On a new class of continuous indices of inequality

Citation for published version:
Chan, T 2022, 'On a new class of continuous indices of inequality', *Mathematical Social Sciences*, vol. 120, pp. 8-23. https://doi.org/10.1016/j.mathsocsci.2022.08.003

Digital Object Identifier (DOI):
10.1016/j.mathsocsci.2022.08.003

Published In:
Mathematical Social Sciences

Mathematical Social Sciences 120 (2022) 8–23

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/mss

On a new class of continuous indices of inequality

Terence Chan *

School of Mathematical & Computer Sciences, Heriot-Watt University, Edinburgh, UK
Maxwell Institute for Mathematical Sciences, Edinburgh, UK

A R T I C L E   I N F O

Article history:
Received 2 December 2021
Received in revised form 18 July 2022
Accepted 23 August 2022
Available online 29 August 2022

Keywords:
Gini index
Measures of inequality

A B S T R A C T

The Gini index is a well-known and long-established measure of inequality for distributions of income and other quantities. However, it has a number of mathematical disadvantages. Firstly, it is discontinuous with respect to all the main modes of convergence of probability measures. Secondly, it relies critically on the finiteness of the mean of the underlying distribution. Finally, even when the underlying distribution has a finite mean, estimation of the Gini index from data can be problematic if the variance of the underlying distribution is infinite. In this paper, we propose a class of inequality indices which are continuous with respect to setwise convergence of probability measures (and hence also with respect to convergence in total variation) and which do not require the underlying distribution to possess any finite moments whatsoever. Moreover, our class of inequality indices can be easily estimated from data and the standard methods of statistical inference can be applied to the estimators.

1. Introduction

The Gini index is a well-known and long-established measure of inequality for distributions of various quantities. It was initially used in studies of income inequality but it can be applied to any general probability distribution, for example distributions of life expectancy, rates of literacy, educational attainment, measures of well-being and many other quantities of interest in the social sciences. However, it has a number of mathematical disadvantages. Firstly, it is discontinuous with respect to all the main modes of convergence of probability measures, so statements along the lines of "similar distributions have similar Gini indices" are not necessarily true. Secondly, it relies critically on the finiteness of the mean of the underlying distribution. Finally, even when the underlying distribution has a finite mean, estimation of the Gini index from data can be problematic if the variance of the underlying distribution is infinite (see, for example, Fontanari et al. (2018) and the references therein). The authors of this last paper propose a parametric method of estimation resulting in an estimator which is asymptotically normal. However, this is at the expense of assuming that the underlying data come from a parametric family of fat-tailed distributions. In this paper, we propose a class of inequality indices which are continuous with respect to setwise convergence of probability measures (and hence also with respect to convergence in total variation) and which do not require the underlying distribution to possess any finite moments whatsoever. Moreover, our class of indices can be estimated from data using the usual standard statistical methods because, unlike the Gini index, they are asymptotically normal even for fat-tailed distributions. The key difference between the standard Gini index and our proposed family of indices is that whereas the Gini index is based on absolute differences of the values of the relevant quantity, our indices are based on the ratio of those values. For this reason, the standard Gini index is not a special case of our family of indices.

Let P denote a probability measure on $[0, \infty)$. (The Gini index can also be defined on distributions on the whole real line, but its interpretation may be problematic in certain situations – we shall explain this in more detail later in this section – which is why we restrict ourselves to distributions supported on $[0, \infty)$.) Let $F(x) = P([0, x])$ denote the associated distribution function. In the sequel, we shall sometimes use P and F interchangeably whenever no ambiguity can arise. We begin with a reminder of the definition of the standard Gini index.

Definition 1.1. The Gini index G_0 of inequality for a probability measure P is defined to be

$$G_0(P) = G_0(F) = \frac{1}{2\mu} \int_0^\infty \int_0^\infty |x - y| \, dF(x) \, dF(y),$$

where $\mu = \int_0^\infty x \, dF(x)$ is the mean of the underlying distribution.

If μ is infinite the Gini index cannot be defined.

* Correspondence to: School of Mathematical & Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK.
E-mail address: t.chan@hw.ac.uk.

https://doi.org/10.1016/j.mathsocsci.2022.08.003
0165-4896/

Alternatively, the Gini index may be expressed in terms of the Lorenz curve

$$L(\alpha) = \frac{1}{\mu}\int_0^\alpha q(x)\,dx = \frac{1}{\mu}\int_0^{q(\alpha)} x\,dF(x), \quad 0 \le \alpha \le 1,$$

where $q(x) = \inf\{y : F(y) \ge x\}$ is the quantile function. (If F is invertible then $q(x) = F^{-1}(x)$.) In the case of a discrete distribution with order statistics $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ and associated probabilities $p_1, p_2, \ldots, p_n$, the Lorenz curve is the piecewise linear function joining the points (F_i, L_i), where

$$F_i = \sum_{j=1}^{i} F(x_{(i)}) \quad \text{and} \quad L_i = \frac{\sum_{j=1}^{i} x_{(i)} F(x_{(i)})}{\mu}.$$

The Lorenz curve gives the cumulative proportion of the relevant quantity that is possessed by the α-percentile of the population; for example, in the case of an income distribution, if $L(0.1) = 0.2$, this means that the lowest-earning 10% of the population collectively earn 20% of the total income earned by the whole population.

The Gini index G_0 may then be expressed as

$$G_0 = 2\int_0^1 (x - L(x))\,dx.$$

(The Lorenz curve $L_e(x) = x$ corresponds to the case of perfect equality where every individual possesses the same amount, so the Gini index is twice the difference in L^1 norm between the Lorenz curve of perfect equality and that of the actual distribution under consideration.)

Observe that if $P((-\infty, 0)) > 0$, there is always in interval $(0, \delta)$ for some $\delta > 0$ on which $L < 0$. Furthermore, if $P((-\infty, 0))$ is sufficiently large, $L < -1$ on quite a large interval and the Gini index may then be large than 1. For example, take a normal distribution $N(\epsilon, 1)$ for small $\epsilon > 0$. In this case,

$$L\left(\frac{1}{2}\right) = \frac{1}{2} - \frac{1}{\epsilon\sqrt{2\pi}},$$

so the Gini index will be considerably larger than 1. This is why we shall only consider distributions supported on $[0, \infty)$.

Before presenting a counter-example showing the discontinuity of the Gini index, we first remind the reader of the definitions of the 3 main modes of convergence of probability measures. The definitions below are in order of weakness of the topology, with the weakest first.

Definition 1.2.

1. A sequence of probability measures P_n is said to *converge weakly* to a probability measure P as $n \to \infty$ if $\int \phi\,dP_n \to \int \phi\,dP$ for all bounded and continuous functions ϕ.
2. A sequence of probability measures P_n is said to *converge setwise* to P if $P_n(A) \to P(A)$ for all measurable sets A.
3. A sequence of probability measures P_n is said to *converge in total variation* to P if $\|P_n - P\|_{\text{tv}} \to 0$, where

$$\|P - Q\|_{\text{tv}} = \sup_{|\phi| \le 1}\left\{\left|\int \phi\,dP - \int \phi\,dQ\right|\right\} = 2\sup_A |P(A) - Q(A)|$$

is the total variation distance between probability measures P and Q. (The first supremum above is taken over all measurable functions ϕ with values in $[-1, 1]$ but which are not necessarily continuous.)

Some authors (for example Jacka and Roberts (1997)) refer to setwise convergence as "strong convergence" but we shall avoid this terminology here because convergence in total variation is even stronger. Clearly, convergence in total variation implies setwise convergence. Lemma 1.1 shows that setwise convergence in turn implies weak convergence. The standard notation for weak convergence is $P_n \Rightarrow P$. In the sequel, we use the notation $P_n \to_{\text{sw}} P$ and $P_n \to_{\text{tv}} P$ to denote respectively setwise and total variation convergence.

We now give a counter-example which shows that the Gini index is discontinuous with respect to all the modes of convergence defined above. Let

$$P_n = \frac{n-1}{n}\delta_1 + \frac{1}{n}\delta_{x_n}, \tag{1.1}$$

where $x_n \to \infty$ as $n \to \infty$ and δ_a denotes unit mass at a. Such a distribution represents a situation where, in a population of n individuals, $n-1$ individuals each owns 1 unit of some quantity while the remaining 1 individual owns x_n units. Then $P_n \to_{\text{tv}} \delta_1$ and hence also converges to δ_1 setwise and weakly. However,

$$G_0(P_n) = \frac{2\frac{n-1}{n}\frac{1}{n}(x_n - 1)}{2(1 - 1/n + x_n/n)} \approx \frac{\frac{x_n}{n}}{1 + \frac{x_n}{n}} \to \begin{cases} 1 & \text{if } \frac{x_n}{n} \to \infty \\ \frac{c}{1+c} & \text{if } \frac{x_n}{n} \to c \\ 0 & \text{if } \frac{x_n}{n} \to 0 \end{cases}$$

whereas $G_0(\delta_1) = 0$. Moreover, if x_n/n oscillates, say

$$\liminf \frac{x_n}{n} = 0 \quad \limsup \frac{x_n}{n} = \infty,$$

(this can happen even if x_n is monotonically increasing) then G_0 will also oscillate:

$$\liminf G_0(P_n) = 0 \quad \limsup G_0(P_n) = 1.$$

Also, the Lorenz curve associated with P_n is the piecewise linear curve joining the points $(0, 0)$, $([n-1]/n,\ [n-1]/[n(1 - 1/n + x_n/n)])$ and $(1, 1)$, which does not always converge to the straight line $L_e(x) = x$. Therefore, any extension of the Gini index based on the Lorenz curve will also be discontinuous. For example, Chakravarty (1988) proposed the following generalization:

$$\tilde{G}_0 = 2\phi^{-1}\left(\int_0^1 \phi(x - L(x))\,dx\right),$$

where ϕ is a strictly increasing function.

The fact that the Gini index is discontinuous is not necessarily a flaw; after all, many fundamental parameters of distributions, such as the mean and variance, are also discontinuous. In fact, in some situations it may even be a positive attribute. The various topologies of convergence each captures some aspect in which two distributions may be different or similar to each other but no topology can possibly capture every aspect of such differences or similarities, therefore it may well be the case that two superficially similar distributions may nevertheless hide important differences in terms of their degree of inequality which the Gini index is able to detect. Consider the example (1.1). One may argue that if the vast majority of the population all earn the same amount, there is a high degree of equality despite the presence of a single multi-billionaire (or a tiny number of multi-billionaires) with ever increasing wealth. On the other hand, one can also argue that, if a single individual earns potentially more than the rest of the population combined, that is a sign of an extremely unequal distribution. Ultimately, which of these points of view is to be preferred depends on what the value 1 represents. If it represents a decent level of income which affords a decent standard of living, then the first point of view has considerable merit. If, on the other hand, it represents a level of income below the poverty line, then the second interpretation is more credible. Given these competing points of view, it is perhaps not entirely unreasonable that the Gini index converges to vastly different values depending on the rate at which the ultra-rich increase their income. However, while different people will have different opinions as to the real level of inequality present in distributions such as (1.1), it would be somewhat unsatisfactory to vacillate in one's opinion between the extremes of nearly perfect equality and extreme inequality for the same sequence of distributions, as the Gini index can do. One may not agree that the distribution P_n is approaching perfect equality in reality but at least a continuous index behaves consistently for such a sequence of distributions.

Despite being discontinuous in the strict technical sense, the Gini index (and many others) have partial continuity properties. While it is true that for every distribution F, one can always find a sequence of distributions F_n such that $F_n \to F$ in whatever sense but $G_0(F_n) \not\to G_o(F)$ (so G_0 is nowhere continuous), it is also the case that for every F, there exists a sequence F_n such that $F_n \to F$ and $G_0(F_n) \to G_0(F)$. This is precisely the point made in Fields (1993), which established the following partial continuity result for the indices I considered in that paper: namely that for certain forms of F_n and F, $F_n \to F$ implies that $I(F_n) \to I(F)$. It may be argued – as Fields (1993) appears to implicitly – that such special cases of F_n and F are more natural or important for applications while examples such as (1.1) are somewhat artificial. Even so, there are many situations where one might wish to make statements along the lines of "similar distributions should have similar inequality indices". For example, a policy-maker may try to defend a certain policy by saying that it changes the income distribution only slightly and so any consequent increase in inequality would only be small, which would be outweighed by the other supposed benefits of the policy. Such a statement may not be justifiable even in the case of continuous indices unless one can quantify their modulus of continuity (see Theorem 2.3). In the case of a discontinuous index, there would be even greater cause for scepticism. It is therefore interesting to ask if it is possible to construct an index of inequality which is continuous.

Not withstanding the aforegoing rather extensive discussion of continuity, it is worth pointing out that continuity is by no means the most important or attractive advantage of the family of indices proposed here. Of equal, or perhaps even greater importance, is the fact that these indices are readily applicable to heavy-tailed or fat-tailed distributions.[1] Fontanari et al. (2018) have already provided some motivation for considering fat-tailed distributions in the context of income inequality. We can give another possible motivation. Heavy-tailed distributions have a long history of being used in the insurance industry to model insurance losses due to major natural disasters and other extreme events (see for example Embrechts et al. (1996)). This kind of modelling has assumed ever greater prominence and importance in recent years as a result of climate change. However, insurance companies are not the only entities which can suffer extreme losses due to climate change; on a macroeconomic level, entire countries or communities will also suffer extreme economic loss and such losses will also have heavy-tailed distributions for the same reasons. Therefore, if one wishes to investigate inequalities in the economic and social damage suffered by different communities as a result of climate change, an inequality measure which can handle all manner of fat-tailed distributions may well prove to be useful.

We end this section with some further results on convergence of probability measures which will be needed later.

The following lemma is part of a well-known result sometimes referred to as the Portmanteau theorem — see Billingsley (1968) and Jacka and Roberts (1997).

Lemma 1.1.

1. *A sequence of probability measures P_n converges weakly to P if and only if*

   $F_n(x) \to F(x)$

   for all x at which F is continuous, where F_n and F are respectively the distribution functions of P_n and P.
2. *A sequence of probability measures P_n converges setwise to P if and only if*

   $$\int \phi \, dP_n \to \int \phi \, dP$$

   for all bounded measurable functions ϕ.

Lemma 1.2. *If $P_n \to_{sw} P$ then $F_n(x) \to F(x)$ and $F_n(x) - F_n(x-) \to F(x) - F(x-)$ for all x.*

Proof. Since $P_n(A) \to P(A)$ for every (measurable) A, the result follows by taking $A = [0, x]$ and $A = \{x\}$ respectively. □

Unfortunately, the converse of Lemma 1.2 does not hold — see Example 2.4 of Feinberg et al. (2014).

The following characterization of continuous linear functions of probability measures is largely well-known but perhaps less so when stated in this form. Nevertheless, it turns out to be easier to give a complete proof here than to find a reference to it in the literature. The proof is given in Appendix A.

Lemma 1.3. *Let P be a probability measure on $\mathbb{R}^d$ and let*

$$\Phi(P) = \int_{\mathbb{R}^d} \phi(x) \, dP(x)$$

for some real-valued measurable function $\phi(x)$ on $\mathbb{R}^d$. Then

1. *Φ is continuous with respect to setwise convergence of probability measures if and only if ϕ is bounded.*
2. *Φ is continuous in total variation distance if and only if ϕ is bounded.*
3. *Φ is continuous with respect to weak convergence of probability measures if and only if ϕ is bounded and continuous.*

2. A new class of inequality indices

Consider inequality indices of the following form:

$$\frac{1}{\Psi(F)} \iint H(x, y) \, dF(x) \, dF(y), \tag{2.2}$$

where $\Psi(F)$ is some arbitrary functional of the distribution F which acts as a normalizing factor so that the index has certain other properties to be discussed later. Apart from the Gini index, many other inequality indices have this form. For example, the variance of logarithms has

$$H(x, y) = |\log x - \log y|^2.$$

The generalized entropy index has

$$H(x, y) = x^\alpha - \text{const.}, \quad \forall y, \ \alpha \neq 0, 1.$$

The extended Gini indices considered in Ebert (2010) have

$$H(x, y) = |x - y|^\alpha, \quad \alpha > 0.$$

However, Lemma 1.3 says that H must be at least a bounded function if such an index is to be continuous with respect to any mode of convergence of probability measures. Another essential property which any sensible inequality index must have is that of scale-invariance; that is, changing the units in which a quantity is measured (e.g. from dollars to euros) cannot change the value of the index. Specifically, let X be a random variable representing

[1] In common with much of the literature, we use the terms "heavy-tailed" and "fat-tailed" interchangeably here. However, in some of the mathematical literature, "heavy-tailed" has a specific technical definition whereas "fat-tailed" tends to be a more generic term used to refer to any distribution which does not have finite lower order (say order 2) moments.

the quantity of interest and suppose $X = cU$. If $F(x)$ is the distribution function of X, then $F_c(x) = F(cx)$ is the distribution function of U. By making the change of variables $x = cu$, $y = cv$ in the integral in (2.2), we see that scale-invariance means

$$\frac{1}{\Psi(F)}\iint H(cu, cv)\,dF_c(u)\,dF_c(v) = \frac{1}{\Psi(F_c)}\iint H(u, v)\,dF_c(u)\,dF_c(v). \tag{2.3}$$

Since this must hold for *all* distributions F, we must have

$$\frac{H(cu, cv)}{\Psi(F)} = \frac{H(u, v)}{\Psi(F_c)}. \tag{2.4}$$

Again, this must hold for *all* F, so

$$H(cu, cv) = \psi(c)H(u, v) \quad \text{and} \quad \Psi(F) = \psi(c)\Psi(F_c) \tag{2.5}$$

for some function $\psi(c)$ which must be bounded because H is bounded. Putting $c = ab$ and iterating (2.5) shows that $\psi(ab) = \psi(a)\psi(b)$. This implies that $\psi(c) = c^p$ for some $p \in \mathbb{R}$, $\psi(c) = 0$ or $\psi(c) = 1$. (We are only concerned with the case that $c > 0$ and not interested in what happens if $c = 0$.) Of these possibilities, $\psi(c) = 0$ is clearly impossible while $\psi(c) = c^p$ is unbounded for $p \neq 0$. Hence we must have $\psi(c) = 1$. But this means $H(cu, cv) = H(u, v)$, so $H(x, y) = H(x/y, 1)$ is a function only of the ratio x/y.

To the properties of continuity and scale-invariance can be added some other natural ones. Firstly, an index I should satisfy $I(F) = 0$ if and only if $F = \delta_a$ for some $a > 0$ (the case of perfect equality). Secondly, larger differences in income should contribute larger components to the index, so $H(x/y, 1)$ should be a strictly increasing function of x/y when $x/y > 1$ and strictly decreasing when $x/y < 1$. Finally, since we are only interested in comparing different levels of income and not the order in which the incomes appear in the index, we should have $H(x, y) = H(y, x)$.

Thus, the only natural indices of the form (2.2) which are continuous and scale-invariant are of the following form. Let h be a non-negative continuous bounded function satisfying the following:

1. $h(1) = 0$,
2. $h(1/x) = h(x)$,
3. h is strictly increasing on $[1, \infty)$ and $h(x) \uparrow 1$ as $x \to \infty$.

Properties (2) and (3) together with the continuity of h imply that $h(0) = 1$. Examples of such functions include

$$1 - e^{-\alpha|\log x|^\beta} \quad \text{and} \quad \frac{|\log x|^\beta}{\alpha + |\log x|^\beta} \quad \alpha, \beta > 0.$$

Next, define a function $H(x, y)$ for $x, y, \geq 0$ as follows:

$$H(x, y) = \begin{cases} h\left(\frac{x}{y}\right) & y > 0 \\ 1 & x > 0, y = 0 \\ 0 & x = y = 0 \end{cases} \tag{2.6}$$

The function $H(x, y)$ is essentially the same as $h(x/y)$ but we need to define it in this way in order to handle distributions which have an atom at 0 in a mathematically rigorous way. Note that H is continuous everywhere in the non-negative quadrant of $\mathbb{R}^2$ except at $(0, 0)$. This discontinuity cannot be removed because $H = 1$ along each axis but $H = 0$ along the diagonal $y = x$.

We propose the following measure G_* as an index of inequality:

$$G_*(H, F) = \frac{\int_0^\infty \int_0^\infty H(x, y)\,dF(x)\,dF(y)}{1 - F(0)^2}. \tag{2.7}$$

Before proceeding further, we should mention some other inequality indices which are not of the form (2.2) but are at least based on that form. These include the Atkinson index (see Atkinson (1970)) which has the form

$$1 - \frac{1}{\mu_F}\left(\int x^\alpha\,dF(x)\right)^{1/\alpha},$$

the Theil index which has the form

$$\frac{1}{\mu_F}\int x\log x\,dF(x) - \log\mu_F,$$

and the log deviation which as the form

$$\log\mu_F - \int \log x\,dF(x)$$

where $\mu_F = \int x\,dF(x)$ is the mean of F. These indices all suffer from the same disadvantages as the Gini index, namely discontinuity and the requirement that F has finite mean. Indeed, we can now summarize the results so far in the following Proposition, which is a direct consequence of Lemma 1.1:

Proposition 2.1. *Consider an inequality index I of the form*

$$I(F) = \frac{1}{\Psi(F)}\Phi\left(\iint H(x, y)\,dF(x)\,dF(y)\right)$$

where $\Psi(F)$ is some functional of the distribution F and Φ is an arbitrary continuous function. Then if H is an unbounded function, the index I is discontinuous with respect to any mode of convergence of probability measures.

A more intriguing condition for discontinuity is Theorem 3.1, which says that discontinuity is a consequence of the Pigou–Dalton transfer principle.

The denominator in (2.7) looks rather odd at first sight. If we simply want to require that $0 \leq G_* \leq 1$, the obvious choice of denominator would simply be 1. However, for an inequality index, we also want $G_*(P_n) \to 1$, where

$$P_n = \frac{n-1}{n}\delta_0 + \frac{1}{n}\delta_1 \tag{2.8}$$

is the archetypal example of increasingly extreme inequality when one individual has everything and everyone else has nothing. In order to achieve this, the denominator cannot include a contribution from any possible atom at 0 and it is the behaviour of the numerator in (2.7) at P_n that provides a clue as to what the appropriately denominator should be. The proof of Proposition 2.2 will provide further clarification.

Lemmas 1.1 and 1.2 imply that, for fixed H, $G_*(P)$ is continuous with respect to setwise convergence and hence also continuous with respect to convergence in total variation but not with respect to weak convergence, at all probability measures P except δ_0 where it is undefined.

We next establish the basic properties of G_*.

Proposition 2.2. *For all functions H of the form (2.6) and all distributions F,*

$$0 \leq G_*(H, F) \leq 1.$$

Moreover, $G_*(F) = 0$ if and only if $F = \delta_a$ for some $a > 0$ and for P_n as in (2.8), $G_*(P_n) \to 1$ for all H.

Proof. The lower bound is obvious, as is the fact that $G_*(F) = 0$ if and only if $F = \delta_a$ for some $a > 0$.

For the proof of the upper bound, we use the notation $\int_{0+} \ldots dF$ to mean that any atom of F at 0 is excluded from the integral. Consider the numerator in (2.7).

$$\begin{aligned}
&\int_0^\infty \int_0^\infty H(x,y)\,dF(y)\,dF(x)\\
&= \int_{0+}^\infty \int_0^\infty H(x,y)\,dF(y)\,dF(x) + F(0)\int_0^\infty H(0,y)\,dF(y)\\
&= \int_{0+}^\infty \int_0^\infty H(x,y)\,dF(y)\,dF(x) + F(0)\left(\int_{0+}^\infty H(0,y)\,dF(y) + 0\right)\\
&= \int_{0+}^\infty \int_0^\infty H(x,y)\,dF(y)\,dF(x) + F(0)(1-F(0))\\
&\le \int_{0+}^\infty 1\,dF(x) + F(0)(1-F(0))\\
&= 1 - F(0) + F(0)(1-F(0)) = 1 - F(0)^2.
\end{aligned}$$

This establishes the upper bound. Finally, for P_n in (2.8),

$$G_*(P_n) = \frac{2(n-1)}{n^2(1-(n-1)^2/n^2)} = \frac{2n-2}{2n-1} \to 1. \quad \square$$

It has already been noted that G_* is continuous with respect to setwise convergence, and hence also continuous in total variation.[2] The following result gives a modulus of continuity for G_*.

Theorem 2.3. *Let F_1 and F_2 be distribution functions. Then for fixed H,*

$$|G_*(F_1) - G_*(F_2)| \le \frac{2\|F_1 - F_2\|_{\mathrm{tv}}}{(1-F_1(0)^2)(1-F_2(0)^2)}.$$

Remark. The topology of setwise convergence is not metrizable so it is more difficult to establish a modulus of continuity for this topology. For this reason, we only give a modulus of continuity in terms of the total variation distance.

Proof. For $k = 1, 2$, let

$$I_k = \int_0^\infty \int_0^\infty H(x,y)\,dF_k(x)\,dF_k(y)$$
$$M_k = 1 - F_k(0)^2.$$

First, note that if $0 \le \phi \le 1$ is a non-negative function bounded by 1, then

$$\left| \int \phi\,dF_1 - \int \phi\,dF_2 \right| \le \frac{\|F_1 - F_2\|_{\mathrm{tv}}}{2}.$$

Firstly,

$$\begin{aligned}
&|I_1 - I_2|\\
&= \left| \int_0^\infty \int_0^\infty H(x,y)\,dF_1(x)\,dF_1(y) \right.\\
&\quad - \int_0^\infty \int_0^\infty H(x,y)\,dF_2(x)\,dF_1(y)\\
&\quad + \int_0^\infty \int_0^\infty H(x,y)\,dF_2(x)\,dF_1(y)\\
&\quad \left. - \int_0^\infty \int_0^\infty H(x,y)\,dF_2(x)\,dF_2(y) \right|\\
&\le \int_0^\infty \left| \int_0^\infty H(x,y)\,dF_1(x) - \int_0^\infty H(x,y)\,dF_2(x) \right| dF_1(y)\\
&\quad + \left| \int_0^\infty \int_0^\infty H(x,y)\,dF_2(x)\,dF_1(y) \right.\\
&\quad \left. - \int_0^\infty \int_0^\infty H(x,y)\,dF_2(x)\,dF_2(y) \right|\\
&\le \frac{1}{2}\int_0^\infty \|F_1 - F_2\|_{\mathrm{tv}}\,dF_1 + \frac{1}{2}\|F_1 - F_2\|_{\mathrm{tv}} = \|F_1 - F_2\|_{\mathrm{tv}}.
\end{aligned}$$

We then have.

$$\begin{aligned}
|G_*(F_1) - G_*(F_2)| &= \left| \frac{I_1}{M_1} - \frac{I_2}{M_2} \right| = \frac{|M_2 I_1 - M_1 I_2|}{M_1 M_2}\\
&= \frac{|M_2(I_1 - I_2) + I_2(M_2 - M_1)|}{M_1 M_2}\\
&\le \frac{|M_2(I_1 - I_2)| + I_2|M_2 - M_1|}{M_1 M_2}\\
&\le \frac{\|F_1 - F_2\|_{\mathrm{tv}} + |M_1 - M_2|}{M_1 M_2}.
\end{aligned} \tag{2.9}$$

Next,

$$\begin{aligned}
|M_1 - M_2| &= |F_2(0)^2 - F_1(0)^2| = |F_1(0) + F_2(0)||F_1(0) - F_2(0)|\\
&\le 2|F_1(0) - F_2(0)| \le \|F_1 - F_2\|_{\mathrm{tv}}.
\end{aligned} \tag{2.10}$$

In the last inequality above, we have used the fact that

$$|F_1(0) - F_2(0)| - \left| \int \phi\,dF_1 - \int \phi\,dF_2 \right|$$

where

$$\phi(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Substituting (2.10) into (2.9) gives the desired result. $\square$

3. Transfer principles

Any sensible measure of equality must satisfy some kind of so-called transfer principle: broadly speaking, a progressive transfer of income from a rich individual to a poorer one should decrease inequality and a regressive transfer from a poor individual to a richer one should increase inequality. Perhaps the simplest, and certainly one that is regarded as fundamental by many economists, is the Pigou–Dalton principle, which can be formalized mathematically as follows. Consider the distribution

$$P = p_1\delta_{a_1} + p_2\delta_{a_2} + \cdots + p_n\delta_{a_n}, \tag{3.1}$$

where $a_1 < a_2 < \cdots < a_n$ and $\sum_{k=1}^n p_k = 1$ with $0 < p_k < 1$. Next, consider the distribution which results from a mean-preserving transfer of amount $\Delta > 0$ from each individual possessing quantity a_i to those possessing $a_j > a_i$ (i.e. a regressive transfer):

$$P' = p_1\delta_{a_1} + \cdots + p_i\delta_{a_i - \Delta} + \cdots + p_j\delta_{a_j + \Delta'} + \cdots + p_n\delta_{a_n}, \tag{3.2}$$

where $\Delta' = p_i\Delta/p_j$, so that $p_i a_i + p_j a_j = p_i(a_i - \Delta) + p_j(a_j + \Delta')$ and P and P' have the same mean. The mean-preserving condition is a natural requirement because it simply says that if one takes away amount Δ from each individual with a_i, the fairest and most natural way to distribute this amount among the individuals with a_j is to divide the total equally among them. We further assume that the transfer does not change the rankings of the individual income levels: that is

$$a_1 < \cdots < a_{i-1} < a_i - \Delta < a_{i+1} < \cdots < a_{j-1} < a_j + \Delta' < a_{j+1} < \cdots < a_n.$$

Then an index of inequality I is said to satisfy the Pigou–Dalton principle if $I(P) < I(P')$.

[2] Suppose $F_n \to_{\mathrm{tv}} F$. Then $F_n \to_{\mathrm{sw}} F$ and since G_* is continuous with respect to setwise convergence, $G_*(F_n) \to G_*(F)$ and hence also continuous in total variation.

We shall now see if G_* satisfies the Pigou–Dalton principle. Observe that

$$G_*(P) - G_*(P') = 2(G_*^<(a_i, a_j, \Delta) + G_*^>(a_i, a_j, \Delta)), \tag{3.3}$$

where

$$\begin{aligned}
G_*^< &= \sum_{k=1}^{i-1} h(a_i/a_k)p_ip_k + \sum_{\substack{k=1\\k\neq i}}^{j-1} h(a_j/a_k)p_jp_k \\
&\quad - \sum_{k=1}^{i-1} h((a_i-\Delta)/a_k)p_ip_k - \sum_{\substack{k=1\\k\neq i}}^{j-1} h((a_j+\Delta')/a_k)p_jp_k \\
&\quad + h(a_j/a_i)p_ip_j - h((a_j+\Delta')/(a_i-\Delta))p_ip_j \\
&= \sum_{k=1}^{i-1}\big(h(a_i/a_k) - h((a_i-\Delta)/a_k)\big)p_ip_k \\
&\quad + \sum_{k=1}^{i-1}\big(h(a_j/a_k) - h((a_j+\Delta')/a_k)\big)p_jp_k \\
&\quad + \sum_{k=i+1}^{j-1}\big(h(a_j/a_k) - h((a_j+\Delta')/a_k)\big)p_jp_k \\
&\quad + h(a_j/a_i)p_ip_j - h((a_j+\Delta')/(a_i-\Delta))p_ip_j
\end{aligned} \tag{3.4}$$

and

$$\begin{aligned}
G_*^> &= \sum_{\substack{k=i+1\\k\neq j}}^{n} h(a_i/a_k)p_ip_k + \sum_{k=j+1}^{n} h(a_j/a_k)p_jp_k \\
&\quad - \sum_{\substack{k=i+1\\k\neq j}}^{n} h((a_i-\Delta)/a_k)p_ip_k - \sum_{k=j+1}^{n} h((a_j+\Delta')/a_k)p_jp_k \\
&= \sum_{k=i+1}^{j-1}\big(h(a_i/a_k) - h((a_i-\Delta)/a_k)\big)p_ip_k \\
&\quad + \sum_{k=j+1}^{n}\big(h(a_i/a_k) - h((a_i-\Delta)/a_k)\big)p_ip_k \\
&\quad + \sum_{k=j+1}^{n}\big(h(a_j/a_k) - h((a_j+\Delta')/a_k)\big)p_jp_k
\end{aligned} \tag{3.5}$$

The terms in $G_*^>$ involve ratios of a_i and a_j to higher amounts a_k while those in $G_*^<$ involve ratios of a_i and a_j to lower amounts a_k. The term involving the ratio of a_i and a_j themselves need only be included in one of $G_*^<$ and $G_*^>$ because it is already multiplied by 2 by virtue of (3.3).

For Δ sufficiently small, the following Taylor approximations hold uniformly for all a_m because there are only finitely many a_m:

$$h\left(\frac{a_i}{a_k}\right) - h\left(\frac{a_i-\Delta}{a_k}\right) \sim h'(a_i/a_k)\Delta/a_k \tag{3.6a}$$

$$h\left(\frac{a_j}{a_k}\right) - h\left(\frac{a_j+\Delta'}{a_k}\right) \sim -h'(a_j/a_k)\Delta'/a_k = -h'(a_j/a_k)\frac{p_i}{p_ja_k}\Delta \tag{3.6b}$$

$$h\left(\frac{a_j}{a_i}\right) - h\left(\frac{a_j+\Delta'}{a_i-\Delta}\right) \sim h'(a_j/a_i)\left(\frac{a_ip_i+a_jp_j}{p_ja_i^2}\right)\Delta. \tag{3.6c}$$

Substituting (3.6) into (3.4) and (3.5) shows that

$$\begin{aligned}
\frac{G_*^< + G_*^>}{p_i\Delta} &= \sum_{k=1}^{i-1}(h'(a_i/a_k) - h'(a_j/a_k))p_k/a_k \\
&\quad - \sum_{k=i+1}^{j-1} h'(a_j/a_k)p_k/a_k + \sum_{i+1}^{j-1} h'(a_i/a_k)p_k/a_k \\
&\quad + \sum_{k=j+1}^{n}(h'(a_i/a_k) - h'(a_j/a_k))p_k/a_k - h'(a_j/a_i)\left(\frac{a_ip_i+a_jp_j}{a_i^2}\right).
\end{aligned} \tag{3.7}$$

Note that the only terms in (3.7) which can be positive are those involving $h'(a_i/a_k) - h'(a_j/a_k)$ because h is decreasing for $x < 1$ and increasing for $x > 1$, so $h'(a_j/a_i) > 0$ and $h'(a_j/a_k) > 0$ for $k < j$ while $h'(a_i/a_k) < 0$ for $k > i$.

It appears that (3.7) is negative (i.e. the Pigou–Dalton principle is satisfied) for some distributions but not others. A natural first question might then be which function h allows the Pigou–Dalton principle to be satisfied for the most distributions. However, it is perhaps more instructive to consider how the Pigou–Dalton principle can fail for G_*.

Consider the following variation of the example at (1.1):

$$P = \left(1-\frac{1}{3n}\right)\delta_1 + \left(1-\frac{1}{3n}\right)\delta_2 + \left(1-\frac{1}{3n}\right)\delta_3 + \frac{1}{n}\delta_n. \tag{3.8}$$

Consider a mean-preserving transfer of Δ from those possessing amount 3 to those possessing amount n, resulting in the following distribution:

$$P' = \left(1-\frac{1}{3n}\right)\delta_1 + \left(1-\frac{1}{3n}\right)\delta_2 + \left(1-\frac{1}{3n}\right)\delta_{3-\Delta} + \frac{1}{n}\delta_{n+(3n-1)\Delta/3}.$$

Then (3.7) becomes

$$\begin{aligned}
&h'(3)\left(1-\frac{1}{3n}\right) + \frac{1}{2}h'(3/2)\left(1-\frac{1}{3n}\right) - h'(n)\left(1-\frac{1}{3n}\right) \\
&\quad - \frac{1}{2}h'(n/2)\left(1-\frac{1}{3n}\right) - h'(n/3)\left(\frac{4-1/n}{9}\right) \\
&\quad \geq \left(1-\frac{1}{3n}\right)(h'(3) + h'(3/2)/2) \\
&\quad - h'(n) - h'(n/2)/2 - \frac{4}{9}h'(n/3).
\end{aligned}$$

Since $h(x)$ is increasing for $x > 1$ and $h(n) \to 1$ as $n \to \infty$, $h'(n) \to 0$ and the above can always be made positive by taking n sufficiently large, so the Pigou–Dalton principle fails.

What is going wrong in the above example is that a transfer from a poorer individual to wealthier one *decreases* inequality among the group with incomes below the poorer individual and also among the group with incomes above the wealthier individual. At the same time, if the difference in income between the poorer and the wealthier individual is very large and if there are very few individuals with incomes between the two, the *increase* in inequality among the small group with incomes between the extremely poor and extremely rich individuals cannot counterbalance the decrease in inequality at the top and bottom ends because this increase in inequality is essentially measured by h' evaluated at a very large income ratio, which has a very small value. This is a direct consequence of the monotonicity and boundedness of h; in fact, it is not necessary for h to be differentiable, merely that $|h(u) - h(v)| \to 0$ as $u, v \to \infty$. (We leave the reader to check that, if on the other hand, a similar transfer is made from those possessing 1 to those possessing n, then Pigou–Dalton is satisfied, for in that case inequality has increased across the board. This property can be generalized in Theorem 3.2.)

The example (3.8) is by no means an isolated artificial example; the same argument works for any distribution with two sufficiently large income groups separated by a large difference. In that case, the increase in inequality resulting from a transfer between the two income groups cannot offset the decrease in inequality within each group. Moreover, the same argument applies more generally to any H which is bounded and strictly increasing in each variable. This is because $\partial H/\partial x$ and $\partial H/\partial y$ must converge to 0, and in these circumstances, it is always possible to construct a counter-example along the lines of (3.8). We have therefore established the following result:

Theorem 3.1. *Any index of the form (2.2) where H is a bounded function which is strictly increasing in each variable cannot satisfy the Pigou–Dalton transfer principle.*

Corollary. *If I is an index of the form (2.2) where H is eventually strictly increasing in each variable and I satisfies the Pigou–Dalton transfer principle, then H is unbounded and hence I is discontinuous with respect to any mode of convergence of probability measures.*

However, if the transfer is between the extremities of a distribution of the form (3.1), a form of Pigou–Dalton transfer principle can be satisfied. This can be done by making the group below the extremely poor individual and the group above the extremely rich individual sufficiently small so that the decrease in inequality among those groups do not dominate as they do in the counter-example just given.

Theorem 3.2. *For the distribution (3.1), consider a mean-preserving transfer resulting in the distribution (3.2), where $\Delta' = p_i\Delta/p_j$. Suppose $|h'(x)| \le C < \infty$ for all x. If p_k for $k < i$ and $k > j$ are sufficiently small, then for all $\Delta > 0$ such that $a_i - \Delta > a_{i-1}$ and $a_j + \Delta' < a_{j+1}$*

$$G_*(P) < G_*(P').$$

In addition, if $h(x)$ is convex for $x < 1$, then the same conclusion holds if p_k are sufficiently small only for $k < i$.

Proof. The only term in (3.4) which can be positive is

$$\sum_{k=1}^{i-1}\big(h(a_i/a_k) - h((a_i-\Delta)/a_k)\big)p_ip_k \tag{3.9}$$

and the only term in (3.5) which can be positive is

$$\sum_{k=j+1}^{n}\big(h(a_j/a_k) - h((a_j+\Delta')/a_k)\big)p_jp_k \tag{3.10}$$

because h is decreasing for $x < 1$ and increasing for $x > 1$. By the Mean Value Theorem,

$$\begin{aligned} &h(a_i/a_k) - h((a_i-\Delta)/a_k)\\ &= h'(a_{i,k})\Delta/a_k \quad \text{for some } (a_i-\Delta)/a_k \le a_{i,k} \le a_i/a_k \qquad (3.11a)\\ &h(a_j/a_k) - h((a_j+\Delta')/a_k)\\ &= -h'(a_{j,k})\Delta'/a_k \quad \text{for some } a_j/a_k \le a_{j,k} \le (a_j+\Delta')/a_k. \qquad (3.11b)\end{aligned}$$

If h' is bounded then the quantities (3.9) and (3.10) can be made sufficiently small by taking p_k sufficiently small for $k < i$ and $k > j$.

Substituting (3.11) into (3.5) gives

$$\begin{aligned}&\sum_{k=j+1}^{n}\big(h(a_i/a_k) - h((a_i+\Delta)/a_k)\big)p_ip_k\\ &\quad+\sum_{k=j+1}^{n}\big(h(a_j/a_k) - h((a_j+\Delta')/a_k)\big)p_jp_k\\ &= p_i\Delta\sum_{k=j+1}^{n}(h'(a_{i,k}) - h'(a_{j,k}))p_k/a_k. \end{aligned}\tag{3.12}$$

For $k \ge j+1$, $a_{i,k} < a_{j,k} < 1$ and if $h(x)$ is convex for $x < 1$ then $h'(a_{i,k}) - h'(a_{j,k}) \le 0$, so in fact the entire quantity in (3.5) is negative. Hence we only need to ensure that (3.9) is sufficiently small in order that $G_*(P) < G_*(P')$ and this can be achieved by taking p_k sufficiently small for $k < i$. □

It can be argued that a failure of the Pigou–Dalton principle in manner described above is not unreasonable. If a very poor individual loses a little while an extremely rich individual gains a little, then their position relative to each other has not changed a great deal. On the other hand a transfer between individuals with similar incomes has a more significant effect on the relative inequality between them. It might therefore be more relevant to consider a local version of the Pigou–Dalton principle: broadly speaking, a transfer from an individual to a wealthier *peer* should increase inequality.

Suppose that $a_j - a_i \le \epsilon$ for some $\epsilon > 0$. By the Mean Value Theorem, $h'(a_i/a_k) - h'(a_j/a_k) = -h''(a_{ij}/a_k)(a_j - a_i)/a_k$ for some $a_i \le a_{ij} \le a_j$. If $h''(a_{ij}/a_k) < 0$, then $h'(a_i/a_k) - h'(a_j/a_k) \le h''_{\max}\epsilon/a_k$, where $h''_{\max} = \max_{k,m} h''(a_m/a_k)$. Finally, for ϵ sufficiently small, there exists some $0 < c < 1$ such that $h'(a_j/a_k) > ch'(1+) > 0$ and $h'(a_i/a_k) < ch'(1-) < 0$ for $i+1 \le k \le j-1$. Therefore, if ϵ is sufficiently small, (3.7) is negative and inequality has increased.

We have now established the following local Pigou–Dalton principle:

Theorem 3.3. *For the distribution (3.1), suppose that for some i, j, $a_j - a_i \le \epsilon$ for some $\epsilon > 0$. Consider a mean-preserving transfer from a_i to a_j, resulting in the distribution (3.2), where $\Delta' = p_i\Delta/p_j$. Then there exists some $\epsilon > 0$ such that for $a_j - a_i \le \epsilon$, $G_*(P) < G_*(P')$ for all Δ sufficiently small so that the asymptotics in (3.6) hold.*

This local transfer principle can be extended to non-local transfers via a sequence of local transfers provided there are sufficiently many values between a_i and a_j.

Corollary. *Suppose that $a_{k+1} - a_k < \epsilon$ for $k = i, i+1, \ldots, j-1$. Consider a mean-preserving transfer from a_i to a_j, resulting in the distribution (3.2), where $\Delta' = p_i\Delta/p_j$. Then there exists some $\epsilon > 0$ such that $G_*(P) < G_*(P')$ for all Δ sufficiently small.*

Proof. Firstly, fix some small $\Delta > 0$. By Theorem 3.3, there exists $\epsilon_i > 0$ such that a mean-preserving transfer of Δ from individual i to $i+1$ results in a distribution with a_i replaced by $a_i - \Delta$ and a_{i+1} replaced by $a_{i+1} + p_i\Delta/p_{i+1}$ which has higher inequality.

Next, consider a mean-preserving transfer of $\tilde{\Delta} := p_i\Delta/p_{i+1}$ from individual $i+1$ to $i+2$, which results in a distribution with $a_{i+1} + p_i\Delta/p_{i+1}$ replaced by a_{i+1} and a_{i+2} replaced by $a_{i+2} + p_i\Delta/p_{i+2}$. For such a transfer, (3.7) now becomes

$$\begin{aligned} D(\tilde{\Delta}) &:= \sum_{k=1}^{i}\big(h'((a_{i+1}+\tilde{\Delta})/a_k) - h'(a_{i+2}/a_k)\big)p_k/a_k\\ &\quad+\sum_{k=i+3}^{n}\big(h'((a_{i+1}+\tilde{\Delta})/a_k) - h'(a_{i+2}/a_k)\big)p_k/a_k\\ &\quad- h'(a_{i+2}/(a_{i+1}+\tilde{\Delta}))\left(\frac{(a_{i+1}+\tilde{\Delta})p_{i+1} + a_{i+2}p_{i+2}}{(a_{i+1}+\tilde{\Delta})^2}\right).\end{aligned}$$

By the same argument as in the proof of Theorem 3.3, there exists ϵ_{i+1} so that $D(\Delta_0) < 0$ for all $\Delta_0 \le \tilde{\Delta}$.

This procedure can be repeated until a final transfer of $p_i\Delta/p_{j-1}$ from $j-1$ to j. The desired result then follows by taking $\epsilon = \min(\epsilon_i, \epsilon_{i+1}, \ldots, \epsilon_{j-1})$ and Δ such that $p_i\Delta/p_{i+k} < \epsilon$ for all $k = 1, \ldots, j-i-1$. □

Admittedly, this local form of Pigou–Dalton principle is rather weak. One of the reasons the Pigou–Dalton principle may be difficult to satisfy is that it is a requirement on the level of transfers between individuals. Some authors have considered other weaker forms of transfer principle which work at the population level. One such transfer principle is the so-called concentration principle - see Ebert (2009, 2010). Consider the following redistribution of income among the entire population with income distribution (3.1). Fix $0 < r \le 1$ and tax everybody at rate r, then divide the total tax thus raised equally among everyone in the population. An individual with income a_i will therefore have income $a_i' = (1-r)a_i + r\bar{a}$ after redistribution, where $\bar{a} = \frac{1}{n}\sum_{i=1}^n a_i$ is the mean income of the distribution. Let P' be the resulting income distribution:

$$P' = p_1\delta_{a_1'} + \cdots + p_n\delta_{a_n'}, \tag{3.13}$$

An index I is said to satisfy the concentration principle if $I(P') \le I(P)$.

Theorem 3.4. *The index G_* satisfies the concentration principle.*

Proof. This is a direct consequence of the following elementary inequality: for $a, b, c > 0$

$$\frac{a+c}{b+c} - \frac{a}{b} \begin{cases} < 0 & \text{if } \frac{a}{b} > 1, \\ > 0 & \text{if } \frac{a}{b} < 1. \end{cases}$$

Now apply this with $a = (1-r)a_i$, $b = (1-r)a_j$ and $c = r\bar{a}$, so $a_i' = a + c$ and $a_j' = b + c$. Since $h(x)$ is decreasing for $x < 1$ and increasing for $x > 1$,

$$h\left(\frac{a_i'}{a_j'}\right) - h\left(\frac{a_i}{a_j}\right) \le 0.$$

Summing over i, j and noting that $P(0) \ge 0$ while $P'(0) = 0$ shows that $G_*(P') \le G_*(P)$. □

We saw earlier that the Pigou–Dalton transfer principle involves considering transfers between groups of individuals and these transfers often have the effect of increasing inequality in some parts of the income distribution while simultaneously decreasing inequality in other parts. The Pigou–Dalton principle effectively says that, in the case of a regressive (resp. progressive) transfer, the resulting increase (resp. decrease) in inequality in the relevant parts of the income distribution must always counter-balance any decrease (resp. increase) in inequality elsewhere in the distribution, and this must hold for all distributions. This seems a rather strong requirement and it is far from obvious that, for the purposes of measuring inequality, this ought to be the case in all circumstances. Whether one regards this as a reasonable requirement would likely depend not only on mathematical considerations but also on one's subjective value judgements as to what constitutes inequality. One of the main motivations for the development of the concentration principle in Ebert (2009) is a series of empirical studies carried out earlier in which respondents were asked various questions relating to their perceptions of inequality and ways of addressing it. One of the findings of these empirical studies is that some properties of inequality measures, including the Pigou–Dalton principle, which are considered desirable or even fundamental in the economics literature, are not generally supported by the participants in these experiments.

The key difference between the concentration principle and Pigou–Dalton principle is that the former is based on population-wide redistributions of wealth and therefore does not involve any detailed balancing between changes in inequality in different parts of the population. Ebert (2010) shows that the Pigou–Dalton principle implies the concentration principle. As we have just seen, the converse is not true.

4. Decomposability

We saw in the last section that the failure of the Pigou–Dalton principle is related to the balance between changes in inequality within certain income groups and opposite changes between those groups. This raises the question of decomposability, which has been considered by many authors, for example see Ebert (2010) and the references therein. Here, we show that the index G_* also admits similar decompositions.

Consider two disjoint intervals $[0, a]$ and (a, ∞). For an income distribution F on $[0, \infty)$, let F_1 be the distribution which is F restricted to $[0, a]$,

$$F_1(x) = \begin{cases} \frac{F(x)}{F(a)} & 0 \le x \le a \\ 1 & x > a. \end{cases}$$

and F_2 be the distribution which is F restricted to (a, ∞),

$$F_2(x) = \begin{cases} 0 & 0 \le x \le a \\ \frac{F(x)-F(a)}{1-F(a)} & x > a. \end{cases}$$

Then G_* admits the following decomposition.

Theorem 4.1. *Suppose $F(0) = 0$ and let $0 < w = F(a) < 1$ be the weight which F puts on $[0, a]$. For $\tilde{x} \in [0, a]$ and $\tilde{y} \in (a, \infty)$, let $\widetilde{P}$ be the distribution*

$$\widetilde{P} = w\delta_{\tilde{x}} + (1-w)\delta_{\tilde{y}}.$$

Then G_* admits the decomposition

$$w^2 G_*(F_1) + (1-w)^2 G_*(F_2) + G_*(\widetilde{P})$$

for some $\tilde{x} \in [0, a]$ and $\tilde{y} \in (a, \infty)$.

The proof relies on the following version of the Mean Value Theorem.

Lemma 4.2. *Let $(\Omega, \mathcal{F})$ be a measurable space, f a real-valued measurable function on Ω and μ a measure on Ω. Let $S \subset \Omega$ be a measurable subset with $0 < \mu(S) < \infty$ for which there exists $-\infty \le a \le b \le +\infty$ such that*

$$a \le f(x) \le b \quad \forall x \in S$$

and for all $y \in (a, b)$, there exists $x \in S$ such that $f(x) = y$. Then there exists some $\tilde{x} \in S$ such that

$$\frac{1}{\mu(S)} \int_S f(x)\,\mu(dx) = f(\tilde{x}).$$

The statement of the above lemma already contains all the ingredients of its own proof, which will be left to the reader as an exercise. Note that, unlike most versions of the Mean Value Theorem, there are no continuity assumptions on f or μ (in fact, there is no assumption that Ω even has a topology). For the purposes of the present application, the measure μ could be purely atomic. Instead, the assumptions in Lemma 4.2 are mostly restrictions on the set S.

Proof of Theorem 4.1. Since $F = wF_1 + (1-w)F_2$, we have

$$\begin{aligned} G_*(F) &= \int_0^\infty \int_0^\infty H(x,y)\,dF(x)\,dF(y) \\ &= w^2 \int_0^a \int_0^a H(x,y)\,dF_1(x)\,dF_1(y) \\ &\quad + (1-w)^2 \int_a^\infty \int_a^\infty H(x,y)\,dF_2(x)\,dF_2(y) \\ &\quad + 2w(1-w) \int_a^\infty \int_0^a H(x,y)\,dF_1(x)\,dF_2(y) \\ &= w^2 G_*(F_1) + (1-w)^2 G_*(F_2) \\ &\quad + 2w(1-w) \int_a^\infty \int_0^a H(x,y)\,dF_1(x)\,dF_2(y). \end{aligned} \tag{4.1}$$

By Lemma 4.2 applied to $\mu = F_1 \times F_2$ and $S = [0,a] \times (a,\infty)$, there exist $\tilde{x} \in [0,a]$ and $\tilde{y} \in (a,\infty)$ such that

$$\int_a^\infty \int_0^a H(x,y)\,dF_1(x)\,dF_2(y) = H(\tilde{x},\tilde{y}).$$

But

$$2w(1-w)H(\tilde{x},\tilde{y}) = G_*(\widetilde{P}).$$

Substituting into (4.1) gives the desired result. □

If $F(0) > 0$, a similar decomposition also holds but there will be an extra coefficient depending on $F(0)$ in front of the between-groups component $G_*(\widetilde{P})$.

The values $\tilde{x}$ and $\tilde{y}$ may be interpreted as representatives of the typical or overall incomes in the income groups $[0,a]$ and (a,∞). However, unless H is linear, they are not the means of the respective groups. For this reason, the between-groups component $G_*(\widetilde{P})$ will in general depend on the inequalities within each group; in particular, mean-preserving transfers within each group will not leave $G_*(\widetilde{P})$ unchanged because $\tilde{x}$ and $\tilde{y}$ depend on F_1 and F_2 respectively. Theorem 4.1 is therefore a weaker form of decomposition than that considered by several authors, for example Shorrocks (1980) and Foster and Shneyerov (1999). Indeed, Theorem 2 of Shorrocks (1980) and Proposition 2 of Foster and Shneyerov (1999) give necessary and sufficient conditions for the between-groups component to be independent of inequalities within each group for the respective types of decomposition considered in those papers. The index G_* does not satisfy any of those conditions.

The property that within-groups transfers leave the between-groups component unchanged is often useful but not always desirable. Ebert (2010) presents a simple example which illustrates this point. Consider 2 income groups each with 2 individuals having incomes $(10, 20)$ and $(15, 15)$ respectively. The mean of both groups is 15 and if this is chosen as the representative income of each group, then the between-groups inequality is 0. On the other hand, if the second income group is changed to $(5, 25)$ by a transfer of 10 within that group, the mean income does not change so the between-groups inequality is still 0. However, Ebert (2010) argues that since Group 1 no longer contains the poorest and richest individuals in the population, the between-groups inequality has in fact changed significantly. Moreover, any further regressive transfer within Group 2 would have the effect of moving both individuals in Group 2 further away from each individual in Group 1 and therefore between-groups inequality should increase rather than remain unchanged at 0. This example is used by the author to motivate the type of decompositions considered in that paper. This type of decomposition does not involve any smoothing of the distributions of income groups but instead uses all the information about the between-groups inequality via a direct pairwise comparison of incomes from each group. The index G_* also satisfies this type of decomposition.

Suppose that the p_i in the distribution (3.1) are rational and that $a_i > 0$ so that $P(0) = 0$, then P may be written as

$$P = \frac{1}{N}(\delta_{x_1} + \cdots + \delta_{x_N}) \tag{4.2}$$

for some N, where the incomes x_i may not be distinct. If $a_1 = x_1 = 0$, a slightly different treatment is required. In this case, P may be written as

$$P = p_0\delta_0 + \frac{1}{K}(\delta_{x_1} + \cdots + \delta_{x_K}). \tag{4.3}$$

For $x_i, x_j > 0$, let

$$Q(x_i,x_j) = \frac{1}{2}\delta_{x_i} + \frac{1}{2}\delta_{x_j}$$

and for $x_j > 0$, let

$$R(0,x_j) = \frac{1}{p_o + 1/K}\left(p_0\delta_0 + \frac{1}{K}\delta_{x_j}\right).$$

We are now in a position to establish the following decomposition:

Theorem 4.3.

1. *If $P(0) = 0$, consider a partition of $\{x_1, x_2, \ldots, x_N\}$ into two disjoint income groups $I_1 = \{x_{i_1}, x_{i_2}, \ldots, x_{i_n}\}$ and $I_2 = \{x_{j_1}, x_{j_2}, \ldots, x_{j_m}\}$ where $m+n = N$. For $k = 1, 2$, let P_k be the measure P at (4.2) restricted to income group I_k, normalized by $P(I_k)$. Then*

$$G_*(P) = \frac{n^2}{N^2}G_*(P_1) + \frac{m^2}{N^2}G_*(P_2) + \frac{4}{N^2}\sum_{r=1}^{n}\sum_{s=1}^{m} G_*(Q(x_{i_r}, x_{j_s})). \tag{4.4}$$

2. *If $P(0) = p_0 > 0$, consider income groups $I_1 = \{0, x_{i_1}, x_{i_2}, \ldots, x_{i_n}\}$ and I_2 as before, where $m+n = K$. For $k = 1, 2$, let P_k be the measure P at (4.3) restricted to income group I_k, normalized by $P(I_k)$. Then*

$$\begin{aligned} G_*(P) &= P(I_1)^2\left(\frac{1-P_1(0)^2}{1-p_0^2}\right)G_*(P_1) + P(I_2)^2 G_*(P_2) \\ &\quad + \left(p_0 + \frac{1}{K}\right)^2 \sum_{s=1}^{m} G_*(R(0,x_{j_s})) \\ &\quad + \frac{4}{K^2(1-p_0^2)}\sum_{r=1}^{n}\sum_{s=1}^{m} G_*(Q(x_{i_r}, x_{j_s})). \end{aligned} \tag{4.5}$$

Proof. The proof is a matter of simple algebra. For the proof of (1), first note that

$$G_*(P_1) = \frac{1}{n^2}\sum_{r=1}^{n}\sum_{s=1}^{n} H(x_{i_r}, x_{i_s})$$

$$G_*(P_2) = \frac{1}{m^2}\sum_{r=1}^{m}\sum_{s=1}^{m} H(x_{j_r}, x_{j_s})$$

$$G_*(Q(x_{i_r}, x_{j_s})) = \frac{1}{2}H(x_{i_r}, x_{j_s})$$

Then

$$\begin{aligned} G_*(P) &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} H(x_i,x_j) \\ &= \frac{1}{N^2}\left[\sum_{r=1}^{n}\sum_{s=1}^{n} H(x_{i_r}, x_{i_s}) + \sum_{r=1}^{m}\sum_{s=1}^{m} H(x_{j_r}, x_{j_s})\right. \end{aligned}$$

$$+ 2\sum_{r=1}^{n}\sum_{s=1}^{m}H(x_{i_r}, x_{j_s})\Bigg]$$
$$= \frac{n^2}{N^2}G_*(P_1) + \frac{m^2}{N^2}G_*(P_2) + \frac{4}{N^2}\sum_{r=1}^{n}\sum_{s=1}^{m}G_*(Q(x_{i_r}, x_{j_s})).$$

Similarly, for the proof of (2), we have

$$\begin{aligned}
G_*(P) = \frac{1}{1-p_0^2}\Bigg[& 2\sum_{r=1}^{n}\frac{p_0}{K}H(0, x_{i_r}) + \frac{1}{K^2}\sum_{r=1}^{n}\sum_{s=1}^{n}H(x_{i_r}, x_{i_s}) \\
& + \frac{1}{K^2}\sum_{r=1}^{m}\sum_{s=1}^{m}H(x_{j_r}, x_{j_s}) \\
& + 2\sum_{s=1}^{m}\frac{p_0}{K}H(0, x_{j_s}) + \frac{2}{K^2}\sum_{r=1}^{n}\sum_{s=1}^{m}H(x_{i_r}, x_{j_s})\Bigg] \\
= \frac{1}{1-p_0^2}\Bigg[& P(I_1)^2\left(\frac{2}{P(I_1)^2}\sum_{r=1}^{n}\frac{p_0}{K}H(0, x_{i_r})\right. \\
& \left. + \frac{1}{P(I_1)^2K^2}\sum_{r=1}^{n}\sum_{s=1}^{n}H(x_{i_r}, x_{i_s})\right) \\
& + P(I_2)^2\frac{1}{P(I_2)^2K^2}\sum_{r=1}^{m}\sum_{s=1}^{m}H(x_{j_r}, x_{j_s}) \\
& + 2\left(p_0 + \frac{1}{K}\right)^2\sum_{s=1}^{m}\frac{p_0/K}{(p_0+1/K)^2}H(0, x_{j_s}) \\
& + \frac{4}{K^2}\sum_{r=1}^{n}\sum_{s=1}^{m}\frac{1}{2}H(x_{i_r}, x_{j_s})\Bigg] \\
= P(I_1)^2 & \left(\frac{1-P_1(0)^2}{1-p_0^2}\right)G_*(P_1) + P(I_2)^2G_*(P_2) \\
+ & \left(p_0 + \frac{1}{K}\right)^2\sum_{s=1}^{m}G_*(R(0, x_{j_s})) \\
+ & \frac{4}{K^2(1-p_0^2)}\sum_{r=1}^{n}\sum_{s=1}^{m}G_*(Q(x_{i_r}, x_{j_s})). \quad \square
\end{aligned}$$

The decomposition (4.4) is precisely the property DEC in Ebert (2010) while (4.5) is analogous to the property $\widehat{\text{DEC}}$, applicable to indices normalized to lie between 0 and 1 which are treated separately in Ebert (2010) and referred to by the author as relative indices. The relative indices considered in that paper are all normalized by functions of the mean of the underlying income distribution whereas the normalization of G_* is a function of $P(0)$. For this reason, the weighting functions of each component in the decomposition $\widehat{\text{DEC}}$ of Ebert (2010) are functions of the mean while those in (4.5) are functions of $P(0)$.

Although the decompositions Theorems 4.1 and 4.3 are stated for 2 income groups, both results may be extended by iteration to any finite number of income groups.

5. An application to real data

In this section, we compare the behaviour of G_* with the standard Gini index G_0 on some real-life data. For a sample of observations $X_1, X_2, \ldots, X_n$, we shall use the obvious estimators

$$\widehat{G}_0 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|X_i - X_j|}{2n^2\bar{X}} \tag{5.1}$$

$$\widehat{G}_* = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}H(X_i, X_j)}{n^2(1-\bar{F}_0^2)} \tag{5.2}$$

for G_0 and G_* respectively, where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i, \quad \bar{F}_0 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i=0\}}. \tag{5.3}$$

We shall examine in greater detail the properties of the estimator $\widehat{G}_*$ – in particular the asymptotic normality – later in Section 6. For now, we restrict ourselves to some simple comparisons with the standard Gini index.

We use data provided by HMRC, the UK tax authority, which give the percentiles from 1 to 99% of the total pre-tax income of UK tax payers for the years 2000–2019, except for 2009 for which no data are available.[3] We should emphasize at this point that the exercise undertaken here is in no way intended as an in-depth study of UK income distribution; rather it is intended merely to be a simple comparison of our inequality indices with the Gini index. While these data are by no means the most comprehensive (e.g. they include only tax payers and those in receipt of tax credits) or the most reliable (e.g. the extent of under-declaration of income or other forms of tax fraud is unknown), they nevertheless include a large sample, are easily accessible and provide a reasonable level of detail (many other data sets group the level of income into wide bands).

We employ a rather simplistic and obvious method to recover the income distribution from the percentile data. For example, the data for 2019 look like:

Percentile	Income (GBP)
1	12,100
2	12,300
⋮	⋮
99	175,000

Thus 1% of the sample earned £12,100 or less and we shall assume that everyone in the lowest 1 percentile earned £12,100. Another 1% earned between £12,100 and £12,300 and we shall assume that everyone in that group earned the mid-point of that range, £12,200, and so on. Finally 1% earned £175,000 or more and we shall likewise assume that everyone in the top 1% earned £175,000. The inequality indices for each year are plotted in Fig. 5.1, where we see that, again, all three indices show a very similar pattern of trends over time.

6. Further observations on estimation of the index

So far, we have seen that in situations where the Gini index is well-defined and easily estimated, the indices proposed here behave in similar ways to the Gini index. However, as mentioned in the introduction, estimation of G_0 is more difficult when the underlying distribution has infinite variance; in particular, the estimator $\widehat{G}_0$ at (5.1) cannot be regarded as a reliable estimator. Fontanari et al. (2018) seek to develop other methods of estimation in order to overcome this difficulty. This problem does not arise with the estimator $\widehat{G}_*$ at (5.2), at least in the case where $F(0)$ is known, for the simple reason that the function H used in (2.7) is bounded. Indeed, there is no requirement that the underlying distribution should have any finite moments at all for this very reason. It turns out that, when $F(0)$ is known, $\widehat{G}_*$ is in fact asymptotically normally distributed so one can calculate approximate confidence intervals for $\widehat{G}_*$ in the usual way.

[3] The data can be downloaded from www.gov.uk/government/statistics/percentile-points-from-1-to-99-for-total-income-before-and-after-tax.

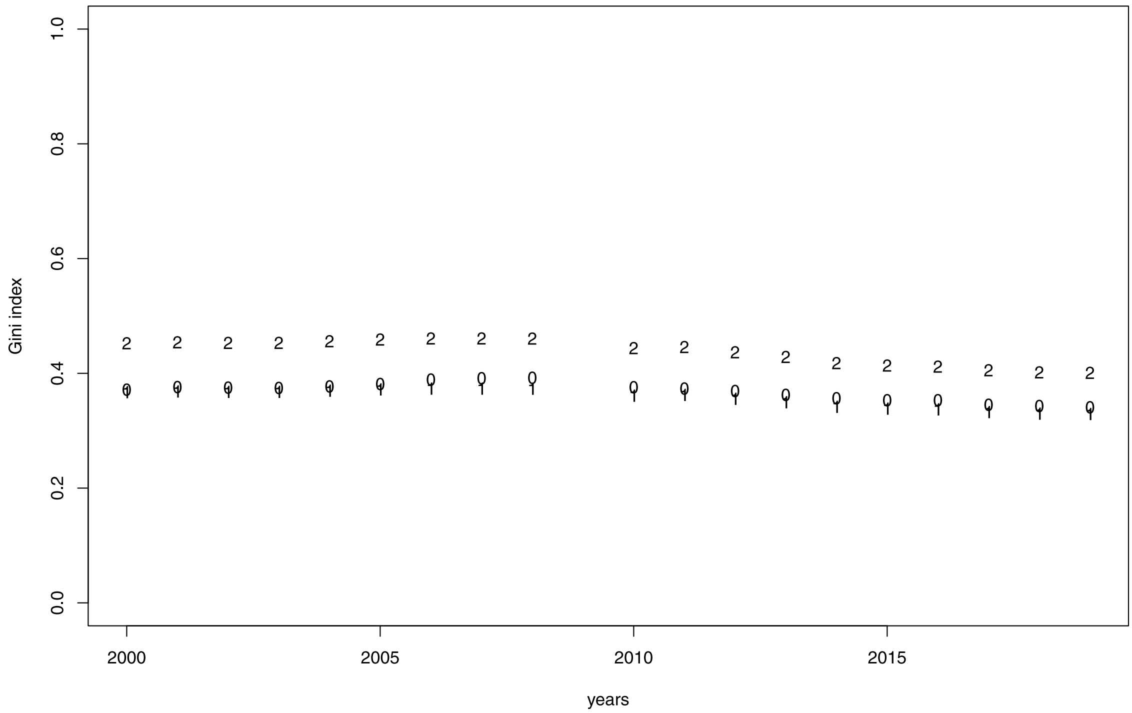

Fig. 5.1. Indices for UK income 2000–2019, "0"= $\widehat{G}_0$, "1"= $\widehat{G}_*(H_1)$ "2"= $\widehat{G}_*(H_2)$.

Before establishing the asymptotic normality of $\widehat{G}_*$, we first compare the behaviour of $\widehat{G}_*$ and $\widehat{G}_0$ on the modulus of a one-sided α-stable distribution whose characteristic function is

$$\psi(\theta) = \mathbb{E}[e^{i\theta X}] = -|\theta|^\alpha[1 - i\text{sgn}(\theta)\tan(\pi\alpha/2)]$$

for $\alpha = 3/2$. This distribution has finite mean but infinite variance and because it has positive probability of being negative, we simply take the modulus. The results are based on 100 samples each consisting of 100 simulated observations from the above stable distribution. We calculate the various inequality indices for the 100 samples and plot the results in Figs. 6.1 to 6.3.

Fig. 6.1 shows the values of $\widehat{G}_0$ for the 100 samples. As expected, the infinite variance of the stable distribution results in a high degree of variation, with some estimates about double the value of some other estimates. Fig. 6.2 shows that increasing the sample size to 1000 results in a noticeable improvement, with significantly less variation in the estimates, but still not ideal.

Fig. 6.3 shows the values of $\widehat{G}_*(H_1)$ and $\widehat{G}_*(H_2)$ for the 100 samples. The estimates seem far more reliable, as one would expect from (6.3).

We next repeat the exercise for the stable distribution with $\alpha = 3/4$. This distribution is supported on $[0, \infty)$ but has infinite mean. Fig. 6.4 shows the values of $\widehat{G}_*(H_1)$ and $\widehat{G}_*(H_2)$ in this case and we see that the estimates are still well-behaved. (There is no point in calculating $\widehat{G}_0$ because the infinite mean renders the results meaningless.)

We next establish the asymptotic normality of $\widehat{G}_*$ in the case that $F(0)$ is known Without loss of generality, we may assume that $F(0) = 0$. In this case

$$\widehat{G}_* = \sum_{i=1}^{n}\sum_{j=1}^{n} H(X_i, X_j) = 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} H(X_i, X_j). \tag{6.1}$$

We begin with some preliminary estimates on the variance of $\widehat{G}_*$. This involves some tedious and messy calculations which are presented in Appendix B. The essential point is that

$$\text{Var}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} H(X_i, X_j)\right) \sim O(n^3). \tag{6.2}$$

Therefore

$$\text{Var}(\widehat{G}_*) \sim O\left(\frac{1}{n}\right). \tag{6.3}$$

This is in contrast to the case of sample means of i.i.d. observations, where the variance is $O(1/n^2)$. Thus, the dependencies among the $H(X_i, X_j)$ have increased the variance significantly.

We shall apply a generalization of the Central Limit Theorem to m-dependent random variables, which is defined as follows:

Definition 6.1. A sequence of random variables $X_1, X_2, \ldots$, is said to be *m-dependent* if $X_1, X_2, \ldots, X_i$ and $X_j, X_{j+1}, \ldots$ are independent whenever $j - i > m$.

Romano and Wolf (2000) established the following generalized Central Limit Theorem, which we state here in a slightly simplified form as it applies to the present situation.

Theorem 6.1. *For each $N = 1, 2, \ldots$, let $Y_1^{(N)}, Y_2^{(N)}, \ldots, Y_N^{(N)}$ be a sequence of m_N-dependent random variables with mean 0. Define*

$$V_{N,k,a}^2 := Var\left(\sum_{i=a}^{a+k-1} Y_i^{(N)}\right)$$

$$V_N^2 := V_{N,N,1}^2 = Var\left(\sum_{i=1}^{N} Y_i^{(N)}\right).$$

Assume the following conditions hold: for some $\delta > 0$ and some $-1 \le \gamma < 1$,

$$\mathbb{E}[|Y_i^{(N)}|^{2+\delta}] \le \Delta_N, \tag{6.4}$$

$$\frac{V_{N,k,a}^2}{k^{1+\gamma}} \le K_N \qquad \forall a, \quad \forall k \ge m_N, \tag{6.5}$$

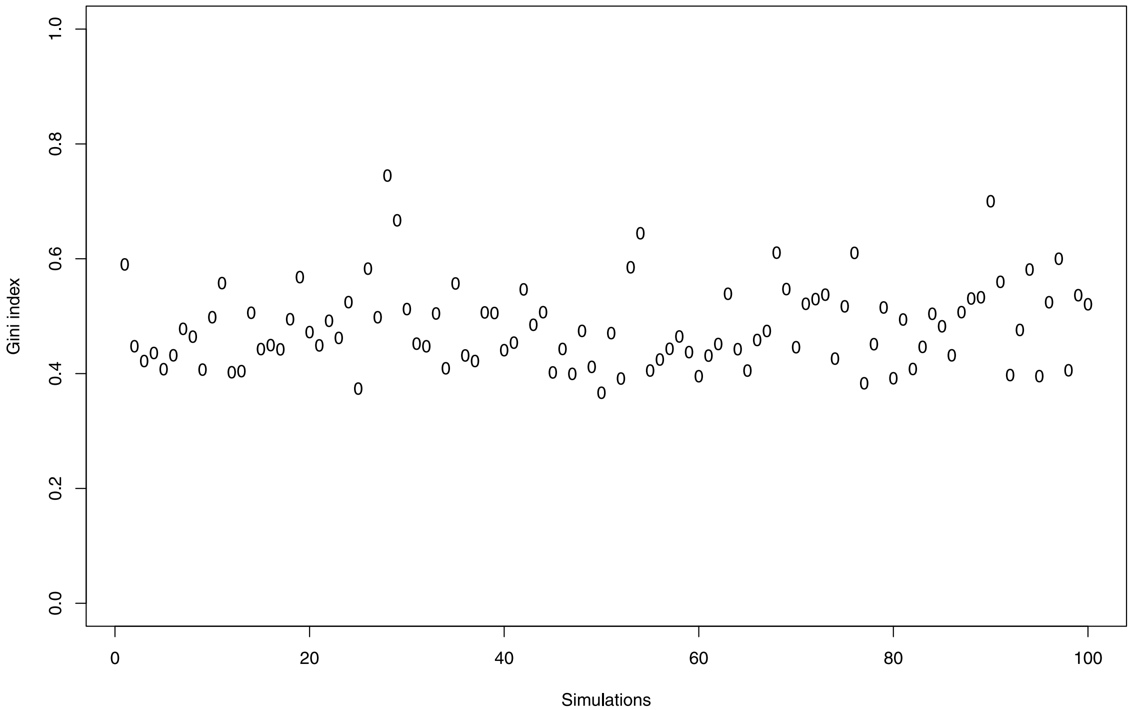

Fig. 6.1. $\widehat{G}_0$ for 100 simulated samples of size 100.

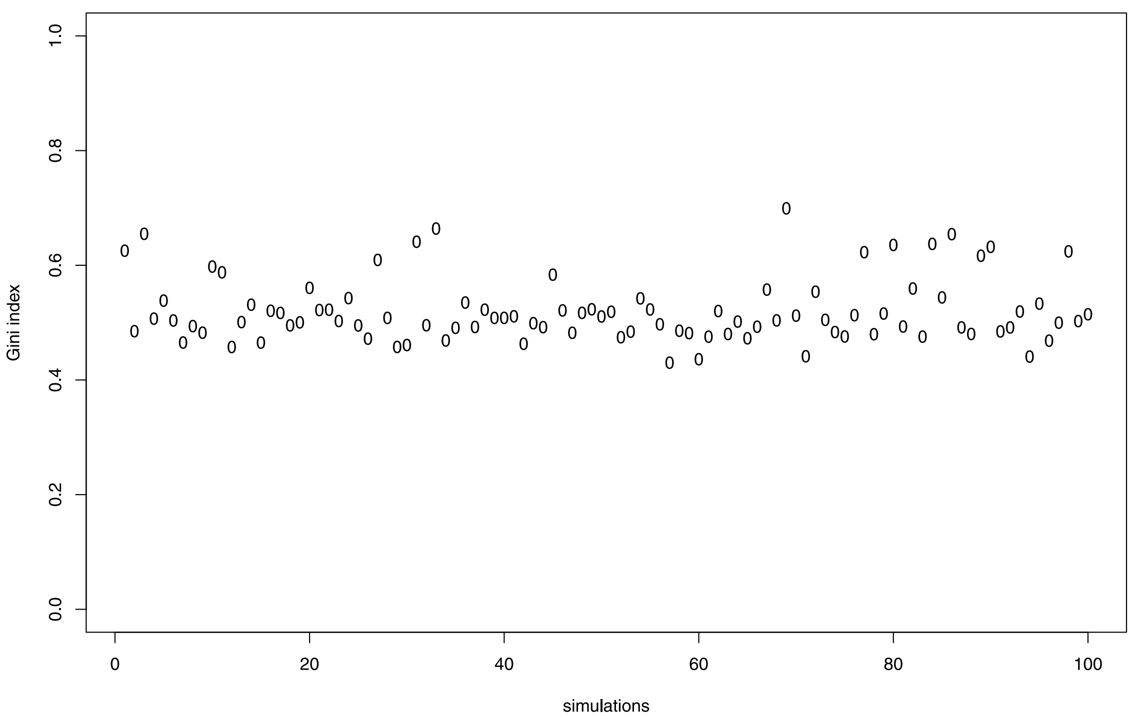

Fig. 6.2. $\widehat{G}_0$ for 100 simulated samples of size 1000.

$$\frac{V_N^2}{Nm_N^{\gamma}} \geq L_N, \tag{6.6}$$

$$\frac{K_N}{L_N} \sim O(1), \tag{6.7}$$

$$\frac{\Delta_N}{L_N^{(2+\delta)/2}} \sim O(1) \tag{6.8}$$

$$\frac{m_N^{1+(1-\gamma)(1+2/\delta)}}{N} \to 0. \tag{6.9}$$

Then

$$\frac{1}{V_N}\sum_{i=1}^{N} Y_i^{(N)} \Rightarrow N(0, 1).$$

To make the application of this theorem easier, it is convenient to relabel the random variables $H(X_i, X_j)$ using a single index. The terms in the second summation in (6.1) have indices (i, j) as

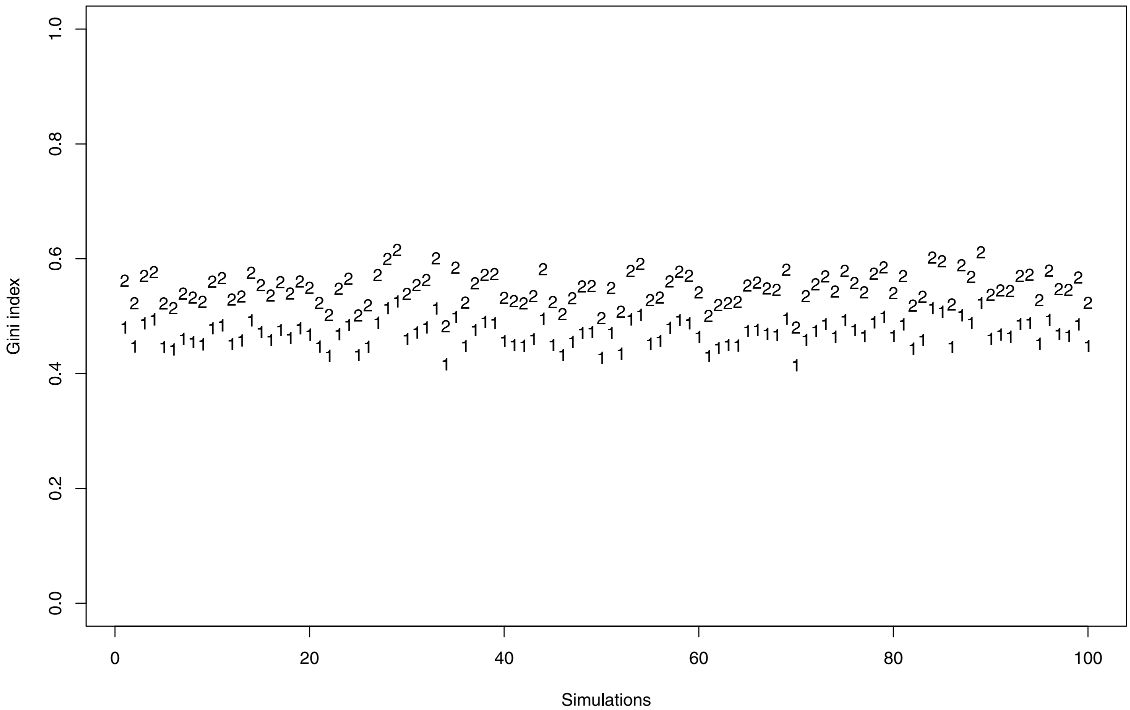

Fig. 6.3. $\widehat{G}_*$ for 100 simulated samples of size 100 $\alpha = 3/2$, "1"= $\widehat{G}_*(H_1)$ "2"= $\widehat{G}_*(H_2)$.

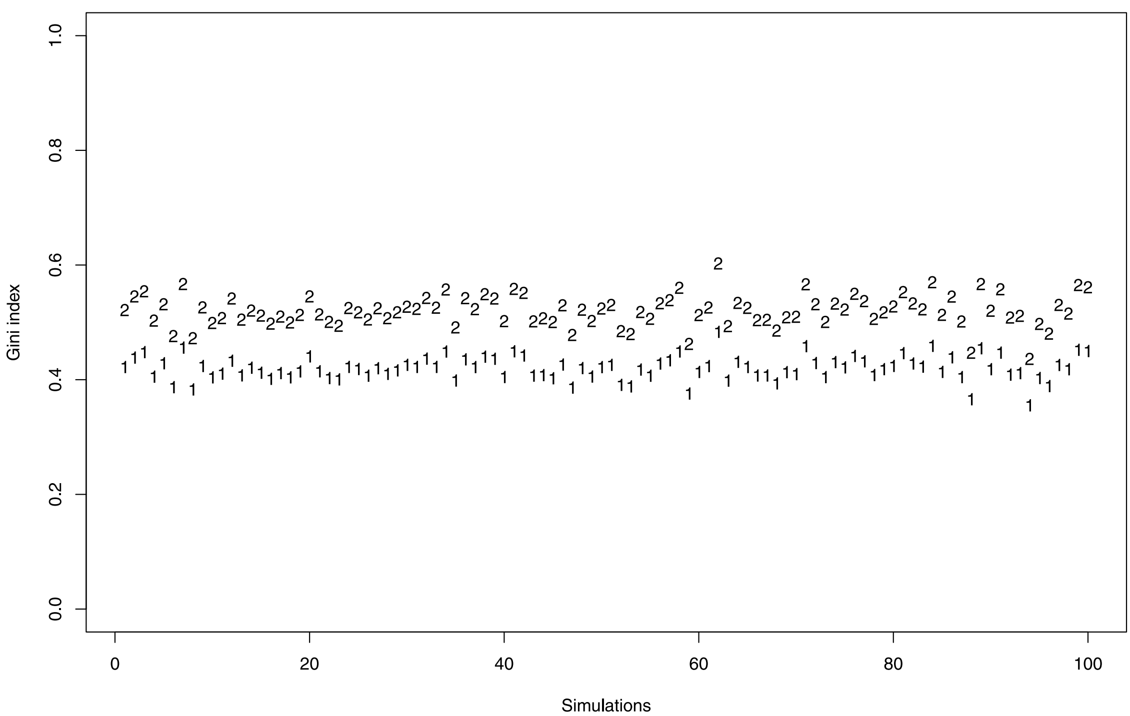

Fig. 6.4. $\widehat{G}_*$ for 100 simulated samples of size 100 $\alpha = 3/4$, "1"= $\widehat{G}_*(H_1)$ "2"= $\widehat{G}_*(H_2)$.

follows:

$$\begin{array}{lllll}(2,1) & & & & \\ (3,1) & (3,2) & & & \\ (4,1) & (4,2) & (4,3) & & \\ (5,1) & (5,2) & (5,3) & (5,4) & \\ \vdots & \vdots & \vdots & \vdots & \\ (n,1) & (n,2) & (n,3) & \cdots & (n.n-1)\end{array} \tag{6.10}$$

Starting from the top corner and proceeding left to right along each row, we can relabel the (i, j) with $k = 1, 2, \ldots, N$, where $N = n(n-1)/2$ by making the substitution

$$k = i + j + \frac{(i-2)(i-3)}{2} - 2, \quad i > j.$$

Let $Z_k = H(X_i, X_j)$ denote the corresponding relabelled random variables.

Theorem 6.2. *Let* $g = \mathbb{E}[H(X_i, X_j)] = \mathbb{E}[Z_k]$ *and for* $k = 1, 2, \ldots, N$ *let*

$$Y_k^{(N)} = \frac{Z_k - g}{(N/2)^{1/4}}.$$

Then

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} Y_k^{(N)} \Rightarrow N(0, \sigma^2).$$

where

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} Var\left(\sum_{k=1}^{N} Y_k^{(N)}\right). \tag{6.11}$$

Proof. We begin by establishing the m-dependence of $Y_k^{(N)}$. Referring to the diagram (6.10), consider the last row of any set of indices (i, j). To reach the first pair of indices which are distinct from all those in that set, we need to go down 2 rows and then go to the last pair in that row. The largest such gap is between $(n-2, 1)$ (relabelled as $(n^2-7n+14)/2$) and $(n, n-1)$ (relabelled as $n(n-1)/2$). Therefore the $Y_k^{(N)}$ are $(3n-8)$-dependent and $m_N \sim O(\sqrt{N})$.

We now proceed to verify the assumptions of Theorem 6.1 for the sequence $Y_k^{(N)}$. In subsequent calculations, we shall often use the fact that

$$N \sim O(n^2/2), \quad n \sim O(\sqrt{2N}). \tag{6.12}$$

(The constant factors of 2 above are important to bear in mind when passing between the single-index labelling $1, \ldots, N$ and the double-index labelling $i, j = 1, \ldots, n$ which is why they are specified in the asymptotics.)

The moment condition (6.4) is obvious; indeed, since $0 \le Y_k^{(N)} \le 1$, we may take δ and Δ_N to be arbitrary, so condition (6.8) can always be satisfied whatever L_N is.

The calculation at (6.2) shows that $V_N^2 \sim O(N)$. Condition (6.6) can therefore be satisfied with $\gamma = 1/2$ and $L_N \sim O(1/N^{1/4})$. The same kind of calculation as in (6.2) shows that $V_{N,k,a}^2 \sim O(Ck^{3/2}/\sqrt{N})$, so condition (6.5) can be satisfied with $\gamma = 1/2$ and $K_N \sim O(1/\sqrt{N})$. But since $1/\sqrt{N} \le 1/N^{1/4}$, this condition can also be satisfied by taking $K_N \sim O(1/N^{1/4})$. Condition (6.7) is now satisfied.

The most delicate condition is (6.9). Since δ can be made arbitrarily large, $\epsilon = 2/\delta$ can be made arbitrarily small. With $\gamma = 1/2$, we have

$$m_N^{1+(1-\gamma)(1+2/\delta)} \sim O\big((\sqrt{N})^{3/2+\epsilon/2}\big) = O(N^{3/4+\epsilon/4}),$$

and condition (6.9) is therefore satisfied.

Since $V_N^2 \sim O(N)$, the limit (6.11) exists and the desired result follows by an application of Theorem 6.1. □

By Theorem 6.2,

$$\frac{1}{\sqrt{N}}\left(\frac{Z_1 + \cdots + Z_N}{(N/2)^{1/4}}\right)$$

has approximate distribution $N((2N)^{1/4}g, \sigma^2)$ for large N. Translated back to the $i, j = 1, \ldots, n$ labelling,

$$\frac{2(Z_1 + \cdots + Z_N)}{n^{3/2}}$$

has approximate $N(g\sqrt{n}, \sigma^2)$ distribution and therefore

$$\widehat{G}_* = \frac{2(Z_1 + \cdots + Z_N)}{n^2}$$

has approximate $N(g, \sigma^2/n)$ distribution. The variance σ^2 can be estimated from data by estimating the individual terms of the sums in (B.1).

The case that $F(0)$ is unknown and therefore must be estimated from data is more complicated and potentially problematic. For a start, $\widehat{G}_*$ is no longer unbiased although it remains consistent.

The details of this case are given in Appendix B because the calculations involved are again rather tedious and messy. The main difficulty here is that the limiting distribution of $\widehat{G}_*$ has infinite variance. The situation here thus has certain similarities to that of trying to estimate the Gini index of a distribution with infinite variance, as discussed in Fontanari et al. (2018). Nevertheless, approximate confidence intervals can be calculated for the estimator $\widehat{G}_*$. Moreover, the width of the confidence interval is $O(1/\sqrt{n})$ so by increasing the sample size n, it should be possible to increase the confidence level or narrow the confidence interval more quickly than for the estimator $\widehat{G}_0$ at (5.1), whose asymptotic distribution is fat-tailed (see Fontanari et al. (2018)).

7. Conclusion

We have defined a class of inequality indices based on bounded functions of the ratio of values in the underlying distribution, in contrast to the classical Gini index which is based on absolute differences of the values. The use of bounded functions results in a class of inequality indices which are continuous with respect to setwise and total variation convergence of probability measures, and which do not require the underlying distribution to possess any finite moments. These indices can therefore be readily applied to heavy-tailed distributions. Although the continuity of our proposed new indices means that they cannot satisfy the Pigou–Dalton principle in general, they do satisfy a weaker form of the Pigou–Dalton principle as well as the concentration principle.

In most cases (when the underlying distribution does not have an atom at 0), these indices are easy to estimate reliably and the usual methods of statistical inference can be applied because they are asymptotically normal. Finally, in situations where the standard Gini index works well, the new indices proposed here behave in very similar ways as the Gini index.

Finally, this paper raises some further questions which remain unanswered and we end by listing some of these which deserve investigation in future work.

1. We saw in Section 5 that our new indices behave in very similar ways as the Gini index. This raises the question of coherence between indices: for example whether $G_0(F_1) \le G_0(F_2)$ implies $G_*(F_1) \le G_*(F_2)$ or vice versa. This question is closely related to that of whether G_* is coherent with respect to certain partial orderings of distributions, such as Lorenz dominance or stochastic dominance (see for example Muliere and Scarsini (1989), Yitzhaki (1982) and Zheng (2021)), which in turn is related to the Pigou–Dalton principle or other similar transfer principles.
2. The decomposability properties of G_* and their implications merit more detailed investigation. For example, even though G_* does not admit decompositions of the form considered in Shorrocks (1980) and Foster and Shneyerov (1999), does it nevertheless admit a decomposition whose between-groups component is invariant under within-groups transfers?
3. We have not considered to what extent can this new class of indices be extended to multidimensional distributions.

Acknowledgments

I would like to express my gratitude to the referees whose many helpful and insightful comments have immeasurably improved and strengthened this paper from its earlier drafts.

Appendix A. Proof of Lemma 1.3

The "if" parts of all three statements are obvious from Definition 1.2 and Lemma 1.1. For the "only" parts, suppose ϕ is unbounded and without loss of generality, we may assume that it is unbounded from above. Therefore there exists a sequence a_n such that $\phi(a_n) \to \infty$. Next, choose some $b \in \mathbb{R}^d$ for which $|\phi(b)| < \infty$ and consider the following sequence of probability measures:

$$P_n = \left(1 - \frac{1}{\phi(a_n)}\right)\delta_b + \frac{1}{\phi(a_n)}\delta_{a_n}.$$

Then $P_n \to_{\text{tv}} \delta_b$ and hence also setwise and weakly. However $\Phi(P_n) \to \phi(b) + 1 \neq \Phi(\delta_b) = \phi(b)$.

Finally, to complete the proof of the "only if" part of Statement 3, suppose that ϕ is bounded but discontinuous at b and let $x_n \to b$. Then $\delta_{x_n} \Rightarrow \delta_b$ but $\Phi(\delta_{x_n}) = \phi(x_n) \nrightarrow \Phi(\delta_b) = \phi(b)$. □

Appendix B. Limiting distribution of $\widehat{G}_*$ in the case $F(0)$ is unknown

We begin with the detailed calculations of the variance of $\widehat{G}_*$.

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{i=1}^n\sum_{j=1}^n H(X_i,X_j)\right)\\
&= \sum_i\sum_j\sum_k\sum_l \operatorname{Cov}(H(X_i,X_k),H(X_j,X_l))\\
&= \sum_i\sum_j\sum_k \operatorname{Cov}(H(X_i,X_k),H(X_j,X_k))\\
&\quad + \sum_i\sum_j\sum_k\sum_{l\neq k} \operatorname{Cov}(H(X_i,X_k),H(X_j,X_l))\\
&= \sum_i\sum_k \operatorname{Var}(H(X_i,X_k)) + \sum_i\sum_{j\neq i}\sum_k \operatorname{Cov}(H(X_i,X_k),H(X_j,X_k))\\
&\quad + \sum_i\sum_k\sum_{l\neq k} \operatorname{Cov}(H(X_i,X_k),H(X_i,X_l))\\
&\quad + \sum_i\sum_{j\neq i}\sum_k\sum_{l\neq k} \operatorname{Cov}(H(X_i,X_k),H(X_j,X_l))\\
&= \sum_i\sum_k \operatorname{Var}(H(X_i,X_k)) + 2\sum_i\sum_{j\neq i}\sum_k \operatorname{Cov}(H(X_i,X_k),H(X_j,X_k))\\
&\quad + \sum_i\sum_{j\neq i}\sum_{k\neq j}\sum_{l\neq k} \operatorname{Cov}(H(X_i,X_k),H(X_j,X_l))\\
&\quad + \sum_i\sum_{j\neq i}\sum_{l\neq j} \operatorname{Cov}(H(X_i,X_j),H(X_j,X_l))\\
&= \sum_i\sum_k \operatorname{Var}(H(X_i,X_k)) + 2\sum_i\sum_{j\neq i}\sum_k \operatorname{Cov}(H(X_i,X_k),H(X_j,X_k))\\
&\quad + \sum_i\sum_{j\neq i}\sum_{l\neq i\neq j} \operatorname{Cov}(H(X_i,X_j),H(X_j,X))\\
&\quad + \sum_i\sum_{j\neq i} \operatorname{Var}(H(X_i,X_j))\\
&= 2\sum_i\sum_k \operatorname{Var}(H(X_i,X_k))\\
&\quad + 3\sum_i\sum_{j\neq i}\sum_k \operatorname{Cov}(H(X_i,X_k),H(X_j,X_k))\\
&\sim O(n^3).
\end{aligned} \tag{B.1}$$

(In the fifth equality above, we have used the fact that, $H(X_i, X_k)$ and $H(X_j, X_l)$ are independent for $i \neq j \neq k \neq l$.)

By the standard Central Limit Theorem, $\bar{F}_0$ as defined in (5.3) has approximate $\mathrm{N}(F(0),\ F(0)(1-F(0))/n)$ distribution. The numerator and denominator of (5.2) are therefore approximately jointly normal and to specify the distribution, we need to calculate their covariance.

Firstly, observe that

$$\operatorname{Cov}(H(X_i,X_j),\mathbf{1}_{\{X_i=0\}}) = F(0)(1-g).$$

Next,

$$\begin{aligned}
&\operatorname{Cov}\left(\sum_i\sum_j H(X_i,X_j),\ \sum_k \mathbf{1}_{\{X_K=0\}}\right)\\
&= \operatorname{Cov}\left(\sum_i\sum_{j\neq i} H(X_i,X_j),\ \sum_k \mathbf{1}_{\{X_K=0\}}\right)\\
&= \sum_i\left(\sum_{j\neq i} H(X_i,X_j),\ \mathbf{1}_{\{X_i=0\}}\right)\\
&\qquad + \sum_i \operatorname{Cov}\left(\sum_{j\neq i} H(X_i,X_j),\ \sum_{k\neq i}\mathbf{1}_{\{X_K=0\}}\right)\\
&= \sum_i\left(\sum_{j\neq i} H(X_i,X_j),\ \mathbf{1}_{\{X_i=0\}}\right) + \sum_i\sum_{j\neq i}\operatorname{Cov}(H(X_i,X_j),\mathbf{1}_{\{X_j=0\}})\\
&\qquad + \sum_i\sum_{j\neq i}\sum_{k\neq i\neq j}\operatorname{Cov}(H(X_i,X_j),\mathbf{1}_{\{X_k=0\}})\\
&= \sum_i\left(\sum_{j\neq i} H(X_i,X_j),\ \mathbf{1}_{\{X_i=0\}}\right) + \sum_i\sum_{j\neq i}\operatorname{Cov}(H(X_i,X_j),\mathbf{1}_{\{X_j=0\}})\\
&= 2\sum_i\sum_{j\neq i}\operatorname{Cov}(H(X_i,X_j),\mathbf{1}_{\{X_i=0\}}) = 2n(n-1)F(0)(1-g).
\end{aligned}$$

The distribution of $\widehat{G}_*$ at (5.2) can then be described as follows: let (U, V) be multivariate normal with mean $(g, F(0))$ and covariance matrix

$$\begin{pmatrix} \frac{\sigma^2}{n} & \frac{2(n-1)F(0)(1-g)}{n^2} \\ \frac{2(n-1)F(0)(1-g)}{n^2} & \frac{F(0)(1-F(0))}{n} \end{pmatrix}. \tag{B.2}$$

Then $\widehat{G}_*$ has approximate distribution

$$R := \frac{U}{1-V^2}. \tag{B.3}$$

Unfortunately, this distribution has infinite variance. To see this, firstly note that the distribution function of R is given by

$$\begin{aligned}
\mathbb{P}(R \le r) &= \mathbb{P}(R \le r,\ 1-V^2 > 0) + \mathbb{P}(R \le r,\ 1-V^2 < 0)\\
&= \int_{|v|<1}\int_{-\infty}^{r(1-v^2)} \varphi_{U,V}(u,v)\,du\,dv\\
&\quad + \int_{|v|>1}\int_{r(1-v^2)}^{\infty} \varphi_{U,V}(u,v)\,du\,dv,
\end{aligned}$$

where $\varphi_{U,V}$ denotes the joint density function of (U, V). The density function of R is therefore given by

$$\begin{aligned}
\psi_*(r) &= \int_{|v|<1} (1-v^2)\varphi_{U,V}(r(1-v^2),v)\,dv\\
&\quad + \int_{|v|>1} (v^2-1)\varphi_{U,V}(r(1-v^2),v)\,dv.
\end{aligned} \tag{B.4}$$

Consider the second integral above and make a change of variables $y = v^2 - 1$; then

$$\int_{|v|>1} (v^2 - 1)\varphi_{U,V}(r(1 - v^2), v)\,dv$$
$$= \int_0^\infty \frac{y}{\sqrt{y+1}} \varphi_{U,V}(-ry, \sqrt{y+1})\,dy.$$

Next,

$$\int_{-\infty}^\infty r^2 \int_0^\infty \frac{y}{\sqrt{y+1}} \varphi_{U,V}(-ry, \sqrt{y+1})\,dy\,dr$$
$$= \int_0^\infty \frac{y}{\sqrt{y+1}} \int_{-\infty}^\infty r^2 \varphi_{U,V}(-ry, \sqrt{y+1})\,dr\,dy. \tag{B.5}$$

An explicit formula can be obtained for the inner integral with respect to r above by means of rather messy but elementary calculations. (The density $\varphi_{U,V}$ is an exponential whose argument is quadratic in r and, by completing the square in the exponent, can be expressed in terms of the density of a normal distribution whose mean and variance are functions of y, then use $\mathbb{E}[R^2] = \mathrm{Var}(R) + \mathbb{E}[R]^2$.) The essential point is that

$$\int_{-\infty}^\infty r^2 \varphi_{U,V}(-ry, \sqrt{y+1})\,dr \sim O\left(\frac{1}{y^2}\right), \quad y \to 0$$

and therefore the integral (B.5) diverges at 0 and R has infinite variance.

Nevertheless, approximate confidence intervals can be calculated for the estimator $\widehat{G}_*$. It is easy to obtain $(1 - \alpha)$ and $(1 - \beta)$-confidence intervals of the form

$$U - a \le g \le U + a$$
$$V - b \le F(0) \le V + b,$$

where $a, b \sim O(1/\sqrt{n})$. Combining the above results in an asymmetric confidence integral of the form

$$\frac{U - a}{1 - (v - b)^2} \le G_* \le \frac{U + a}{1 - (V + b)^2} \tag{B.6}$$

whose confidence level can then be calculated using

$$\int_{g-a}^{g+a} \int_{F(0)-b}^{F(0)+b} \varphi_{U,V}(u, v)\,dv\,du.$$

The binomial expansion for $(1-(v \pm b)^2)^{-1}$ shows that the width of the confidence interval (B.6) is $O(1/\sqrt{n})$.

References

Atkinson, A.B., 1970. On the measurement of inequality. J. Econom. Theory 2, 244–263.
Billingsley, P., 1968. Convergence of Probability Measures. John Wiley & Sons.
Chakravarty, S.R., 1988. Extended gini indices of inequality. Internat. Econom. Rev. 29, 147–156.
Ebert, U., 2009. Taking empirical evidence seriously: the principle of concentration and the measurement of welfare and inequality. Soc. Choice Welf. 32, 555–574.
Ebert, U., 2010. The decomposition of inequality reconsidered: weakly decomposable measures. Math. Social Sci. 60, 94–103.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1996. Modelling Extreme Events for Insurance and Finance. Springer.
Feinberg, E.A., Kasyanov, P.O., Zgurovsky, M.Z., 2014. Convergence of probability measures and Markov decision models with incomplete information. Proc. Steklov Inst. Math. 287, 96–117.
Fields, G.S., 1993. Inequality in dual economy models. Econom. J. 103, 1228–1235.
Fontanari, A., Taleb, N.N., Cirillo, P., 2018. Gini estimation under infinite variance. Physica A 502, 256–269.
Foster, J.E., Shneyerov, A.A., 1999. A general class of additively decomposable inequality measures. Econom. Theory 14, 89–111.
Jacka, S.D., Roberts, G.O., 1997. On strong forms of weak convergence. Stochastic Process. Appl. 67, 41–53.
Muliere, P., Scarsini, M., 1989. A note on stochastic dominance and inequality measures. J. Econom. Theory 49, 314–323.
Romano, J.P., Wolf, M., 2000. A more general central limit theorem for *m*-dependent random variables with unbounded *m*. Statist. Probab. Lett. 47, 115–124.
Shorrocks, A.F., 1980. The class of additively decomposable inequality measures. Econometrica 48, 613–625.
Yitzhaki, S., 1982. Stochastic dominance, mean variance, and Gini's mean difference. Am. Econ. Rev. 72, 178–185.
Zheng, B., 2021. Stochastic dominance and decomposable measures of inequality and poverty. J. Public Econ. Theory 23, 228–247.