzhaki, S. (1983). On an extension of the Gini inequality index. *International Economic Review* 24, 617-628.
- [10] Pietra, G. (1948). *Studi di Statistica Metodologica*. Giuffrè, Milano.
- [11] Eliazar, I.I., Sokolov, I.M. (2010). Measuring statistical heterogeneity: The Pietra index. *Physica A: Statistical Mechanics and its Applications*, 389, 117-125.

- [12] Sarabia, J.M., Jordá, V. (2014). Explicit expressions on the Pietra index for the generalized function for the size distribution of income *Physica A: Statistical Mechanics and its Applications*, 416, 582-595.
- [13] Sarabia, J.M., Jordá, V., Remuzgo, L. (2017). The Theil Indices in Parametric Families of Income Distributions - A Short Review. *Review of Income and Wealth*, 63, 867-880.
- [14] Acerbi, C., Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7), 1487-1503.
- [15] Murthy, D.N.P., Xie, M., Jiang, R. (2004) *Weibull Models*, Wiley Series in Probability and Statistics, John Wiley and Sons, Hoboken, NJ.

Figure 4: Plots of the Value at Risk (top panel) and the Tail Value at Risk (bottom panel) for new-Pareto type distribution as a function of p , for $\alpha=1.5$, 2 and 5 and $\rho = 1$.

Figure 5: Lorenz curves based on the new-Pareto distribution corresponding to the median annual personal data by state in the United States in 2012 and 2013.