STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, October 2025, pp 1625–1639.
Published online in International Academic Press (www.IAPress.org)

Properties of the Leimkuhler Curve with Its Application in JCR

Vahideh Asghari, Gholamreza Mohtashami Borzadaran*, Hadi Jabbari

Department of Statistics, Ferdowsi University of Mashhad, Iran

Abstract One of the most noticeable ways of illustrating the degree of concentration in a theoretical or empirical frequency distribution is via the Leimkuhler curve. Leimkuhler curve is particularly appropriate in the field of informetrics where the variable of interest is the number of citations, relevant references, borrowing of a monograph, etc. In informetrics, interest usually focuses on the most productive sources, and the equivalent graphical representation is via the Leimkuhler curve. In this paper, some statistical properties of the Leimkuhler curve, a plot of the cumulative proportions of total productivity against the cumulative proportions of sources, where the sources are ordered non-increasingly concerning their productivity levels are discussed. Also, some aspects of the Leimkuhler curve and its connection with other criteria are derived. Finally, several concentration measures are obtained using the data of the impact factors in eight scientific fields.

Keywords Leimkuhler curve, Lorenz curve, Gini index, Weighted distribution, Distortion function, Tail value at risk, Inverse distribution function.

AMS 2010 subject classifications 91B80, 62N05, 91B15, 62P20

DOI: 10.19139/soic-2310-5070-1349

1. Introduction

There is a long history of the illustration of concentration of production in informetrics by both graphical and numerical means. It can be argued that the founding articles of Lotka [26] and Bradford [6] are both essentially concerned with such concentration aspects.
Leimkuhler [24] derived a mathematical form for the Bradford distribution,

$$f(x) = \frac{\log(1+\beta x)}{\log(1+\beta)}, \qquad 0 \le x \le 1, \tag{1}$$

depending on a single parameter interpretable from the verbal formulation of Bradford's law. Sarabia et al. [37] proposed a general methodology for generating new classes of parametric Lorenz curve (LC) and Leimkuhler curve (LKC) that contain the original curve as limiting or particular curve. The mathematical form of Bradford distribution (1) is derived by using this methodology. Trueswell [42] propounded a graphical presentation of the concentration of data concerning book usage, subsequently developed by Burrell [7], who noted its equivalent to the Lorenz curve (Lorenz [25]) in econometrics. The Lorenz curve is a graphical method of illustrating inequality in, for example, income and wealth distribution (Kleiber and Kotz [23]; Sarabia [36]; Sarabia et al. [34]). The Leimkuhler curve plots the cumulative proportion of total productivity against the cumulative proportion of sources (Burrell [10]; Egghe [15]). Leimkuhler's approach is considered a variant of the Lorenz curve of concentration for bibliometric data (Burrell [7], [8]; Rousseau [33]). The difference between the two constructions is only the sorting order of the values (Burrell [9]; Sarabia [35]).

*Correspondence to: Gholamreza Mohtashami Borzadaran (Email:grmohtashami@um.ac.ir). Department of Statistics, Ferdowsi University of Mashhad, Iran.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

The Lorenz curve is defined by points $(p, L(p))$, where p represents the cumulative proportion of the income-receiving units, and $L(p)$ the cumulative proportion of incomes, but the Leimkuhler curve is defined by point $(p, K(p))$, where p represents the cumulative proportion of total productivity, and $K(p)$ the cumulative proportion of sources. The difference between the two is that for the Lorenz curve, the sources (individuals) are arranged in increasing order of productivity (income), while for the Leimkuhler curve, they are arranged in decreasing order. The Leimkuhler curve bears similarity with the industrial concentration curve that graphs cumulative market shares against rank-ordered number of firms, where the firms are ranked from the largest to the smallest. (See Hannah and Kay [21]). In fact, the inverse generalized Lorenz curve also has a similar construction (Jenkins and Lambert [22]). The Gini coefficient (Gini [20]), which is derived from the Lorenz or Leimkuhler curve, is the most well-known and broadly used measure of inequality in giving graphical and numerical summaries of the concentration of bibliometric distributions. The index measures the inequality among values of a frequency distribution (e.g., incomes in economics, or numbers of citations in bibliometrics). The index has been connected to the field of bibliometrics since Carpenter [11] drew the attention to the similarity of Pratt's [31] measure of class concentration in bibliometrics to the Gini coefficient . In the 1990s bibliometricians investigated the possibility of using other measures of concentration and found that the Gini coefficient performs well in the field (Burrell [8]; Egghe [16]).

The contents of this paper are as follows.

Section 2 presents some preliminaries regarding the Leimkuhler and Lorenz curves. The transformation and weighted of the Leimkuhler curve are described in section 3. Section 4 introduces Parametric Leimkuhler curve. We review Leimkuhler ordering in section 5. The distorted Leimkuhler curve is expressed in section 6. We display the relation between risk measure and Leimkuhler curve is expressed in section 7. In section 8 Some connections between the Leimkuhler curve and other criteria of inequality are derived. In section 9 applications with data taken from Thomson Reuters Journal Citation Reports Edition 2019, in eight scientific fields, are considered. Finally, some conclusions are presented in section 10.

2. Preliminaries

Let X be a non-negative random variable with finite and positive mean $E(X) = \mu$. The distribution function and survival function of X are denoted by F and $\bar{F} = 1 - F$ respectively. The quantile function is assigned by $F^{-1}(t) = \inf\{x : F(x) \geq t, t \in (0,1)\}$. The most widely used graphical tool for describing and analyzing the size distribution and wealth is the Lorenz curve that was introduced by Lorenz [25]. For the income random variable X, the Lorenz curve is defined by

$$L(p) = \frac{\int_0^p F^{-1}(t)dt}{\int_0^1 F^{-1}(t)dt}, \quad 0 \leq p \leq 1, \tag{2}$$

the function $L(p)$ is the cumulative percentage of total income held by individuals having the $100p\%$ lowest incomes.

In this paper, the main result is depicted in the terms of the Leimkuhler curve $K(p)$, which is

$$K(p) = \frac{\int_{1-p}^1 F^{-1}(t)dt}{\int_0^1 F^{-1}(t)dt}, \quad 0 \leq p \leq 1, \tag{3}$$

that represent the proportion of total productivity that accrues to sources having the $100p\%$ largest productivity. The Leimkuhler curve is a reverse-mirror image of the Lorenz curve reflected through the diagonal 45-degree line. The definition of the Lorenz and Leimkuhler curves $L(p)$ and $K(p)$ imply that these curves are coupled by the connection

$$K(p) + L(1-p) = 1 \tag{4}$$

The Lorenz and Leimkuhler curves $L(p)$ and $K(p)$ are monotone increasing from the value zero (i.e. $L(0) = 0$ and $K(0) = 0$) to the value one (i.e. $L(1) = 1$ and $K(1) = 1$). Consequently, the Lorenz and Leimkuhler curves are bounded from below and from above as follows:

$$0 \leq L(p) \leq p \leq K(p) \leq 1. \tag{5}$$

The Lorenz curve $L(p)$ is bounded from below by the floor $L(p) = 0$ of the unit square and is bounded from above by the diagonal line $L(p) = p$ of the unit square. The Leimkuhler curve is bounded from below by the diagonal line $K(p) = p$ of the unit square and is bounded from above by the ceiling $K(p) = 1$ of the unit square. These bounds manifest the two socio-economic extremes of human societies: perfect equality and perfect inequality. Perfect equality is the case of a purely egalitarian distribution of wealth. In terms of the Lorenz curves, perfect equality is characterized by the diagonal line of the unit square: $L(p) = p = K(p) \ (0 \leq p \leq 1)$. In the context of societies with infinitely large populations, perfect inequality constitutes 0% of the society's population has 100% of the society's overall wealth. In terms of the Lorenz and Leimkuhler curves, perfect inequality is characterized by the floor and the ceiling of the unit square: $L(p) = 0 \ (0 \leq p < 1)$ and $K(p) = 1 \ (0 < p \leq 1)$.

Moreover, the Lorenz curve is convex and the Leimkuhler curve is concave, because The first derivative of LKC w.r.t. p implies $K'(p) = \frac{1}{\mu}F^{-1}(1-p)$, for all $0 \leq p \leq 1$ and $K''(p) = \frac{-1}{\mu f(x)}$, where $F(x) = 1 - p$.
The Gini index is the best known and the most widely applied inequality index in economics and informetrics in particular and across science in general. The Gini index G is depicted as the area between the Leimkuhler and Lorenz curves. The Gini coefficient can theoretically range from 0 (complete equality) to 1 (complete inequality) as:

$$G = \int_0^1 [K(p) - L(p)]dp = 1 - \frac{\int_0^\infty \bar{F}(x)^2 dx}{E(x)}. \tag{6}$$

A low Gini index indicates more equal productivity distribution while a high Gini index indicates more unequal productivity distribution. As an illustration, consider a classical Pareto distribution with a distribution function

$$F(x) = 1 - (\frac{x}{\sigma})^{-\alpha}, \ x \geq \sigma \tag{7}$$

where $\sigma > 0$ is a scale parameter and $\alpha > 0$ is a shape parameter. The Lorenz and Leimkuhler curves of the classical Pareto distribution respectively is given by

$$L(p;\alpha) = 1 - (1-p)^{1-\frac{1}{\alpha}}, \ 0 \leq p \leq 1,$$
$$K(p;\alpha) = p^{1-\frac{1}{\alpha}}, \ 0 \leq p \leq 1.$$

Using relation (6), the Gini index of the classical Pareto distribution is

$$G(\alpha) = \frac{1}{2\alpha - 1}.$$

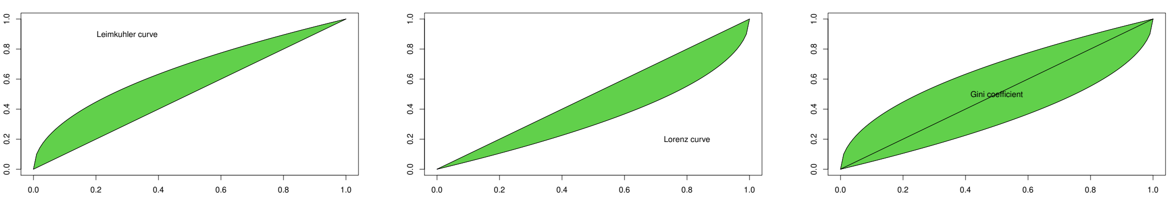

Figure 1. Plot of the Lorenz and Leimkuhler curves and Gini index of the classical Pareto distribution for $\alpha = 2$.

Figure 1 presents three panels, each depicting visualizations related to the Lorenz curve, Leimkuhler curve, and the Gini index for a classical Pareto distribution with the shape parameter $\alpha = 2$. The Lorenz curve (highlighted in the middle panel) is a graphical representation used to show the cumulative distribution of wealth or income in a population. A perfectly equal distribution would follow the line of equality (a 45-degree diagonal line), but the Lorenz curve generally bows below this line, indicating inequality. The farther the curve is from the line of equality, the higher the level of inequality. The Leimkuhler curve (shown in the first panel) is closely related to the Lorenz curve. The area between the line of equality and the Leimkuhler curve can be used to measure the inequality in rank-based productivity distribution. The third panel combines the concepts of the Lorenz and Leimkuhler curves, showing the area between the Lorenz curve and the line of equality, which represents the Gini coefficient.

Intuitively productivity distribution of the random variable and it's LKC representation is equivalent. So, we can obtain the pdf from LKC.

Proposition 1
Suppose that $K(p)$ is a Leimkuhler curve and consider $K''(p)$ exists and is negative everywhere in an interval (x_1, x_2) then, F has a finite probability density in the interval $(\mu K'(x_1^+), \mu K'(x_2^+))$ which is given by

$$f(x) = \frac{-1}{\mu K''(p)\bar{F}(x)}. \tag{8}$$

Proof
The first derivative of LKC w.r.t. p implies $K'(p) = \frac{1}{\mu}F^{-1}(1-p)$, for all $0 \leq p \leq 1$, where $F(x) = 1-p$, leads to $K'(\bar{F}(x)) = \frac{x}{\mu}$. So, $K''(\bar{F}(x))(-f(x)) = \frac{1}{\mu}$ equivalent to $f(x) = \frac{-1}{\mu K''(p)\bar{F}(x)}$. □

This method is a version of Proposition 1 which is given by Burrell [9].

The ratio of the Lorenz curve at p to its value at $1-p$ measures the fraction of income that the lowest $100p\%$ of the population have relative to the upper $100p\%$ via the Leimkuhler curve. For the status of the lower and upper-income group Gastwirth [19] defined

$$J(p) = \frac{L(p)}{1-L(1-p)} = \frac{L(p)}{K(p)} \tag{9}$$

as the ratio of the total income of poorest p^{th} fraction of population to the total income of the highest p^{th} fraction and the ratio of the total income of the middle p^{th} fraction of the population to that of the upper p^{th} fraction is reflected by

$$J_m(p) = \frac{L(0.5+\frac{p}{2}) - L(0.5-\frac{p}{2})}{1-L(1-p)} = \frac{K(0.5+\frac{p}{2}) - K(0.5-\frac{p}{2})}{K(p)}. \tag{10}$$

The measure of polarization as named Wolfson's index of bipolarization originally proposed for a population divided into two groups in view of the median (Wolfson [43]). This measure can be expressed as

$$W_X = \frac{2\mu}{m}(2K(0.5) - 1 - G) \tag{11}$$

where μ, m and G are mean, median and Gini index respectively.

3. Transformation and weighted version of Leimkuhler curve

Transformations are used to present data on different scale in statistics. The nature of a transformation determines how the scale of an untransformed variable will be affected. Consider a random variable X and let $h : R \to R$ be a transformation $Y = h(X)$, then we wish to know the Leimkuhler of this transformation $(Y = h(X))$ and its link with the Leimkuhler curve of X in the general case and special cases as below:

Proposition 2
Let X be a random variable with expectation μ and Leimkuhler curve $K(.)$. Then the Leimkuhler curve of $Y = h(X)$ (h is invertible) is

$$K^*(p) = \frac{1}{E(h(X))} \int_{F^{-1}(1-p)}^{\infty} h(z)f(z)dz.$$

Proof
Via definition of LKC we have

$$K^*(p) = \frac{1}{E(Y)} \int_{1-p}^{1} D^{-1}(t)dt,$$

where D and F are distribution functions of Y and X respectively. On nothing that $D^{-1}(t) = h(F^{-1}(t))$,

$$\begin{aligned} K^*(p) &= \frac{1}{E(Y)} \int_{1-p}^{1} h(F^{-1}(t))dt \\ &= \frac{1}{E(h(X))} \int_{F^{-1}(1-p)}^{\infty} h(z)f(z)dz. \end{aligned}$$

□

Remark 1
When $Y = aX + b$, then

$$\begin{aligned} K^*(p) &= \frac{1}{a\mu + b}[a \int_{F^{-1}(1-p)}^{\infty} zf(z)dz + b(1 - F(F^{-1}(1-p)))] \\ &= \frac{a\mu K(p) + bp}{a\mu + b} \\ &= wp + (1-w)K(p) \end{aligned}$$

where $w = \frac{b}{a\mu + b}$. In this case $b = 0$ ($a = 1$) implies $K^*(p) = K(p)$ ($K^*(p) = \frac{\mu K(p) + bp}{\mu + b}$), which is Sarabia [35] achievement.

Weighted distribution with a generally weighted $w(.)$ defined by Rao [32] and Patil and Rao [29], [30] as $f_w(x) = \frac{w(x)f(x)}{E(w(X))}$ where f is the referenced distribution.
The distribution function of w can be represented as

$$F_w(x) = \frac{1}{E(w(X))} \int_0^{F(x)} w(F^{-1}(u))du = L_{w(x)}(F(x)) \ \ (K_{w(x)}(F(x))) \tag{12}$$

if $w(.)$ is increasing (decreasing). For the case $w(x) = \phi(F(x))$ where includes distribution such as order statistics, record value, Jones model, ..., etc, we have

$$F_w(x) = \frac{1}{\int_0^1 \phi(t)dt} \int_0^{F(x)} \phi(u)du = L_{\phi(F)}(F(x)) \ \ (K_{\phi(F)}(F(x)))$$

if $\phi(.)$ is increasing (decreasing). In view of Bartoszewicz [3], we have the following assertions for Leimkuhler curve in weighted cases:

- If the odd number of $w_i(.),\ i = 1, 2, \dots n$ be decreasing then, for $w(x) = w_1(w_2(\dots(w_n(x))))$, $F_w(x) = K_w(F(x))$, which can be proved by induction.

- For the case $w^*(x) = \prod_{i=1}^{n} w_i(x)$, when all $w_i(.)$ be decreasing, then $F_{w^*}(x) = K_{w^*}(F(x))$, but this relation for other choices may be correct, which is depending on $w_i's$.

For example, If X has a classical Pareto distribution with a scale parameter ($\sigma > 0$) and a shape parameter ($\alpha > 0$), we take $w(x) = x^\beta$ ($\beta < 0$), then From (12)

$$F_w(x) = K_w(F(x)) = 1 - (\frac{x}{\sigma})^{\beta-\alpha}.$$

Now let $w(x) = w_1(w_2(\ldots(w_n(x)))) = x^B$ be decreasing, where $w_i(x) = x^{b_i}$, $(b_i < 0)$, $i = 1, 2, \ldots n$ (n be a odd number), $B = \prod_{i=1}^{n} b_i$, then

$$F_w(x) = K_w(F(x)) = 1 - (\frac{x}{\sigma})^{B-\alpha}.$$

For the decreasing function $w^*(x) = \prod_{i=1}^{n} w_i(x) = x^{B^*}$, where $w_i(x) = x^{b_i}$, $(b_i < 0)$, $B^* = \sum_{i=1}^{n}(b_i)$, then

$$F_{w^*}(x) = K_{w^*}(F(x)) = 1 - (\frac{x}{\sigma})^{B^*-\alpha}.$$

4. Parametric Leimkuhler curve

Finding an appropriate functional form of LC and LKC is an important practical and theoretical problem. In this section, some functional forms of LKC are presented:

Proposition 3
If $K(p)$ is a given LKC, then $K_\alpha(p) = [K(p)]^\alpha$ is also a LKC if and only if

$$0 < \alpha < 1 - \frac{K(p).K''(p)}{(K'(p))^2} \quad for\ \ all\ \ 0 < p < 1. \tag{13}$$

Proof
The proof is similar to which mentioned in Proposition 2 of Burrell [9]. □

- Let $K(p) = p$, then $K_\alpha(p) = p^\alpha$ is also a LKC if and only if $\alpha \in [0, 1]$.
- Let $K_1(p)$ and $K_2(p)$ be two LKCs, the functional form $K_{\alpha,\beta}(p) = 1 - (1 - K_1(p))^\alpha (1 - K_2(p))^\beta$, where $\alpha, \beta \geq 1$ is a genuine LKC, because it is easy to see that $K_{\alpha,\beta}(0) = 0, K_{\alpha,\beta}(1) = 1$, $K'_{\alpha,\beta}(p) \geq 0$ and $K''_{\alpha,\beta}(p) \leq 0, p \in (0, 1)$.
- Let $K(p)$ be a strictly increasing baseline Leimkuhler curve, $K_{\alpha,\gamma}(p) = [1 - K^{-1}(1 - p)]^\alpha p^\gamma$, where $\alpha > 0$ and $\gamma \geq 1$, then it is a new functional form that a genuine LKC.

Some of the Leimkuhler forms characterize special distribution as follows:

Proposition 4
Suppose that X is a non-negative random variable, have uniform distribution $\mathcal{U}(0, 2c\mu_X)$ if and only if $K(p) = a + bp + cp^2$ for $c < 0, b > 0, a \in R$.

Proof
On using the conditions of LKC, it is trivial. □

In view of Sarabia et al. [39] we obtain following Proposition:

Proposition 5
Let $K_i(p), i = 1, 2$ be two LKCs then the functional form $K_{12}(p) = K_1(K_2(p)), 0 \leq p \leq 1$ defines a new genuine LKC.

Proof
The following conditions imply the proof:

$(i) K_{12}(0) = K_1(K_2(0)) = K_1(0) = 0, \quad K_{12}(1) = K_1(K_2(1)) = K_1(1) = 1$
$(ii) K'_{12}(p) = (K_1(K_2(p))' = K'_1(K_2(p)).K'_2(p) \geq 0$
$(iii) K''_{12}(p) = K'_2(p).K''_1(K_2(p)) + K''_2(p).K'_1(K_2(p)) \leq 0, \quad for \ \ all \ \ 0 < p < 1.$

□

The extension of the Proposition can be expressed as below:

- Let $K_i(p), i = 1, 2, ..., k$ be LKC then the functional form $K_{12..k}(p) = K_1(K_2...K_k(p))$ is a new genuine LKC too, which can be achieved via mathematical induction supposing that $K_{k-1}(p) = K_1(K_2(...K_{k-1}(p)))$ is a LKC.
- Let G_1, G_2 be two Gini indices corresponding to Leimkuhler curves $K_1(p)$ and $K_2(p)$. If $K_1(p) < K_2(p)$ for every $p \in (0, 1)$. Then $G_1 < G_2$. i.e. Gini index is increasing functional of LKC.

In view of hybrid models of the LC (Ogwang et al. [28]) we have this assertion for LKC:

Proposition 6
Let $Y_1 = K_1(p)$ and $Y_2 = K_2(p)$ be two Leimkuhler curves with Gini indices G_1 and G_2 respectively. Then a convex combination $Y = K(p) = \delta K_1(p) + (1-\delta)K_2(p)$ of two models is a LKC, for all $0 \leq \delta \leq 1$ and its corresponding Gini index is $G = \delta G_1 + (1-\delta)G_2$.

Proof
The proof follows from the conditions of LKC:
$K(0) = \delta K_1(0) + (1-\delta)K_2(0) = 0$ and $K(1) = \delta K_1(1) + (1-\delta)K_2(1) = 1$ and $K'(p) = \delta K'_1(p) + (1-\delta)K'_2(p) \geq 0$ beacause $K'_1(p), K'_2(p) \geq 0$
$K''(p) = \delta K''_1(p) + (1-\delta)K''_2(p) \leq 0$ because $K''_1(p), K''_2(p) \leq 0$. □

A possible solution for obtaining better fit consists in building more complex models combining some of the classical models using the convex linear combination of LKCs.

Remark 2
Let $K_1(p), K_2(p), \ldots, K_m(p)$ be Leimkuhler curves, the following $K(p)$ are LKCs for $i = 1, 2, \ldots, m$.

(i) $K(p) = \sum_{i=1}^{m} w_i K_i(p), \ \ w_i \geq 0, \ \ \sum_{i=1}^{m} w_i = 1.$
(ii) $K(p) = 1 - \prod_{i=1}^{m}(1 - K_i(p))^{\alpha}, \ \ \alpha \geq 0.$
(iii) $K(p) = min\{K_1(p), K_2(p), \ldots, K_m(p)\}, \ \ p \in [0, 1].$

Let X be a non-negative integer real value of random variable with probability mass function $p_j = P(X = j); j = 0, 1, 2, 3, \cdots$ where $\sum_{j=0}^{\infty} p_j = 1$, then probability generating function of it can be expressed as $A_X(s) = E(s^X) = \sum_{j=0}^{\infty} p_j s^j$ which is convergent for $|s| < 1$. We state assertions related to the Leimkuhler curve and moment generating function via the similar arguments in Sarabia et al. [37]. Notes based on Leimkuhler curve the below Proposition is derived.

Proposition 7
Let $K_0(p)$ be a LKC curve where the first and second derivative of it w.r.t. p exist. Consider X be a discrete random variable which takes value $1, 2, 3, \cdots$ with probability $\pi_j = P(X = j) \geq 0$ for $j = 1, 2, \cdots$ and $A_X(.)$ be the corresponding generating function of X which the first and second derivative of it w.r.t. s exist. Then $\tilde{K}(p, \pi) = 1 - A_X(1 - K_0(p))$ defines a LKC.

Proof
We have $\tilde{K}(0, \pi) = 1 - A_X(1 - K_0(0)) = 1 - A(1) = 0$,
$\tilde{K}(1, \pi) = 1 - A_X(1 - K_0(1)) = 1 - A(0) = 1$,
$\tilde{K}'(p, \pi) = K'_0(p)A'_X(1 - K_0(p)) > 0$ and

$\tilde{K}''(p,\pi) = K_0''(p)A_X'(1-K_0(p)) - (K_0'(p))^2 A_X''(1-K_0(p)) < 0,$
hence $\tilde{K}(p,\pi)$ is a LKC. □

As an example assume that $\pi_j = \dfrac{1}{n},\ \ j = 1,2,\dots$ when $n = 1,2,\dots$ and $A_X(z) = \dfrac{1-z^n}{n(1-z)}$, then $\tilde{K}(p,\pi) = 1 - A_X(1-K_0(p)) = 1 - \dfrac{1-(1-K_0(p))^n}{nK_0(p)}$.

5. Leimkuhler ordering

We can use some stochastic ordering about LKC. If two Leimkuhler curves do not intersect, they can be ordered.

Definition 1
Suppose that X_1 and X_2 be two non-negative random variables with positive finite expectation, the random variable X_1 is said to be at least as unequal as X_2 in the Leimkuhler sense $(X_1 \leq_{LKC} X_2)$ if $K_1(p) \leq K_2(p)$ for all $p \in [0,1]$, that is

$$X_1 \leq_{LKC} X_2 \Longleftrightarrow K_1(p) \leq K_2(p), \quad 0 \leq p \leq 1. \tag{14}$$

Another well-known partial order to compare the skewness of two probability distribution is the star-shaped order. The star-shaped order is stronger than the Leimkuhler order.
Let X_1 and X_2 be two non-negative random variables with distribution functions F_1 and F_2 respectively. Also, star-shaped ordering is defined as follows (Arnold et al. [1]):

Definition 2
X_1 is star-shaped with respect to X_2, $(X_1 \leq_\star X_2)$ if $\dfrac{F_1^{-1}(x)}{F_2^{-1}(x)}$ is a non-increasing function of x.

The star-shaped ordering implies Leimkuhler ordering as seen in below:

Proposition 8
Suppose that X_1 and X_2 be two non-negative random variables with positive finite expectation. If $X_1 \leq_\star X_2$, then $X_1 \leq_{LKC} X_2$ and $K_1(p) \leq K_2(p)$ for all $p \in [0,1]$.

Proof
The proof is similar to that of Theorem 9.3 of Sarabia [36] about LC. □

In view of Theorem 1 of Moothathu [27] we have the following property:

Lemma 1
Let X_1 and X_2 be two positive random variables with quantile functions F_1^{-1} and F_2^{-1}, and Leimkuhler curves $K_1(p)$ and $K_2(p)$ respectively and let $U(t) = \frac{F_1^{-1}(t)}{F_2^{-1}(t)}$. Then, for every p in $(0,1)$,
$(i) K_1(p) > K_2(p) \ \ if \ \ U(t)\ is\ increasing\ in (0,1).$
$(ii) K_1(p) < K_2(p) \ \ if \ \ U(t)\ is\ decreasing\ in (0,1).$
$(iii) K_1(p) = K_2(p) \ \ if \ \ U(t) = C, constant\ in (0,1).$

Proof
The proof is similar to that of Theorem 1 of Fellman [18].

□

6. Distorted Leimkuhler curve

A distortion function is known as a probability transformation function. Risk theory based on distorted probability can be considered as a dual theory of choice under risk in the sense that it uses the notion of distortion function as opposed to the utility function used in classical utility theory. Here we use the concept of distortion function which is an increasing function $h : [0,1] \to [0,1]$ such that $h(0) = 0$ and $h(1) = 1$; in connection with LKC. Some aspects of LKC related to distortion cases are as below:

Proposition 9
Let $K(p)$ be a LKC and h be a twice differentiable distortion function such that $h''(t) \geq 0,\ t \in (0,1)$. Then $\tilde{K}(p) = h(K(p)), p \in [0,1]$ can not defined a LKC; always.

Proof
Simply see $\tilde{K}(0) = 0$ and $\tilde{K}(1) = 1$. The first derivative of $\tilde{K}(p)$ is $\tilde{K}'(p) = h'(K(p))K'(p)$ where h and K are increasing implies $\tilde{K}'(p) \geq 0$, with second derivative of $\tilde{K}'(p)$ conclude $\tilde{K}''(p) = h''(K(p))(K'(p))^2 + h'(K(p))K''(p)$ since $h'' > 0$, it should be $\tilde{K}''(p) \nleq 0$. □

Based on the above Proposition and its arguments some results are as a remark.

Remark 3
Let $K(p)$ and $h(p)$ be Leimkuhler curve and distorted function respectively, then

(i) For $\tilde{K}(p) = 1 - h(1 - K(p))$, $\tilde{K}(p)$ is also a LKC.
(ii) Let $h(p) = K_1(p)$ be a LKC as a distorted function, then $\tilde{K}(p) = K_1(K(1-p))$ defines also a LKC. As an example, consider $K_1(p) = \dfrac{p}{1-(1-\theta)(1-p)}$, $\theta \in (0,1]$, $p \in [0,1]$ then $\tilde{K}(p) = \dfrac{K(p)}{1-(1-\theta)(1-K(p))} = K_1(K(p))$ is also a LKC.
(iii) Let X be a discrete random variable take integer value with probability $p_j = P(X = j) > 0$ and A be the probability generating function parallel to Sordo et al. [41],
$$\tilde{K}(p) = 1 - (1 - K(p))^{\alpha} A(1 - K(p))$$
with $\alpha \geq 1$ is also a LKC. For example let $X \sim P(\theta)$, then $A(t) = E(t^X) = exp\{-(1-t)\theta\}$ so, $\tilde{K}(p) = 1 - (1 - K(p))^{\alpha} exp\{-\theta K(p)\}$ is also a LKC.
(iv) Let $h_i(.), K_i(p)(i = 1, \ldots, k)$ be k distorted functions and k Leimkuhler curves respectively, then
$$\tilde{K}(p) = 1 - \prod_{i=1}^{k} h_i(1 - K_i(p))$$
is a LKC.

For two distorted functions the following Proposition is noticeable:

Proposition 10
Let $h_i(.), K_i(p), (i = 1, 2)$ be two distorted functions and two LKCs, then
$$\tilde{K}(p) = 1 - h_1(1 - K_1(p))h_2(1 - K2(p))$$
is a LKC.

Proof
On noting $K_i(p)\ \ i = 1, 2$ are LKCs, then $\tilde{K}(0) = 0$ and $\tilde{K}(1) = 1$.
$\tilde{K}'(p) = K_1'(p)h_1'(1 - K_1(p))h_2(1 - K_2(p)) + K_2'(p)h_1(1 - K_1(p))h_2'(1 - K_2(p))$ which is positive. Drive twice of $\tilde{K}'$ w.r.t. p implies
$$\begin{aligned}\tilde{K}''(p) = {} & K_1''(p)h_1'(1 - K_1(p))h_2(1 - K_2(p)) \\ & - (K_1'(p))^2 h_1''(1 - K_1(p))h_2(1 - K_2(p)) \\ & - 2K_1'(p)K_2'(p)h_1'(1 - K_1(p))h_2'(1 - K_2(p)) \\ & + K_2''(p)h_1(1 - K_1(p))h_2'(1 - K_2(p)) \\ & - (K_2'(p))^2 h_1(1 - K_1(p))h_2''(1 - K_2(p))\end{aligned}$$

is negative. So, $\tilde{K}(.)$ is a LKC. □

7. Risk measure and Leimkuhler curve

Is the Leimkuhler curve can be related to risk measure? For answering this question we have studied this relation via the following parallel to Lorenz curve arguments:

- Tail Value at risk is considered as the arithmetic mean of the VaRs of X from p on. $TVaR(X;p) = \frac{\mu}{1-p}K(1-p), \quad p \in (0,1)$.
- Conditional tail expectation as $CTE(X;p) = E(X|X > VaR(X;p)) = \frac{\mu}{1-p}K(F^{-1}(p))$ and conditional value at risk can be as

$$\begin{aligned} CVaR(X;p) &= E(X - VaR(X;p)|X > VaR(X;p)) \\ &= \frac{\mu}{1-p}K(F^{-1}(p)) - F^{-1}(p), \quad p \in (0,1). \end{aligned}$$

- Expected shortfall can be expressed by

$$\begin{aligned} ES(X;p) &= \int_{F^{-1}(p)}^{\infty} (x - F^{-1}(p))f(x)dx \\ &= \mu K(1-p) - (1-p)F^{-1}(p), \quad p \in (0,1). \end{aligned}$$

- The expected proportional shortfall function defined by Belzunce et al. [5] as $EPS(X;p) = E((\frac{X - VaR(X;p)}{VaR(X;p)})_+)$ where $(x)_+ = max\{x, 0\}$. It is expressible in terms of LKC also:

$$EPS(X;p) = \frac{\mu K(1-p) - (1-p)F^{-1}(p)}{F^{-1}(p)}, \quad p \in (0,1).$$

- A relative version of $TVaR$ is given by

$$TVaRD = \frac{TVaR(X;p)}{E(X)} = \frac{K(1-p)}{1-p}.$$

- Given two random variables X and Y with respective distribution function F and G, X is less than Y in stochastic order if $VaR(X;p) \leq VaR(Y;p)$ which implies $F^{-1}(p) \leq G^{-1}(p), \quad for \ p \in (0,1)$. It can be led to $E(X)K_X(p) \leq E(Y)K_Y(p)$ and if $E(X) = E(Y)$ it leads to $K_X(p) \leq K_Y(p)$.
- The tail value-at-risk is closely related to the increasing convex order that Sordo [40] obtained increasing convex ordering can be characterized $TVaR$ as

$$X \leq_{icx} Y \Rightarrow TVaR(X;p) \leq TVaR(Y;p) \Rightarrow E(X)K_X(p) \leq E(Y)K_Y(p), \ \ p \in (0,1).$$

- Let X and Y be two non-negative random variables with interval support; if $X \leq_{PS} Y$, then $K_Y(p) \geq K_X(p)$. Note that $X \leq_{PS} Y$ if $EPS(X;p) \leq EPS(Y;p), \quad$ for $\quad p \in D_X \cap D_Y$.

8. Connection between Leimkuhler curve and other criteria

The links between the Leimkuhler curve and other criteria which are important in economics and informetrics are as below:

- Zenga [45] introduced a new inequality curve $Z(p)$, based on the ratio between the lower mean $M^{-}(p)$ and the upper mean $M^{+}(p)$ of non-negative random variable by $Z(p) = 1 - \dfrac{M^{-}(p)}{M^{+}(p)}$, $p \in (0,1)$ where $M^{-}(p) = \dfrac{\int_0^p F^{-1}(t)dt}{p}$ and $M^{+}(p) = \dfrac{\int_p^1 F^{-1}(t)dt}{1-p}$.
Based on Zenga [45], it can be proved the following link between $Z(p)$ and $K(p)$ curves:

$$K(p) = \frac{p}{1-(1-p)[Z(1-p)]}$$

- Crow [13] defined the population double cumulative curve as a Lorenz curve with origin moved to $(1,1)$ and axes rotated 180° in the plane of the paper. It is applied in weather modification. Also, it is similar to the Leimkuhler curve which is applicable in informetrics studies.
- The Right Concentration (RC) curve is defined by Belzunce et al. [4] as

$$RC(p) = \frac{1}{F^{-1}(p)} \int_{F^{-1}(p)}^{+\infty} \overline{F}(x)dx + 1$$

where can be expressed the LKC in related to RC curve as

$$K(p) = \frac{F^{-1}(1-p)}{\mu}[RC(1-p) - p].$$

- Arora et al. [2] defined the concept of cumulative mean income curve (COMIC) which has an important role in measuring inequality. It is the mean income of the lower of the $\%100p$ of the income as

$$C(p) = \frac{\mu(1-K(1-p))}{p}, \quad 0 < p < 1.$$

which shows link with Leimkuhler curve.
- An important generalization of the Gini index is defined by Donaldson and Weymark [14] and Yitzaki [44] as $G(\nu) = 1 - \nu(\nu-1)\int_0^1 (1-p)^{\nu-2} L(p)dp$. It can be expressed via LKC as below:

$$G(\nu) = 1 - \nu(\nu-1)\int_0^1 (y)^{\nu-2}(1-K(y))dy.$$

- Pietra index is another important equality measure that is defined as the maximal vertical deviation between the Lorenz curve and the egalitarian line.

$$P_L = \max_{0\le p\le 1}\{p - L(p)\}.$$

It can be obtained via LKC as

$$P_L = \max_{0\le p\le 1}\{K(p) - p\}.$$

- The E-gini index which is defined by Charkravaty [12] can be expressed in terms of LKC

$$\text{E-gini} = 2\left[\int_0^1 (K(y)-y)^{\alpha} dy\right]^{\frac{1}{\alpha}}.$$

- A hill curve ($I(u)$) is an inequality index where seen is Eliazar [17]. A vertical distances between LKC and LC as

$$I(u) = K(u) - L(u), \quad 0 < u < 1$$

and another version of it can be defined as

$$I(u) = L^{-1}(u) - K^{-1}(u), \quad 0 < u < 1.$$

9. Application

We use data on impact factors from the latest available (2019) edition of Thomson Reuters Journal Citation Reports (JCR). Following Sarabia et al. [38] we use impact factors for scientific journals belonging to the following scientific fields: Chemistry, Economics, Education, Information Science and Library Science (abbreviated as Information SLS in table), Mathematics, Neuroscience, Psychology and Physics. A field of research is considered by a JCR subject category or (for instance in the case of Mathematics) by a number of JCR subject categories taken together.
We have considered only positive figures of the impact factor, that is, we have excluded journals with zero impact factor.
Table 1 reports some summary statistics for the eight fields, including the number of journals with non-zero impact, the mean, the quartiles, the Gini and Zenga indexes.
Comparing our data with data used by Sarabia et al. [38], we observe that the number of journals index in JCR has increased between JCR 2010 Edition and JCR 2019 edition for each science category analyzed; the increases lie in the range between 15% and 60%. Other results are close to those reported by Sarabia et al. [38].
The highest mean value corresponds to the Neuroscience field with a value of 3.870 and the lowest to the Mathematics field with a value of 1.365. The Gini index is interpreted as a measure of concentration that ranges from 0 to 1 where 0 corresponds to perfect equality (every paper in a journal receives the same number of citations) and 1 corresponds to perfect inequality (all citations are received only by one single paper). Also, the Zenga index is another measure of concentration. Three main differences between Gini and Zenga indexes. First, the Zenga index increases more rapidly for low values of the variation and decreases more slowly when the variation approaches intermediate values from above. Second, the Zenga index seems to be better predicted by the variation. Third, although the Zenga index is always higher than the Gini one, the ordering of some pairs of cases may be inverted. The highest Gini and Zenga values correspond to Physics with a value of 0.539 and 0.832 respectively, that higher Gini and Zenga indexes indicate greater inequality and the lowest Gini and Zenga values correspond to Education with a value of 0.314 and 0.656 that indicate smaller inequality.

Table 1. Summary statistics of the IF (empirical values) for selected fields, N denoted the number of journals with positive impact factor.

Field	N	Mean	First quartile	Median	Third quartile	Std. Dev.	Gini index	Zenga index
Chemistry	654	3.828	1.417	2.406	4.181	5.044	0.503	0.821
Economic	371	1.805	0.846	1.414	2.240	1.485	0.399	0.739
Education	347	1.735	0.997	1.559	2.189	1.065	0.314	0.656
Information SLS	87	2.046	0.788	1.580	2.715	1.624	0.427	0.783
Mathematics	749	1.365	0.688	1.053	1.642	1.208	0.373	0.704
Neuroscience	271	3.870	2.224	3.056	4.426	3.397	0.371	0.715
Psychology	761	2.349	1.156	1.869	2.703	2.352	0.402	0.742
Physics	482	3.494	1.224	1.970	3.163	5.306	0.539	0.832

Figure 2 displays four panels, each panel includes the Lorenz curve and the Leimkuhler curve. The Gini coefficient represented by the area between the Lorenz curve and the line of equality. A larger area between the Lorenz curve and the line of equality indicates greater inequality. Chemistry appears to have the highest level of inequality among the four fields, as indicated by the larger Gini coefficient, while Education shows the least. Economics and Mathematics fall between the two, with Economics displaying more inequality than Mathematics. Figure 3 presents Lorenz and Leimkuhler curves for Information, Physics, Psychology, and Neurosciences. Neurosciences has the most equal distribution of productivity or resources, with the smallest Gini coefficient. Psychology and Information show increasing levels of inequality, with Information displaying more inequality than Psychology. Physics has the highest inequality, as indicated by the largest Gini coefficient.

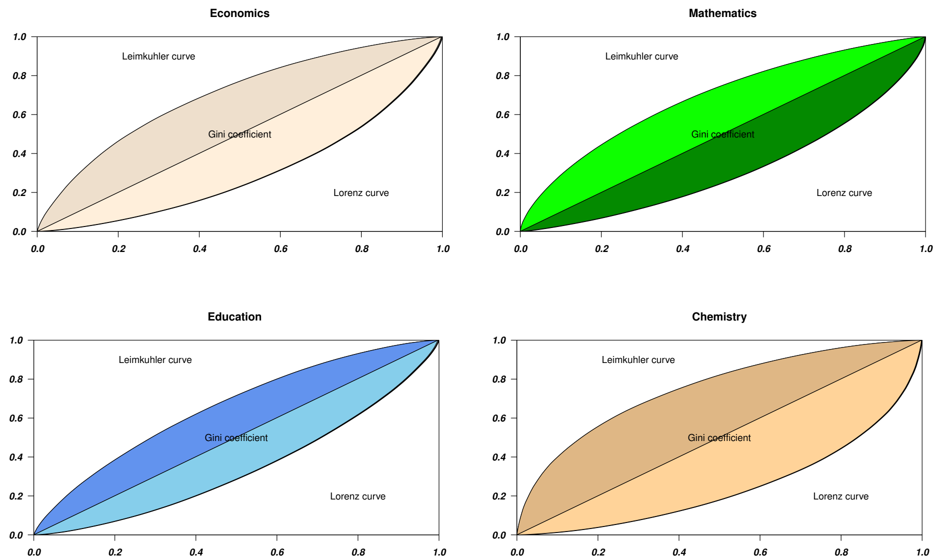

Figure 2. Lorenz and Leimkuhler curves for the impact factor in four categories: Economics, Mathematics, Education and Chemistry

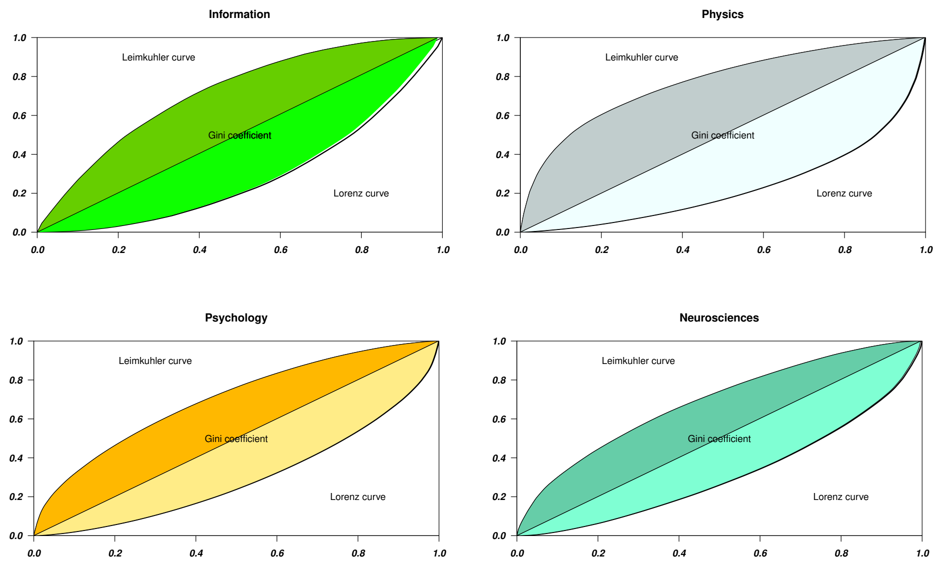

Figure 3. Lorenz and Leimkuhler curves for the impact factor in four categories: Information science and Library science, Physics, Psychology and Neurosciences

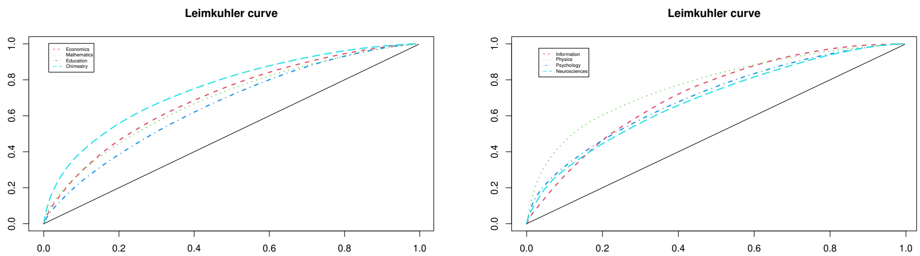

Figure 4. Leimkuhler curves for the impact factor in four categories: Economics, Mathematics, Education and Chemistry (left) and Information science and Library science, Physics, Psychology and Neurosciences (right)

Figure 4 presents two graphs, both depicting Leimkuhler curves for the impact factors of journals in different academic categories. In the left figure, the Leimkuhler curve for Chemistry and the Leimkuhler curve for Education are the farthest and closest curves to the equality line, respectively, confirming that they have the highest and lowest Gini coefficients, as shown in Figure 2. Similarly, in the right figure, the Leimkuhler curves for Physics and Neurosciences are, respectively, the farthest and the closest to the equality line, which confirms that they have the highest and lowest Gini coefficients, as shown in Figure 3.

10. Conclusion

Leimkuhler curve as an inequality measure which is mostly applied in informetrics has properties that are achieved some of them here.
Also, for variant families and cases, the connections with the Leimkuhler curve are obtained with attached remarks. Applications with data taken from Thomson Reuters Journal Citation Reports Edition 2019, in eight scientific fields are considered.

Acknowledgement

The authors would like to thank the Associate Editor and referees for constructive suggestions which have helped to improve the paper significantly.

REFERENCES

1. B. C. Arnold, and J.M. Sarabia, *Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics* Cham: Springer, Vol. 7, 2018.
2. S. Arora, K. Jain, and S. Pundir, *On cumulated mean income curve*, Model Assisted Statistics and Applications, vol. 1, no. 2, pp. 107–114, 2005.
3. J. Bartoszewicz, *On a representation of weighted distributions*, Statistics and Probability Letters, vol. 79, no. 15, pp. 1690–1694, 2009.
4. F. Belzunce, J. F. Pinar, J. M. Ruiz, and M. A. Sordo, *Comparison of concentration of several parametric families of income distributions*, In Proceeding of the International Conference on Mathematical and Statistical Modelling in Honor of Enrique Castillo, 2006.
5. F. Belzunce, J. F. Pinar, J. M. Ruiz, and M. A. Sordo, *Comparison of risks based on the expected proportional shortfall*, Insurance: Mathematics and Economics, vol. 51, no. 2, pp. 292–302, 2012.
6. S. C. Bradford, *Sources of information on specific subjects*, Engineering, vol. 137, pp. 85–86, 1934.
7. Q. L. Burrell, The 80/20 rule: Library lore or statistical law? Journal of Documentation, vol. 41, no. 1, pp. 24–39, 1985.
8. Q. L. Burrell, *The Bradford distribution and the Gini index*, Scientometrics, vol. 21, pp. 181–194, 1991.
9. Q. L. Burrell, *Symmetry and other transformation features of Lorenz/Leimkuhler representations of informetric data*, Information Processing and Management, vol. 41, no. 6, pp. 1317–1329, 2005.
10. Q. L. Burrell, *On Egghe's version of continuous concentration theory*, Journal of the American Society for Information Science and Technology, vol. 57, no. 10, pp. 1406–1411, 2006.

11. M. P. Carpenter, *Similarity of Pratt's measure of class concentration to the Gini index*, Journal of the American Society for Information Science, vol. 30, no. 2, pp. 108–110, 1979.
12. S. R. Chakravarty, *Extended Gini indices of inequality*, International Economic Review, pp. 147–156, 1988.
13. E. L. Crow, *Double cumulative and Lorenz curves in weather modification*, Journal of Applied Meteorology, vol. 21, no. 8, pp. 1063–1070, 1982.
14. D. Donaldson, and J. A. Weymark, *A single-parameter generalization of the Gini indices of inequality*, Journal of Economic Theory, vol. 22, no. 1, pp. 67–86, 1980.
15. L. Egghe, *Applications of the theory of Bradford's law to the calculation of Leimkuhler's law and to the completion of bibliographies*, Journal of the American Society for Information Science, vol. 41, no. 7, pp. 469–492, 1990.
16. L. Egghe, *Duality aspects of the Gini index for general information production processes* Information Processing and Management, vol. 28, no. 1, pp. 35–44, 1992.
17. I. Eliazar, *A tour of inequality*, Annals of Physics, vol. 389, pp. 306–332, 2018.
18. J. Fellman, *The effect of transformation on Lorenz curve* Econometrica, vol. 44, pp. 823–824, 1976.
19. J. L. Gastwirth, *Measures of economic inequality focusing on the status of the lower and middle income groups* Statistics and Public Policy, vol. 3, no. 1, pp. 1–9, 2016.
20. C. Gini, *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche.[Fasc. I.]* Tipogr. di P. Cuppini.
21. L. Hannah, and J. A. Kay, *Concentration in modern industry: Theory, measurement and the UK experience*, Springer, 1977.
22. S. P. Jenkins, and P. J. Lambert, *Three 'I's of poverty curves, with an analysis of UK poverty trends* Oxford economic papers, vol. 49, no. 3, pp. 317–327, 1997.
23. C. Kleiber, and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences* John Wiley and Sons, 2003.
24. F. F. Leimkuhler, *The Bradford distribution* Journal of Documentation, vol. 23, no. 3, pp. 197–207, 1967.
25. M. O. Lorenz, *Methods of measuring the concentration of wealth* Publications of the American Statistical Association, vol. 9, no. 70, pp. 209–219, 1905.
26. A. J. Lotka, *The frequency distribution of scientific productivity* Journal of the Washington Academy of Sciences, vol. 16, no. 12, pp. 317–323, 1926.
27. T. S. K. Moothathu, *On a sufficient condition for two non-intersecting Lorenz curves* Sankhyā: The Indian Journal of Statistics. vol. 53, pp. 268–274, 1991.
28. T. Ogwang, and U. G. Rao, *Hybrid models of the Lorenz curve* Economics Letters, vol. 69, no. 1, pp. 39-44, 2000.
29. G. P. Patil, and C. R. Rao, *The weighted distributions: A survey of their applications* Applications of statistics, pp. 383–405, 1977.
30. G. P. Patil, and C. R. Rao, *Weighted distributions and size-biased sampling with applications to wildlife populations and human families*, Biometrics, pp. 179–189, 1978.
(1978). .
31. A. D. Pratt, (1977) *A measure of class concentration in bibliometrics*, Journal of the American Society for Information Science, vol. 28, no. 5, pp. 285–292, 1977.
32. C. R. Rao, *On discrete distributions arising out of methods of ascertainment* Sankhyā: The Indian Journal of Statistics, Series A, pp. 311–324, 1965.
33. R. Rousseau, *Double exponential models for first-citation processes* Scientometrics, vol. 30, no. 1, pp. 213–227, 1994.
34. J. M. Sarabia, E. Castillo, and D. J. Slottje, *An ordered family of Lorenz curves* Journal of Econometrics vol. 91, pp. 43–60, 1999.
35. J. M. Sarabia, *A general definition of the Leimkuhler curve* Journal of Informetrics, vol. 2, no. 2, pp. 156–163, 2008.
36. J. M. Sarabia, *A general definition of the Leimkuhler curve* Modeling income distributions and Lorenz curves, Springer, pp. 167–190, 2008a.
37. J. M. Sarabia, E. Gómez-Déniz, M. Sarabia, and, F. Prieto, *A general method for generating parametric Lorenz and Leimkuhler curves* Journal of Informetrics, vol. 4, no. 4, pp. 524–539, 2010.
38. J. M. Sarabia, F. Prieto, and C. Trueba, *Modeling the probabilistic distribution of the impact factor* Journal of Informetrics, vol. 6, no. 1, pp. 66–79, 2012.
39. J. M. Sarabia, V. Jordá, and C. Trueba, *The Lamé class of Lorenz curves*, Communications in Statistics-Theory and Methods, vol. 46, no. 11, pp. 5311–5326, 2017.
40. M. A. Sordo, *Characterizations of classes of risk measures by dispersive orders* Insurance: Mathematics and Economics, vol. 42, no. 3, pp. 1028–1034, 2008.
41. M. A. Sordo, J. Navarro, and J. M. Sarabia, *Distorted Lorenz curves: models and comparisons*, Social Choice and Welfare, vol. 42, no. 4, pp. 761–780, 2014.
42. R. L. Trueswell, *Some behavioral patterns of library users: The 80/20 rule*, Wilson Library Bulletin, vol. 43, pp. 458–461, 1969.
43. M. C. Wolfson, *When inequalities diverge* The American Economic Review, vol. 84, no. 2, pp. 353–358, 1994.
44. S. Yitzhaki, *On an extension of the Gini inequality index*, International Economic Review, pp. 617–628, 1983.
45. M. Zenga, *Inequality curve and inequality index based on the ratios between lower and upper arithmetic means*, Statistica & Applicazioni, vol. 5, no. 1, pp. 3–27, 2007.