ANALYSIS OF INDICES OF ECONOMIC INEQUALITY FROM A MATHEMATICAL POINT OF VIEW[1]

Ričardas Zitikis[2]

Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, Ontario, Canada N6A 5B7 (*E-mail: zitikis@stats.uwo.ca*)
and
Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Ontario, Canada K1S 5B6.

A number of indices of economic inequality have been proposed in the literature. Their constructions are based on various econometric motives and justifications such as axioms of fairness. In this paper we analize the indices stepping slightly aside from their econometric meanings and adopting a mathematical approach that treats the indices as distances – in some functional spaces – between the egalitarian and actual Lorenz curves. More specifically, starting with, and being guided by, the econometric definitions of various indices, we modify the indices in such a way that the resulting ones become natural from the mathematical point of view. It turns out that some of the new "mathematical" indices coincide with the corresponding well known "econometric" ones, some appear to be only asymptotically equivalent, and some turn out to have different asymptotic behaviour when the sample size indefinitely increases.

1. Introduction and preliminaries

Suppose we are interested in the distribution of income in a society. We randomly select n individuals from the society and record their incomes: $x_1, x_2, \ldots, x_n$. The incomes can be either negative (if individuals are in debt) or non-negative numbers. We

[1]This paper is based on the author's invited lecture at the 11th Indonesian Mathematics Conference (July 22nd -25th, 2002) at the State University of Malang, Malang, Indonesia.

[2]Research partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada individual research grant at the University of Western Ontario, as well as by the Royal Netherlands Academy of Arts and Sciences (KNAW) under the auspices of the binational Extended Program in Applied Mathematics (EPAM 2000/2004) between Indonesia and The Netherlands.

then order the incomes in ascending order obtaining: $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$. Given these ordered values, we finally calculate, for any $k = 0, 1, \ldots, n$, the proportions of income possessed by the least fortunate $(k/n) \times 100\%$ individuals. In mathematical terms, these proportions are:

$$l_{k,n} := \left(\sum_{i=1}^{k} x_{i:n} \right) \Big/ \left(\sum_{i=1}^{n} x_{i:n} \right), \quad k = 0, 1, \ldots, n. \tag{1.1}$$

Note that $l_{k,n}$ are well defined provided that the denominator in (1.1) is not zero, which is equivalent to

$$\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_n \neq 0. \tag{1.2}$$

Throughout the paper we therefore assume (1.2) unless otherwise is explicitly indicated. Note that when $k = 0$, then the sum in (1.2) is empty and, therefore, equals 0 by definition. Consequently, $l_{0,n} = 0$. Note also that $l_{n,n} = 1$.

To visualize proportions (1.1), we follow a suggestion of Lorenz (1905) and plot the points $(k/n, l_{k,n})$, $k = 0, 1, \ldots, n$, on the real plane, finally connecting them by straight lines. As the result of this, we obtain the curve L_n called the (empirical) Lorenz curve. The curve L_n is well defined on the entire interval $[0, 1]$, with values $L_n(0) = 0$ and $L_n(1) = 1$. In mathematical terms, L_n can be written as follows:

$$L_n(t) = \begin{cases} C_n(t) \big/ (n\bar{x}), & 0 \leq t < 1, \\ 1, & t = 1, \end{cases} \tag{1.3}$$

where

$$C_n(t) := \sum_{i=1}^{[tn]} x_{i:n} + (tn - [tn])\, x_{[tn]+1:n}$$

with $[tn]$ denoting the largest integer not exceeding tn. Comparing formulae (1.3) and (1.1), we see that $L_n(k/n) = l_{k,n}$ for any $k = 0, 1, \ldots, n$.

The function C_n is convex. Consequently, if $\bar{x} > 0$, then L_n is also convex, and if $\bar{x} < 0$, then L_n is concave. Since $L_n(0) = 0$ and $L_n(1) = 1$, we therefore conclude that the Lorenz curve L_n is always either below (when $\bar{x} > 0$) or above (when $\bar{x} < 0$) the diagonal

$$I(t) := t, \quad 0 \leq t \leq 1.$$

The diagonal I, on the other hand, is also an empirical Lorenz curve. Indeed, assuming that all the incomes $x_1, x_2, \ldots, x_n$ are equal, we obtain from formula (1.3) that the corresponding Lorenz curve L_n is identically equal to the diagonal I. Thus, the interpretation of I as the "egalitarian" Lorenz curve.

Based on the discussion above, it now becomes natural to measure the economic inequality present in the sample $x_1, x_2, \dots, x_n$ by using some "distance"

$$d(I, L_n) \tag{1.4}$$

between the egalitarian Lorenz curve I and the actual one L_n. If the sample size is sufficiently large, then we may even consider $d(I, L_n)$ as a measure of economic inequality in the whole population. We shall briefly touch upon the latter subject in Section 7 below.

When introducing $d(I, L_n)$ in (1.4), we wrote *distance* in quotation marks. This was intended to indicate that in the context of the current paper we are not actually concerned whether $d(\cdot, \cdot)$ is, or is not, a distance (i.e. metric) on the set of all empirical Lorenz curves. The main idea behind the construction of $d(\cdot, \cdot)$ is based on the fact that we are merely interested in measuring the distance between I and L_n. This implies that we are really interested only in the functionals

$$\mathcal{D}(L_n) := d(I, L_n)$$

defined on the set of all empirical Lorenz curves L_n. It is natural to require the functional $\mathcal{D}$ be such that

1) $\mathcal{D}(L_n) \geq 0$,
2) $\mathcal{D}(L_n) = 0$ if $L_n = I$,
3) $\mathcal{D}(L_n^*) \geq \mathcal{D}(L_n^{**})$ whenever $L_n^* \leq L_n^{**}$.

In what follows we shall construct and discuss a number of such functionals $\mathcal{D}$.

2. A prelude to classical indices

When thinking about comparing two curves – for example, I and L_n in the context of this paper – at the very outset we are usually interested in the maximal distance between the two curves. This suggests, for example, using

$$\mathcal{D}_\infty(L_n) := \sup_{0 \leq t \leq 1} |t - L_n(t)|$$

as a measure of deviation between I and L_n. Naturally, as a measure of economic inequality, $\mathcal{D}_\infty(L_n)$ may not be highly informative. The reason for this is that the difference $t - L_n(t)$ is always negligible near the two end-points $t = 0$ and $t = 1$ irrespectively of the values of $x_1, x_2, \dots, x_n$. The situation can, nevertheless, be rectified by making the difference $t - L_n(t)$ more, or even less, visible by employing various weight function $w : (0, 1) \to (0, \infty)$ and modifying $\mathcal{D}_\infty(L_n)$ as follows:

$$\mathcal{D}_{\infty,w}(L_n) := \sup_{0 \leq t \leq 1} |t - L_n(t)| w(t).$$

Note that $\mathcal{D}_{\infty,w}(L_n)$ is well defined and finite for any weight function $w : (0,1) \to (0,\infty)$ satisfying the assumption

$$\sup_{0\le t\le 1} t(1-t)w(t) < \infty.$$

This is so because the difference $t - L_n(t)$ is asymptotically equivalent to t when $t \downarrow 0$ and to $1-t$ when $t \uparrow 1$. We may therefore choose, for example, to work with weight functions of the form $t \mapsto t^{p_0}(1-t)^{p_1}$, where $p_0, p_1 \ge -1$ are some fixed parameters. Note that if $-1 \le p_1 < 0$, then we emphasize the difference between I and L_n near 1 and de-emphasizes it when $p_1 > 0$.

3. Unifying classical indices into one

The area between I and L_n appears to be more interesting and fruitful as a measure of economic inequality. In mathematical terms, the area can be written as follows:

$$\mathcal{D}_1(L_n) := \int_0^1 |t - L_n(t)|\, dt. \tag{3.1}$$

In order to emphasize or de-emphasize the smallness of $|t - L_n(t)|$ near the end-points $t = 0$ and $t = 1$, we can may modify the integral of (3.1) in, for example, two ways: as $(\int_0^1 |t - L_n(t)|^p dt)^{1/p}$ for some $p > 0$, or as $\int_0^1 |t - L_n(t)|\, w(t)dt$ for some $w : (0,1) \to (0,\infty)$. Unifying the two approaches gives us the following general index of economic inequality:

$$\mathcal{D}_{p,w}(L_n) := \left(\int_0^1 |t - L_n(t)|^p\, w(t)dt\right)^{1/p},$$

where $0 < p < \infty$ is fixed and $w : (0,1) \to (0,\infty)$ is such that

$$\int_0^1 t^p(1-t)^p w(t)dt < \infty. \tag{3.2}$$

The index $\mathcal{D}_{p,w}(L_n)$ covers – as special cases – many of the well known and widely used indices of economic inequality. We shall now discuss some of them in greater detail, assuming throughout the rest of the paper that

$$\bar{x} > 0. \tag{3.3}$$

In this case, as we have already mentioned above, the Lorenz curve L_n is on or below the diagonal I over the entire interval $[0,1]$.

Example 3.1. Let $p = 1$ and $w(t) \equiv 2$. Then

$$\begin{aligned}\mathcal{D}_{p,w}(L_n) &= 2\int_0^1 (t - L_n(t))\, dt \\ &= G_n, \qquad (3.4)\end{aligned}$$

where G_n is the Gini coefficient

$$G_n := \frac{1}{2\bar{X}n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|X_i - X_j|. \tag{3.5}$$

The Gini coefficient G_n has played a central role in measuring economic inequality since its introduction by Corrado Gini at the beginning of the 20th century. For historical and bibliographical notes on the subject we refer to Giorgi (1990, 1993).

EXAMPLE 3.2. Let $p > 0$ and $w(t) \equiv 1$. Then the following equality

$$\mathcal{D}_{p,w}(L_n) = G_{n,p}$$

holds, where $G_{n,p}$ is the E-Gini index

$$G_{n,p} := 2\left(\int_0^1 (t - L_n(t))^p \, dt\right)^{1/p} \tag{3.6}$$

of Chakravarty (1988). Note that when $p = 1$, then $G_{n,p}$ is the classical Gini coefficient G_n. It should also be noted that Chakravarty (1988) actually introduced the following general index: $2\phi^{-1}(\int_0^1 \phi(t - L_n(t))dt)$, where ϕ is a strictly increasing function such that $\phi(0) = 0$. When $\phi(x) = x^p$ for some $p > 0$, then the latter index reduces to (3.6) as it is noted on on p.150 of Chakravarty (1988).

EXAMPLE 3.3. Let $\nu > 0$ be fixed. If $p = 1$ and $w(t) = \nu(\nu - 1)(1 - t)^{\nu-2}$, then

$$\begin{aligned}\mathcal{D}_{p,w}(L_n) &= \nu(\nu - 1)\int_0^1 (t - L_n(t))\,(1 - t)^{\nu-2}dt \\ &= I_{n,\nu}, \end{aligned} \tag{3.7}$$

where $I_{n,\nu}$ is the S-Gini index

$$I_{n,\nu} := 1 - \frac{1}{\bar{X}n^{\nu}}\sum_{i=1}^{n}\left((n - i + 1)^{\nu} - (n - i)^{\nu}\right)X_{i:n}.$$

Comparing (3.7) with (3.4) shows that when $\nu = 2$, then the S-Gini index $I_{n,\nu}$ is the Gini coefficient G_n. As a measure of economic inequality, $I_{n,\nu}$ can be found in Kakwani (1980), Donaldson and Weymark (1980), Weymark (1980/81). We also refer to Yitzhaki (1983) for a closely related work in the area.

EXAMPLE 3.4. Let $p = 1$. Then $\mathcal{D}_{p,w}(L_n)$ becomes

$$\int_0^1 (t - L_n(t))\,w(t)dt. \tag{3.8}$$

As a measure of economic inequality, the integral (3.8) appears in formula (3) on p.806 of Mehran (1976). The integral (3.8) also appears in Nygård and Sandström (1988, 1989) where it is called the weighted Lorenz area. If we define w by the formula

$w(t) = w_1(t) \big/ \int_0^1 s w_1(s) ds$ where $w_1 : (0,1) \to (0,\infty)$ is such that $\int_0^1 s w_1(s) ds \in (0,\infty)$, then

$$\begin{aligned} \mathcal{D}_{1,w}(L_n) &= \frac{1}{\int_0^1 s w_1(s) ds} \int_0^1 (t - L_n(t))\, w_1(t) dt \\ &=: G_n(w_1). \end{aligned}$$

The quantity $G_n(w_1)$ is known in the literature (cf. Shorrocks and Slottje, 1995, or p.142 of Sen, 1997, as a more convenient reference) as the generalized Gini index. Note that when $w_1(s) \equiv 1$, then $G_n(w_1)$ is the classical Gini coefficient G_n. If $w_1(s) = (1-t)^{\nu-2}$ for some $\nu > 1$, then $G_n(w_1)$ is the S-Gini index $I_{n,\nu}$.

4. Understanding the Gini index: "two areas" vs "one"

Comparing (3.4) with (3.1), we see that the only difference between the "econometric" measure $2\int_0^1 (t - L_n(t))\, dt$ and the "mathematical" one $\int_0^1 (t - L_n(t))\, dt$ is the constant 2 in front of the "econometric" one. In other words, the econometric measure suggests the choice $p = 1$ and $w(t) \equiv 2$ in the definition of $\mathcal{D}_{p,w}(L_n)$, whereas the mathematical suggests the simplest possible choice: $p = 1$ and $w(t) \equiv 1$. Thus the question: why do we choose to work with *two* areas instead of the seemingly natural *one* area? In order to explain the phenomenon, we proceed as follows.

Assume for the rest of this paper that

$$x_1, x_2, \ldots, x_n \geq 0. \tag{4.1}$$

In view of assumption (3.3), there is at least one (strictly) positive x_j among the x's of (4.1). Thus, the Lorenz curve L_n is well defined and lies on or below the diagonal I over the whole interval $[0,1]$. Furthermore, L_n lies on or above the curve

$$O_n(t) := \begin{cases} 0, & 0 \leq t < 1 - n^{-1}, \\ tn - (n-1), & 1 - n^{-1} \leq t \leq 1. \end{cases}$$

The just defined deterministic curve O_n appears to be also a Lorenz curve. Indeed, let all x's of (4.1), except only one, be equal to zero. Then the corresponding Lorenz curve is exactly O_n. Letting n tend to infinity, we obtain that O_n pointwise converges to the degenerate curve

$$O(t) := \begin{cases} 0, & 0 \leq t < 1, \\ 1, & t = 1. \end{cases}$$

Consequently, for any sample size n, and for any Lorenz curve L_n, the area between I and L_n never exceeds the area between I and O. This proves the following inequalities:

$$0 \equiv \mathcal{D}_1(I) \leq \mathcal{D}_1(L_n) \leq \mathcal{D}_1(O) \equiv 1/2.$$

Consequently, the "econometric" Gini coefficient

$$G_n = \mathcal{D}_1(L_n)/\mathcal{D}_1(O)$$

is the normalized "mathematical" index $\mathcal{D}_1(L_n)$.

5. Normalizing the unified index

Following the idea of previous Section 4, we now normalize the unified index $\mathcal{D}_{p,w}(L_n)$ and obtain the following general coefficient of economic inequality:

$$\mathcal{G}_{n,p}(w) := \mathcal{D}_{p,w}(L_n)/\mathcal{D}_{p,w}(O) \tag{5.1}$$

The coefficient is is well defined for any fixed $p > 0$ and for any function $w : (0,1) \to (0,\infty)$ such that

$$\int_0^1 t^p w(t)dt \in (0,\infty). \tag{5.2}$$

Note that $0 \le \mathcal{G}_{n,p}(w) \le 1$ for any Lorenz curve L_n. Furthermore, if $L_n = I$, then $\mathcal{G}_{n,p}(w) = 0$. If $L_n = O_n$, then $\mathcal{G}_{n,p}(w) \uparrow 1$ when $n \to \infty$.

We shall now consider several special cases of the "normalized unified" Gini coefficient $\mathcal{G}_{n,p}(w)$. Note how simple and natural (at least from the mathematical point of view) the choices of p and w will be in following Examples 5.1-5.3.

Example 5.1. When $p = 1$ and $w(t) \equiv 1$, then

$$\mathcal{G}_{n,p}(w) = G_n,$$

the classical Gini coefficient.

Example 5.2. When $p > 0$ and $w(t) \equiv 1$, then

$$\mathcal{G}_{n,p}(w) = \mathcal{G}^*_{n,p},$$

where

$$\mathcal{G}^*_{n,p} := (p+1)^{1/p} \left(\int_0^1 (t - L_n(t))^p \, dt \right)^{1/p}.$$

When $p = 1$, then $\mathcal{G}^*_{n,p}$ is the Gini coefficient G_n. We note that from the mathematical point of view the definition of $\mathcal{G}^*_{n,p}$ is more natural than that of the E-Gini index $G_{n,p}$. We call $\mathcal{G}^*_{n,p}$ the "normalized" E-Gini coefficient. Note that $\mathcal{G}^*_{n,p}$ is $(p+1)^{1/p}/2$ times the E-Gini index $G_{n,p}$.

Example 5.3. Let $p = 1$, and let $w : (0,1) \to (0,\infty)$ be such that $\int_0^1 tw(t)dt \in (0,\infty)$. Then

$$\mathcal{G}_{n,p}(w) = G_n(w),$$

the general Gini index of Shorrocks and Slottje (1995). Choosing the weight function $w(t) = (1-t)^{\nu-2}$ with some $\nu > 1$, we obtain

$$\mathcal{G}_{n,1}(w) = I_{n,\nu},$$

the S-Gini index.

6. Modifying the normalized index

The discussion above demonstrates that the "normalized unified" Gini coefficient $\mathcal{G}_{n,p}(w)$ is a natural measure of economic inequality covering a number of well known and widely used coefficients/indices. From the mathematical point of view, however, $\mathcal{G}_{n,p}(w)$ is still somewhat artificial. The reason for this can be explained as follows.

Any (empirical) Lorenz curve L_n lies between the two "extreme" ones: I and O_n. Therefore, $\mathcal{D}_{p,w}(I) \le \mathcal{D}_{p,w}(L_n) \le \mathcal{D}_{p,w}(O_n)$. The latter bounds suggest the following natural normalization

$$\mathcal{H}_{n,p}(w) := \mathcal{D}_{p,w}(L_n)/\mathcal{D}_{p,w}(O_n) \tag{6.1}$$

for $\mathcal{D}_{p,w}(L_n)$, instead of that in (5.1). The coefficient $\mathcal{H}_{n,p}(w)$ is well defined for any $p > 0$ and for any $w : (0,1) \to (0,\infty)$ such that

$$\int_0^1 t^p(1-t)^p w(t)dt \in (0,\infty). \tag{6.2}$$

Note that assumption (6.2) is weaker than (5.2). Consequently, $\mathcal{H}_{n,p}(w)$ is well defined for a lager class of weight functions $w : (0,1) \to (0,\infty)$ than the coefficient $\mathcal{G}_{n,p}(w)$. Moreover, since (6.2) and (3.2) are identical, the coefficient $\mathcal{H}_{n,p}(w)$ is well defined under the same assumptions as $\mathcal{D}_{p,w}(L_n)$, rendering the outmost generality of $\mathcal{H}_{n,p}(w)$.

Some basic properties of $\mathcal{H}_{n,p}(w)$ now follow. First, $0 \le \mathcal{H}_{n,p}(w) \le 1$ for any Lorenz curve L_n. Second, if $L_n = I$, then $\mathcal{H}_{n,p}(w) = 0$. Third, if $L_n = O_n$, then $\mathcal{H}_{n,p}(w) = 1$. In view of these three properties, we can now argue that $\mathcal{H}_{n,p}(w)$ is a more natural coefficient than $\mathcal{G}_{n,p}(w)$, since the latter one satisfies the third property above only asymptotically, when $n \to \infty$.

In the following three examples we shall carefully analyze the coefficient $\mathcal{H}_{n,p}(w)$ in three special cases related to, respectively, the Gini, E-Gini, and S-Gini indices.

Example 6.1. When $p = 1$ and $w(t) = 1$, then straightforward calculations show that the normalizing constant $\mathcal{D}_{p,w}(O_n)$ equals $2^{-1}(1-n^{-1})$. Thus, the equality

$$\mathcal{H}_{n,p}(w) = H_n,$$

where H_n is the "modified normalized" Gini coefficient:

$$\begin{aligned} H_n &:= \frac{n}{n-1} 2 \int_0^1 (t - L_n(t))\, dt \\ &= \frac{n}{n-1} G_n \\ &= \frac{1}{2\bar{X} n(n-1)} \sum_{i=1}^n \sum_{j=1}^n |X_i - X_j|. \end{aligned} \tag{6.3}$$

Note that if representation (3.5) connects G_n with the theory of V-statistics, then equality (6.3) connects H_n with the theory of U-statistics. It is clear, however, that from the asymptotic point of view when $n \to \infty$, both H_n and G_n are equivalent.

Example 6.2. When $p > 0$ and $w(t) = 1$, then the normalizing constant $\mathcal{D}_{p,w}(O_n)$ equals $(p+1)^{-1/p}(1 - n^{-1})$. Consequently,

$$\mathcal{H}_{n,p}(w) = H_{n,p},$$

where $H_{n,p}$ is the "modified normalized" E-Gini coefficient:

$$\begin{aligned} H_{n,p} &:= \frac{n}{n-1} (p+1)^{1/p} \left(\int_0^1 (t - L_n(t))^p\, dt \right)^{1/p} \\ &= \frac{n}{n-1} G_{n,p}. \end{aligned} \tag{6.4}$$

When $p = 1$, then $H_{n,p}$ is the "modified normalized" Gini coefficient H_n. Equation (6.4) demonstrates that $H_{n,p}$ and $G_{n,p}$ are asymptotically equivalent when $n \to \infty$.

Example 6.3. Let $p = 1$ and $w(t) = (1-t)^{\nu - 2}$ for some $\nu > 0$. Straightforward calculations prove that

$$\mathcal{D}_{p,w}(O_n) = \begin{cases} \frac{1}{\nu(1-\nu)} \left(n^{1-\nu} - 1 \right), & 0 < \nu < 1, \\ \log n, & \nu = 1, \\ \frac{1}{\nu(\nu-1)} \left(1 - \frac{1}{n^{\nu-1}} \right), & \nu > 1. \end{cases}$$

Note the completely different asymptotic behaviour of $\mathcal{D}_{p,w}(O_n)$ when $n \to \infty$ in the three cases: $0 < \nu < 1$, $\nu = 1$, and $\nu > 1$. For this reason, we shall now separately look at the coefficient $\mathcal{H}_{n,p}(w)$ in these three cases. As a by-product of this, we shall also introduce the "modified normalized" S-Gini index $J_{n,\nu}$ for any value $0 < \nu < \infty$.

Case $\nu > 1$. We have the following equality

$$\mathcal{H}_{n,p}(w) = J_{n,\nu},$$

where $J_{n,\nu}$, $\nu > 1$, is the "modified normalized" S-Gini index defined as follows:

$$\begin{aligned} J_{n,\nu} &:= \frac{n^{\nu-1}}{n^{\nu-1}-1}\nu(\nu-1)\int_0^1 (t - L_n(t))\,(1-t)^{\nu-2}dt \\ &= \frac{n^{\nu-1}}{n^{\nu-1}-1} I_{n,\nu} \qquad (6.5)\\ &= \frac{n^{\nu-1}}{n^{\nu-1}-1}\left(1 - \frac{1}{\bar{X}n^{\nu}}\sum_{i=1}^{n}\left((n-i+1)^{\nu} - (n-i)^{\nu}\right)X_{i:n}\right). \end{aligned}$$

Equation (6.5) shows that, when $\nu > 1$, then the two S-Gini indices $J_{n,\nu}$ and $I_{n,\nu}$ are asymptotically equivalent when $n \to \infty$.

Case $\nu = 1$. We have the following equality

$$\mathcal{H}_{n,p}(w) = J_{n,1},$$

where $J_{n,1}$ is the "modified normalized" S-Gini index defined as follows:

$$\begin{aligned} J_{n,1} &:= \frac{1}{\log n}\int_0^1 (t - L_n(t))\,(1-t)^{-1}dt \\ &= 1 - \frac{1}{\bar{X}n\log n}\sum_{i=1}^{n}\left((n-i+1)\log(n-i+1) - (n-i)\log(n-i)\right)X_{i:n}. \end{aligned}$$

Note that the corresponding S-Gini index $I_{n,1}$ is identically 0.

Case $0 < \nu < 1$. We have the following equality

$$\mathcal{H}_{n,p}(w) = J_{n,\nu},$$

where $J_{n,\nu}$, $0 < \nu < 1$, is the "modified normalized" S-Gini index defined as follows:

$$\begin{aligned} J_{n,\nu} &:= \frac{1}{n^{1-\nu}-1}\nu(1-\nu)\int_0^1 (t - L_n(t))\,(1-t)^{\nu-2}dt \\ &= -\frac{1}{n^{1-\nu}-1} I_{n,\nu} \qquad (6.6)\\ &= -\frac{1}{n^{1-\nu}-1}\left(1 - \frac{1}{\bar{X}n^{\nu}}\sum_{i=1}^{n}\left((n-i+1)^{\nu} - (n-i)^{\nu}\right)X_{i:n}\right). \end{aligned}$$

From equation (6.6) we see that when $0 < \nu < 1$, then $J_{n,\nu}$ and $I_{n,\nu}$ have different asymptotic behaviour when $n \to \infty$.

7. Open problems

Let F denote the cumulative distribution function (cdf) F of income X in the society under consideration. Under some assumptions on F, p, and w, we can demonstrate that both coefficients $\mathcal{G}_{n,p}(w)$ and $\mathcal{H}_{n,p}(w)$ converge to the (theoretical) "normalized unified" Gini coefficient

$$\mathcal{G}_{\infty,p}(w) := \mathcal{D}_{p,w}(L_F)/\mathcal{D}_{p,w}(O). \tag{7.1}$$

In the definition above, L_F denotes the (theoretical) Lorenz curve defined by the fomula (cf. Gastwirth, 1971):

$$L_F(t) := \frac{1}{\mu}\int_0^t F^{-1}(s)ds,$$

where F^{-1} is the quantile function $F^{-1}(t) := \inf\{x : \ F(x) \geq t\}$ and $\mu := \mathbf{E}(X)$ is the mean of X. We assume throughout that μ is finite, non-zero, and positive.

Choosing various p and w in formula (7.1), we arrive at the theoretical counterparts to the empirical indices of economic inequality discussed in previous sections. For example, when $p = 1$, then $\mathcal{G}_{\infty,p}(w)$ is the generalized Gini index

$$G_\infty(w) := \frac{1}{\int_0^1 sw(s)ds}\int_0^1 (t - L_F(t))\, w(t)dt \tag{7.2}$$

of Shorrocks and Slottje (1995). For a convenient reference concerning $G_\infty(w)$, we refer to p.142 of Sen (1997).

With the notations above, we are now in the position to discuss some open problems concerning the large sample asymptotic behaviour of $\mathcal{G}_{n,p}(w)$ and $\mathcal{H}_{n,p}(w)$. We start with a problem concerning weak and strong consistencies of the two coefficients.

Open problem 7.1. Under what (minimal) assumptions on F, p, and w, do the two coefficients $\mathcal{G}_{n,p}(w)$ and $\mathcal{H}_{n,p}(w)$ converge to $\mathcal{G}_{\infty,p}(w)$ in probability, or almost surely, when $n \to \infty$?

Assuming that a solution to open problem 7.1 has been obtained, we are then interested in the following problem.

Open problem 7.2. What is the asymptotic distribution of the appropriately normalized differences $\mathcal{G}_{n,p}(w) - \mathcal{G}_{\infty,p}(w)$ and $\mathcal{H}_{n,p}(w) - \mathcal{G}_{\infty,p}(w)$? Under what minimal assumptions does the asymptotic distribution hold?

Given the close relationship of the two coefficients $\mathcal{G}_{n,p}(w)$ and $\mathcal{H}_{n,p}(w)$ to a number of well known indices of economic inequality, it is natural to expect that open problems 7.1 and 7.2 have known partial solutions. Indeed, asymptotic consistency and normality of, for example, the classical Gini coefficient G_n can be found in Hoeffding (1948). Asymptotic results concerning the S- and E-Gini indices can also be found in numerous econometric and statistical papers. For references and recent results on the topic, we refer, for example, to Barrett and Donald (2000), Zitikis and Gastwirth (2002), Zitikis (2000, 2002). We note in this regard that Barrett and Donald (2000) employ empirical and quantile processes point of view and obtain desired asymptotic results for a large class of indices, including the S- and E-Gini indices. The papers by Zitikis and Gastwirth (2002) and Zitikis (2000, 2002) aim at asymptotic results for indices under minimal assumptions on the cdf F. In Zitikis and Gastwirth (2002) and Zitikis (2002) we argue, for example, that the theory of L-statistics is a most natural tool for investigating the S-Gini index, whereas in Zitikis (2002) we suggest using the so-called general Vervaat process V_n (cf., e.g., Zitikis, 1998) when investigating the E-Gini index. In this regard we believe that the Vervaat process V_n will also appear as a most appropriate and powerful tool for deriving desired asymptotic results for $\mathcal{G}_{n,p}(w)$ and $\mathcal{H}_{n,p}(w)$ under minimal assumptions on F, p, and w.

Open problem 7.3. Let w satisfy (6.2) but fail to satisfy (5.2). What is the asymptotic behaviour – in probability or almost surely – of the appropriately centered and normalized coefficient $\mathcal{H}_{n,p}(w)$ when $n \to \infty$?

8. Concluding remarks and summary

In this paper we demonstrate that a number of indices of economic inequality are special and natural cases of the "normalized unified" Gini coefficient $\mathcal{G}_{n,p}(w)$ defined in (5.1). The choices of the parameter $p > 0$ and the weight function $w : (0,1) \to (0,\infty)$ leading to well known indices of economic inequality are also natural from the mathematical point of view. Furthermore, in this paper we propose the "modified normalized" Gini coefficient $\mathcal{H}_{n,p}(w)$ defined in (6.1). The coefficient $\mathcal{H}_{n,p}(w)$ satisfies the following three properties: 1) it is always in the interval $[0,1]$, 2) equals 0 in the "egalitarian" case, and 3) equals 1 in the case of "extreme inequality."

We conclude this section, and the paper as well, with the note that the present research is a result of a careful mathematical analysis of well known indices of economic inequality. We have already mentioned a number of papers directly related to the subject. Now we shall mention a few books. Various econometric aspects of constructing measures of economic inequality can be found, for example, in the monographs by Amiel and Cowell (1999), Champernowne and Cowell (1998), Kakwani (1980a), Nygård and Sandström (1981), Sen (1997), as well as in the handbook by Silber (1999). Various statistical and probabilistic tools for analyzing such measures of economic inequality – especially from the asymptotic point of view – can be found, for example, in Csörgő (1983), Csörgő, Csörgő, and Horváth (1986), Helmers (1982), Serfling (1980), Shorack (2000), Shorack and Wellner (1986),

References

Amiel, Y. and Cowell, F.A. (1999). *Thinking About Inequality.* Cambridge University Press, Cambridge.

Barrett, G.F. and Donald, S.G. (2000). Statistical inference with generalized Gini indices of inequality and poverty. *Discussion Paper* 2002/01. School of Economics. University of New South Wales. This paper can be downloaded at: http://economics.web.unsw.edu.au/people/gbarrett/papers/ginitext.pdf

Chakravarty, S. R. (1988). Extended Gini indices of inequality. *International Economic Review,* 29, 147–156.

Champernowne, D. G. and Cowell, F.A. (1998). *Economic Inequality and Income Distribution.* Cambridge University Press, Cambridge.

Csörgő, M. (1983). *Quantile Processes with Statistical Applications.* SIAM, Philadelphia.

Csörgő, M., Csörgő, S. and Horváth, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes.* Springer-Verlag, Berlin.

Donaldson, D. and Weymark, J. A. (1980). A single-parameter generalization of the Gini indices of inequality. *Journal of Economic Theory,* 22, 67–86.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica,* 39, 1037-1039.

Giorgi, G.-M. (1990). Bibliographic portrait of the Gini concentration ratio. *Metron,* 48, 183-221.

Giorgi, G.-M. (1993). A fresh look at the topical interest of the Gini concentration ratio. *Metron,* 51, 83-98.

Helmers, R. (1982). *Edgeworth Expansions for Linear Combinations of Order Statistics.* Mathematisch Centrum, Amsterdam.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statisitcs,* 19, 293–325.

Kakwani, N. (1980). On a class of poverty measures. *Econometrica,* 48, 437-446.

Kakwani, N. C. (1980a). *Income Inequality and Poverty.* Oxford University Press, New York.

Lorenz, M.O. (1905). Methods of measuring the concentration of wealth. *Journal of the American Statistical Association,* 9, 209-219.

Mehran, F. (1976). Linear measures of income inequality. *Econometrica,* 44, 805-809.

Nygård, F. and Sandström, A. (1981). *Measuring Income Inequality.* Almqvist & Wiksell, Stockholm.

Nygård, F. and Sandström, A. (1988). The weighted mean difference. *Metron,* 46, 21-31.

NYGÅRD, F. AND SANDSTRÖM, A. (1989). Income inequality measures based on sample surveys. *Journal of Econometrics,* 42, 81-95.

SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley and Sons, New York.

SEN, A. (1997). *On Economic Inequality.* (Expanded edition with a substantial annexe by J. E. Foster and A. Sen.) Clarendon Press, Oxford.

SHORACK, G. R. (2000). *Probability for Statisticians.* Springer-Verlag, New York.

SHORACK, G. R. AND WELLNER, J.A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley and Sons, New York.

SHORROCKS, A. F. AND SLOTTJE, D. J. (1995). Approximating unanimity orderings: An application to Lorenz dominance. *Discussion Paper.* University of Essex.

SILBER, J. (1999). *Handbook on Income Inequality Measurement.* Kluwer Academic Publishers, Boston.

WEYMARK, J. A. (1980/81). Generalized Gini inequality indices. *Mathematical Social Sciences,* 1, 409-430.

YITZHAKI, S. (1983). On an extension of the Gini inequality index. *International Economic Review,* 24, 617–628.

ZITIKIS, R. (1998). The Vervaat process. In: *Asymptotic Methods in Probability and Statistics – A Volume in Honour of Miklós Csörgő* (ed. B. Szyszkowicz), 667–694. Elsevier, North-Holland, Amsterdam.

ZITIKIS, R. (2000). Asymptotic estimation of the E-Gini index. (Submitted for publication.)

ZITIKIS, R. (2002). Large sample estimation of a family of economic inequality indices. To appear in: *Pakistan Journal of Statistics.* (A special issue honoring Dr. S. Ejaz Ahmed's contributions to the development of statistics.)

ZITIKIS, R. AND GASTWIRTH, J.L. (2002). Asymptotic distribution of the S-Gini index. To appear in: *Australian and New Zealand Journal of Statistics.*