Article

f-Gintropy: An Entropic Distance Ranking Based on the Gini Index

Tamás Sándor Biró [1,2,*] **, András Telcs** [1,3] **, Máté Józsa** [2] **and Zoltán Néda** [2]

[1] Wigner Research Centre for Physics, Konkoly-Thege M. 29-33, 1121 Budapest, Hungary; telcs.andras@wigner.hu

[2] Department of Physics, Babeş-Bolyai University, Str. M. Kogalniceanu 1, 400084 Cluj-Napoca, Romania; mate.jozsa@ubbcluj.ro (M.J.); zoltan.neda@ubbcluj.ro (Z.N.)

[3] Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Műegyetem rkp. 3, 1111 Budapest, Hungary

[*] Correspondence: biro.tamas@wigner.hu; Tel.: +36-20-435-1283

Abstract: We consider an entropic distance analog quantity based on the density of the Gini index in the Lorenz map, i.e., gintropy. Such a quantity might be used for pairwise mapping and ranking between various countries and regions based on income and wealth inequality. Its generalization to f-gintropy, using a function of the income or wealth value, distinguishes between regional inequalities more sensitively than the original construction.

Keywords: entropy; Gini index; gintropy

1. Introduction

The study of social and economic inequality seems to belong to social sciences and humanities. However, based mostly on economic and econophysics studies, analyses of income and wealth data in terms of mathematical quantities such as the Lorenz curve [1] and the related Gini index [2–4] are trying to grasp an "objective" measure of overall inequality in distributions [5–9].

Given that most interesting distributions have infinite support, the norms on their space are not equivalent. That is why there are numerous measures of distance between distributions. Clearly, the Kullback–Leibler divergence is the most popular one. It is well known that to produce a good estimate of the KL divergence using a limited sample is not easy. That is one of the reasons why several generalization and particular alternatives have been developed. Some are for a specific family of distributions [10,11]; others are better suited to a specific field [12,13]. Our work falls into the latter category, aiming to propose a divergence well adapted to economic or wealth inequalities. The adaptation and generalization of divergences should meet certain criteria The reader can find an overview and new advances in [14,15] in that direction. More recent developments can be found in [16–20], and a recent review in [21].

Certainly, the Gini index is only one among many possible approaches and does not replace the study of the distributions themselves [22]. In this sense, the density of the Gini index in the Lorenz curve—the gintropy—whose precise definition will be given below, includes the total information on the probability density function (PDF) of such distributions.

In an earlier work we introduced this quantity called gintropy [23] and demonstrated its entropy-like properties: for all income PDFs it is non-negative, has a single maximum exactly at the average value and shows overall the corresponding convexity in terms of the cumulative rich population as well as in terms of the cumulative wealth. Based on this property, we propose that gintropy may better reflect the inequality concentrated on the middle class with average income for which much more statistical data are available than for the richest segment of the income and wealth distributions.

In this paper we attempt to generalize this concept to a gintropic distance measure between two PDFs, making it possible to draw a distance graph between countries with varying inequality. By doing so we also investigate the generalization potential in replacing the obvious variable, the income value x, with its monotonic function, $f(x)$. That provides the analyzers with some freedom in using differently weighted *f-gintropy* concepts, e.g., using x^2 or any other quantity behaving coherently with the original one but possibly magnifying some effects. Indeed, we observe a finer distinguishing power by using functions of x, describe in a later section.

This article is constructed as follows: First, we review the definitions of gintropy and f-gintropy, controlling their entropy-like properties. Then we construct the f-gintropic divergence measure in analogy to the entropic divergence and study its basic mathematical properties. Finally, based on available data on regional and national income distributions, we develop an inequality distance map. We suggest using these maps as a "first sight" tool for judging the real effect of inequalities on the population parts near the average.

2. Gintropy, f-Gintropy and Gintropic Divergence

The Gini index is a number between zero and one characterizing the inequality inherent in income and wealth distributions. Its definition,

$$G \equiv \frac{\langle |x - y| \rangle}{\langle x + y \rangle} = \frac{\int_0^\infty \int_0^\infty |x - y| P(x) P(y) dx dy}{\int_0^\infty \int_0^\infty (x + y) P(x) P(y) dx dy}, \tag{1}$$

can be used for deriving expressions of G in terms of cumulative distributions of the population and the wealth or income possessed by them. These cumulatives,

$$\overline{C}(x) \equiv \int_x^\infty P(x) dx, \qquad \overline{F}(x) \equiv \frac{1}{\langle x \rangle} \int_x^\infty x P(x) dx, \tag{2}$$

plotted against each other present the Lorenz curve in the $\overline{F} - \overline{C}$ plane. This curve is restricted to the unit square, since $\overline{C}(0) = 1$, $\overline{F}(0) = 1$ are the total normalized values, while $\overline{C}(\infty) = 0$ and $\overline{F}(\infty) = 0$ both vanish.

The gintropy, $\sigma \equiv \overline{F} - \overline{C}$, occurs as a density of the Gini index in this map and shows entropy-like features. Based on this, to any probability density function (PDF), such as $P(x)$, one may construct an individual gintropy $\sigma(x)$. It is interesting to note that the half of the Gini index also can be viewed as the integral of gintropy over any of the cumulatives:

$$G/2 = \int_0^1 \sigma d\overline{C} = \int_0^1 \sigma d\overline{F}. \tag{3}$$

As such the $G/2$ value represents the area between the Lorenz curve and the diagonal in the $\overline{C} - \overline{F}$ plane.

The function $\sigma(\overline{C})$ is non-negative, and shows a definite convexity. Furthermore $\sigma(x)$ is maximal at $x = \langle x \rangle$. Indeed, for several common PDFs $\sigma(\overline{C})$, it resembles entropy formulas suggested so far. For example, for an exponential PDF the classical formula, $\sigma = -\overline{C} \ln \overline{C}$, arises; for a Pareto distribution $P(x)$, we obtain a Tsallis q-entropy formula $\sigma(\overline{C}) = (\overline{C}^q - \overline{C})/(1 - q)$. The particular value, $q = 2$, resembles the entanglement entropy, $S = \text{Tr}(\rho - \rho^2)$ with ρ being the density operator, used in some quantum computing calculations. These analogies are elaborated in [23].

In the present paper we attempt to generalize this construction to a possible use of another quantity's cumulative tail more integral than the original variable, x. Moreover, based on this, we construct an entropic divergence measure for the purpose of using it as

an "inequality distance" between PDFs based on their generalized Lorenz curve behavior, comprised in the quantity *f-gintropy*.

Let $P(x)$ be an absolute continuous PDF over $[0, \infty)$. To begin with, we consider only P and do not use any subscript related to it. Later, when considering two different PDFs and constructing their f-gintropic distance measure, we shall retain such indices.

Let f be a monotonic, non-negative differentiable function with $f(0) = 0$. Assume that the expectation value,

$$\langle f \rangle = \int_0^\infty f(x)P(x)dx < \infty, \tag{4}$$

exists. Let the truncated f-moment be defined as

$$\overline{F}_f(x) \equiv \frac{1}{\langle f \rangle} \int_x^\infty f(y)P(y)dy \tag{5}$$

and we use the original tail-cumulative $\overline{C}(x)$ defined in Equation (2). Inverting the above definitions, one notes that $d\overline{C}/dx = -P(x)$ and $d\overline{F}_f/dx = -f(x)P(x)/\langle f \rangle$. Hence, also $d\overline{F}_f/d\overline{C} = f(x)/\langle f \rangle$.

The f-Gini index is constructed based on the geometry of the Lorenz curve. First, one easily proves that

$$\int_0^1 \overline{F}_f d\overline{C} + \int_0^1 \overline{C} d\overline{F}_f = 1. \tag{6}$$

Second, the corresponding Gini index is defined as the difference between these two integrals

$$G_f \equiv \int_0^1 \overline{F}_f d\overline{C} - \int_0^1 \overline{C} d\overline{F}_f. \tag{7}$$

Let us first prove the statement about the sum in Equation (6). The first integral can be expanded as

$$\int_0^1 \overline{F}_f d\overline{C} = \int_0^\infty \overline{F}_f(x)P(x)dx = \int_0^\infty dx P(x) \int_x^\infty dy P(y)\frac{f(y)}{\langle f \rangle}, \tag{8}$$

while the second similarily as

$$\int_0^1 \overline{C} d\overline{F}_f = \int_0^\infty \overline{C}(x)\frac{f(x)}{\langle f \rangle}P(x)dx = \int_0^\infty dx P(x)\frac{f(x)}{\langle f \rangle} \int_x^\infty P(y)dy. \tag{9}$$

The second integral can be rewritten by exchanging the dummy integration variables x and y among each other,

$$\int_0^1 \overline{C} d\overline{F}_f = \int_0^\infty dy P(y)\frac{f(y)}{\langle f \rangle} \int_y^\infty P(x)dx. \tag{10}$$

Then we interchange the order of integration to arrive at

$$\int_0^1 \overline{C} d\overline{F}_f = \int_0^\infty dx P(x) \int_0^x dy P(y)\frac{f(y)}{\langle f \rangle}. \tag{11}$$

Indeed, the sum of Equations (8) and (11) becomes

$$\int_0^1 \overline{F}_f d\overline{C} + \int_0^1 \overline{C} d\overline{F}_f = \int_0^\infty dx P(x) \left[\int_0^x dy P(y) \frac{f(y)}{\langle f \rangle} + \int_x^\infty dy P(y) \frac{f(y)}{\langle f \rangle} \right]. \tag{12}$$

Here, obviously the two integrals inside the square bracket can be unified in a single integral proving that

$$\int_0^1 \overline{F}_f d\overline{C} + \int_0^1 \overline{C} d\overline{F}_f = \int_0^\infty dx P(x) \int_0^\infty dy P(y) \frac{f(y)}{\langle f \rangle} = \int_0^\infty dx P(x) \frac{\langle f \rangle}{\langle f \rangle} = 1. \tag{13}$$

Inspecting a general Lorenz curve drawn in the unit square on the $\overline{C} - \overline{F}_f$ plane, it is obvious that the area between the curve and diagonal represents the half of the Gini index, defined in Equation (7), as the difference of integrals.

We are interested further in its equivalent forms. We would like to view G_f also as an expectation value of something. We utilize the forms (8) and (9) to write

$$G_f = \int_0^\infty dx P(x) \int_x^\infty dy P(y) \frac{f(y) - f(x)}{\langle f \rangle}. \tag{14}$$

Exchanging the integration variables again and reordering the integrations, an equivalent expression emerges

$$G_f = \int_0^\infty dx P(x) \int_0^x dy P(y) \frac{f(x) - f(y)}{\langle f \rangle}. \tag{15}$$

Now we fix a few requirements for the function $f(x)$. To map $\overline{F}_f$ between zero and one $f(x), \geq 0$ is necessary. To interpret the f-Gini index as an expectation of an absolute value of difference, the strict monotonity property, $f(x) > f(y)$ for $x > y$ and equivalently $f(x) < f(y)$ for $x < y$ is required. Fulfilling these requirements, one obtains that G_f is also the half of the sum of Equations (14) and (15):

$$G_f = \frac{\langle |f(x) - f(y)| \rangle}{\langle |f(x) + f(y)| \rangle}. \tag{16}$$

This agrees with the original definition for $f(x) = x$ and follows its form ready for data evaluation.

Based on the above, it is natural to introduce f-gintropy as

$$\sigma_f \equiv \overline{F}_f - \overline{C}. \tag{17}$$

This quantity has the following properties, based on the already discussed restrictions on $f(x)$, i.e., $f(x) \geq 0$ and $f'(x) > 0$:

- It is never negative, $\sigma_f \geq 0$.
- It is zero for $x = 0$ and $x = \infty$ only, or at $\overline{C} = 0$ and $\overline{C} = 1$.
- It is maximal at the $x = x_m$ value, fulfilling $f(x_m) = \langle f \rangle$.
- It has a single maximum only, either as a function of x or $\overline{C}$ or $\overline{F}_f$.
- The f-gintropy, σ_f, like the entropy, is everywhere concave on $\overline{C}$ and $\overline{F}_f$.

We give a short analysis of the above statements as follows.

$$\sigma_f(x) = \frac{1}{\langle f \rangle} \int_x^\infty dy P(y)[f(y) - \langle f \rangle] \tag{18}$$

is positive if $f(x) > \langle f \rangle$, since for all $y > x$ in the above integral one has $f(y) > f(x) > \langle f \rangle$. For the opposite choice of x, when $f(x) \leq \langle f \rangle$, we rewrite the gintropy expression as

$$\sigma_f(x) = \left[1 - \int_0^x dy P(y) \frac{f(y)}{\langle f \rangle} \right] - \left[1 - \int_0^x dy P(y) \right] = \frac{1}{\langle f \rangle} \int_0^x dy P(y) [\langle f \rangle - f(y)]. \quad (19)$$

Again, due to the monotonity of f for all $y \leq x$ in the integration range, we have $f(y) \leq f(x) \leq \langle f \rangle$. That finishes the proof of $\sigma_f \geq 0$. It is also obvious from the above expressions that equality occurs only if using $x = 0$ or $x \to \infty$.

Let us turn now to the discussion of concavity. For that purpose we calculate the first and second derivatives of gintropy w.r.s.p to $\overline{C}$ and $\overline{F}_f$. These are obtained in turn from the x-derivative,

$$\frac{d\sigma_f}{dx} = \frac{d\overline{F}_f}{dx} - \frac{d\overline{C}}{dx} = \left(1 - \frac{f(x)}{\langle f \rangle} \right) P(x). \quad (20)$$

This quantity changes sign exactly where $f(x_m) = \langle f \rangle$ and nowhere else.

The $\overline{C}$- and $\overline{F}_f$-derivatives can be obtained using the $d\overline{C}/dx$ and $d\overline{F}_f/dx$ values shown earlier in the text. One calculates

$$\frac{d\sigma_f}{d\overline{C}} = \frac{f(x)}{\langle f \rangle} - 1, \quad \text{and} \quad \frac{d^2\sigma_f}{d\overline{C}^2} = -\frac{f'(x)}{P(x)\langle f \rangle}. \quad (21)$$

Due to the strict monotonity of f, we have $f'(x) > 0$, and therefore, the second derivative of f-gintropy is always negative. We similarly arrive at the same conclusion in terms of F_f:

$$\frac{d\sigma_f}{d\overline{F}_f} = 1 - \frac{\langle f \rangle}{f(x)}, \quad \text{and} \quad \frac{d^2\sigma_f}{d\overline{F}_f^2} = -\frac{f'(x)\langle f \rangle}{P(x)f^3(x)} < 0. \quad (22)$$

Let us recall that the Gini index and its generalization can be written as an integral over the gintropy, σ_f.

We built the f-gintropic divergence measure following the construction pattern of the Kullback–Leibler divergence. It is also in line with the idea how Csiszár's f-divergence has been introduced [24,25]. For recent advances of the f-divergence and its estimate see, e.g., [26,27]. The main difference is that while the *f-entropy* due to Csiszár replaces the logarithm of the probability with a general function in the entropy formula, we generalize the tail cut integral of the income to a function of it. The latter function will be restricted, as discussed later.

Instead of using the PDFs, we suggest applying a similar formula to the gintropies. Since the original gintropy function is not normalized, we use its normalized version

$$\widehat{\sigma}(x) \equiv \frac{\sigma(x)}{\langle \sigma \rangle} \quad (23)$$

with

$$\langle \sigma \rangle = \int_0^\infty \sigma(x) P(x) dx = \int_0^1 \sigma d\overline{C} = G/2. \quad (24)$$

Then $\widehat{\sigma}(\overline{C})$ behaves as a PDF mathematically: its integral over the possible $\overline{C}$-range is unity, and it is everywhere non-negative. Analogous to the definition of classical entropic divergences, we propose a Kullback–Leibler type Gini divergence between the distributions $P_1(x)$ and $P_2(x)$ as

$$\text{KLGD}_f(P_1 || P_2) = \int_0^1 \widehat{\sigma}^{(2)} \log\left(\frac{\widehat{\sigma}^{(2)}}{\widehat{\sigma}^{(1)}} \right) d\overline{C}. \quad (25)$$
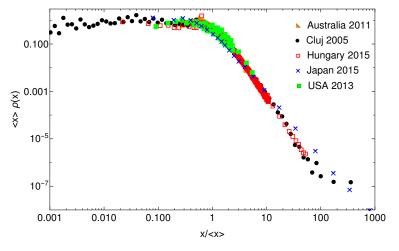
It is straightforward to generalize further by using a general convex function s instead of the logarithm with $s(1) = 0$ and $s'' < 0$. One immediately realizes that $KLGD_{s,f}(P_1||P_2) = 0$ iff $P_1 = P_2$ almost surely and $KLGD_{s,f}(P_1||P_2) \geq 0$ in general, based on the convexity inherent in the definition.

Analogous to the definition of mutual information, the mutual gintropy, MG, can also be introduced as the gintropic divergence between the joint distribution and the product of the subdistributions:

$$MG(P_1; P_2) = KLGD_{s,f}(P_1 \otimes P_2 || P_{1,2}). \tag{26}$$

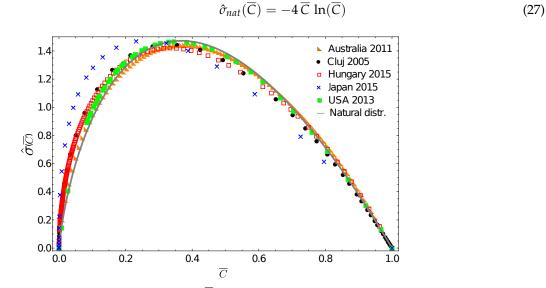## 3. Example: Gintropic Distance Ranking of Income Data

We will exemplify the use of gintropic distance reconsidering the income distribution data studied in [28]. Data for Australia (2011), Hungary (2015), Japan (2015), USA (2013) and Cluj county, Romania (2005) are used. These are the same data as the ones used in [28], and they were selected for statistical studies due to their free availability and higher resolutions in the income intervals, allowing for a precise construction of relevant quantities. The data for Cluj county are exhaustive from a social security database, containing the income of all registered employee from that region [29]. To use a common scale and to collapse the distributions as much as possible, we consider for each case the normalized income, i.e., income relative to the average value. The probability density functions of these distributions for the whole normalized income ranges are plotted in Figure 1. They look very similar on log-log scale, and only the data for Japan seems to have a different scaling in the high income region. On such a log-log representation, the probability density functions of normalized income for Hungary and Cluj county (Romania) seem to collapse perfectly, which is definitely a consequence of the closely related economic history of these two countries.

Figure 1. Probability density function for the distributions of the normalized income for some countries and geographical regions. The income for each region is normalized to the respective average value. Please note that we use log-log scales.

One can learn from population biology that in many cases the use of the probability density function is not the best methodology to illustrate the relevant differences in abundance [30] . For heavy-tailed distributions, the generally used log-log scales (such as in Figure 1) is appropriate to illustrate scaling, but it does not offer relevant information for those parts of the distribution where the majority of the individuals are. Therefore in illustrating these differences, instead of the mathematically well-defined probability density function, a special frequency histogram, the Preston plot [31], is used. When illustrating social inequalities in form of income, we are in a similar situation. As an alternative to the probability density, from the income distribution data one can construct a normalized gintropy $\hat{\sigma}(\overline{C}) = \sigma(\overline{C})/\langle \sigma \rangle$ by estimating numerically both cumulative functions, $\overline{C}(x)$ and $\overline{F}(x)$, from the data themselves. In contrast to the probability density function, this quantity

will peak in the vicinity of the average income, i.e., characterizing the income regions where the majority of the population is to be found. Figure 2 presents these gintropy functions in comparison to the one expected for the natural (exponential) distribution:

$$\hat{\sigma}_{nat}(\overline{C}) = -4\,\overline{C}\,\ln(\overline{C}) \tag{27}$$

Figure 2. Normalized gintropy $\hat{\sigma}(\overline{C})$ calculated from the income distribution data in comparison with the one expected for the natural distribution (27).

Figure 2 again reveals the fact that the income distribution for Japan seems to be very different, and surprisingly, we see that the gintropy curve for Australia is the closest to the one for the natural distribution. The gintropies, $\hat{\sigma}(\overline{C})$, for Hungary and Cluj county are also quite close to that of the natural distribution. All these results confirm again what we have known for a long time: both income and wealth distributions tend to have an exponential shape when restricted to the middle class of a society [32,33].

Gintropic distances can be calculated using the well-known Kullback–Leibler divergences for the $\hat{\sigma}(\overline{C})$ functions.

$$\text{KLGD}(P_i||P_j) = \int_0^1 \hat{\sigma}_i(\overline{C}) \log\left(\frac{\hat{\sigma}_i(\overline{C})}{\hat{\sigma}_j(\overline{C})}\right) d\overline{C} \tag{28}$$

Approximating the above integral from the interpolated experimental data for $\hat{\sigma}(\overline{C})$, we thus determine a quantity that characterizes the gintropic differences between income distributions among different countries. In Table 1 we indicate these values and also consider the distance relative to the normalized gintropy which is characteristic to the natural distribution $\hat{\sigma}_{nat}(\overline{C})$. This table again suggests that while focusing on the relevant (larger) part of the society, the income distribution in Australia and the USA are close to the expected natural distribution. On the other hand, Hungary, Romania and Japan seem to be different: an interesting and unexpected result. Such a methodology based on the gintropy instead of the commonly used probability density function could definitely be useful in cases where one prefers to group countries in clusters according to their most abundant income categories.

One can now go further and test whether the $\hat{\sigma}(\overline{C})$ gintropy, associated to $f(x) = x$, can be fitted with the gintropy for the Tsallis–Pareto distribution with $\langle x \rangle = 1$:

$$P(x) = \frac{1}{q}\left(1 + \frac{1-q}{q}x\right)^{-\frac{2-q}{1-q}} \tag{29}$$
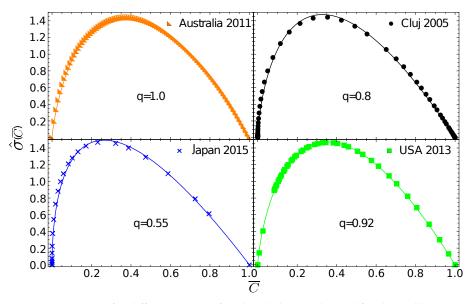
The corresponding normalized gintropy is proportional to the Tsallis q-entropy form [23]:

$$\hat{\sigma}_q(\overline{C}) = 2\,\frac{1+q}{1-q}(\overline{C}^q - \overline{C}). \tag{30}$$

In Figure 3 we fit the experimental gintropy curves with the one given in Equation (30). The best fit parameter, $q = 1$, for Australia again suggests that in this case the natural distribution is relevant since Equation (30) in the limit $q = 1$ leads to Equation (27). For Cluj county and Hungary, the best fit is obtained with $q = 0.8$, for the USA we get $q = 0.92$, and for Japan we obtain $q = 0.55$.

Table 1. Gintropic distances for income distributions calculated by using the (28) gintropic Kullback–Leibler divergences. We also indicate this distance relative to the natural distribution.

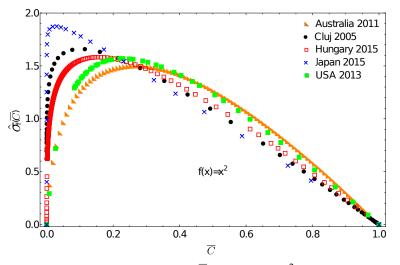| $\times 10^{-3}$	**Natural**	Australia	USA	Cluj	Hungary	Japan
Natural	0	0	0.4	2.9	2.9	17
Australia	0	0	0.4	2.9	2.9	17
USA	0.4	0.4	0	1.1	1.1	13
Cluj	3	3	1.1	0	0	6.4
Hungary	3	3	1.1	0	0	6.4
Japan	19	19	14	6.8	6.8	0

Figure 3. Gintropy for different regions fitted with the one derived for the Tsallis–Pareto distribution $\hat{\sigma}_q(\overline{C})$ (30). In the figures we illustrate the best best fit and also give the best-fit parameter, q. There is no separate panel for Hungary since the experimental gintropy for Hungary and Cluj are very close, as already seen in Figure 2.

One can now use the above discussed income distributions to exemplify the application potential of the f-gintropy. We choose a simple convex function $f(x) = x^2$ and construct the corresponding $\hat{\sigma}_f(\overline{C})$ normalized f-gintropy for all the income distributions. Such a choice is justified if: (i) instead of income one considers some other socioeconomic metrics that depend on its square or (ii) if we intend to amplify differences at large income values. In this latter case, any convex-shaped $f(x)$ function can serve this purpose.

The obtained curves are plotted in Figure 4. A first immediate consequence of using the f-gintropy instead of the usual gintropy is that the $\hat{\sigma}_f(\overline{C})$ curves become more strongly separated. Japan and Australia are again the two extremes. In such a representation, one can

distinguish between the income distributions in Hungary and Cluj county that appeared to be very close before. A Kullback–Leibler distance matrix, similar to the one shown in Table 1, is now constructed for the f-gintropy with the $f(x) = x^2$ choice. To proceed, we used a second order interpolation method to estimate the Kullback–Leibler divergence for the f-gintropy in the $\hat{\sigma}_f(\overline{C})$ representation. The results are given in Table 2. According to the values from Table 2, Australia and the USA form a clear cluster, and similarly Hungary with Cluj county and Japan continues to be in a separate cluster. While the similarities between Australia and the USA together with the ones between Hungary and Romania are easy to interpret, Japan's position seems to be more surprising. These results might confirm some earlier hypotheses according to which the Japanese taxation and redistribution system is perfectly balanced, neither highly redistributive nor too capitalistic [34]. Interestingly, their income redistribution policies seem to be closer to those in former socialist countries rather than the ones with consecrated free market economies. In such a view, the results obtained here make sense.

Figure 4. Normalized f-gintropy $\hat{\sigma}_f(\overline{C})$ with $f(x) = x^2$ calculated from the income distribution data. Note the more evident separation of the studied geographical regions.

Table 2. f-gintropic distances for income distributions calculated using the (28) generalized Kullback–Leibler divergences for $f(x) = x^2$.

$\times 10^{-3}$	Australia	USA	Cluj	Hungary	Japan
Australia	0	1.8	36	19	62
USA	1.7	0	27	13	50
Cluj	48	39	0	3.8	4
Hungary	22	16	3.5	0	14
Japan	86	74	4.3	17	0

4. Conclusions

In conclusion, we have constructed a generalization of the usual Lorenz-curve based entropy similar to the concept, gintropy, developed by us earlier. This generalization involves weighting functions more general than the original income values, replacing x by $f(x)$, the only requirement being its monotonicity and non-negativity. Entropic properties of such a generalization were demonstrated.

A Kullback–Leibler type entropic divergence was used to inspect clusterization among income distribution PDFs stemming from different regional groups. In the example of detailed data from five different regions (Japan, Australia, USA, Hungary and Cluj) we have demonstrated that: (i) the gintropy emphasizes data near the average income, and (ii) the

use of weight functions rising steeper than x enhances differences not otherwise seen. The former result is practical, since data about extreme high incomes and wealth are usually insufficient or unreliable. The latter may help resolve close looking data in the future, as a change of viewing angle can resolve contours appearing to be the same while they are not.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lorenz, M.O. Methods of measuring the concentration of wealth. *Publ. Am. Stat. Assoc.* **1905**, *9*, 209–219. [CrossRef]
2. Gini, C. Il diverso accrescimento delle classi sociali e la concentrazione della richezza. *G. Econ.* **1909**, *20*, 27–83.
3. Gini, C. On the characteristics of Italian statistics. *J. R. Stat. Soc. Ser. A* **1965**, *128*, 89–109. [CrossRef]
4. Ceriani, L.; Verme, P. The origins of the Gini index: Extracts from Variabilità e Mutabilità by Corrado Gini. *J. Econ. Inequal.* **2012**, *10*, 421–443. [CrossRef]
5. Atkinson, T. On the measurement of inequality. *J. Econ. Theory* **1970**, *2*, 244–263. [CrossRef]
6. Yitzhaki, S. More than a dozen alternative ways of spelling Gini. *Res. Econ. Inequal.* **1998**, *8*, 13–30.
7. Giorgi, G.M.; Nadarajah, S. Bonferroni and Gini indices for various parametric families of distributions. *Metron Int. J. Stat.* **2010**, *LXVIII*, 23–46. [CrossRef]
8. Gosh, A.; Chattapadhyay, N.; Chakrabati, B.K. Inequality in societies, academic institutions and science journals: Gini and k-indices. *Phys. A* **2014**, *410*, 30–34. [CrossRef]
9. Koutsoyiannis, D.; Sargentis, G.-F. Entropy and Wealth. *Entropy* **2021**, *23*, 1356. [CrossRef]
10. Bouhlel, N.; Dziri, A. Kullback–Leibler divergence between multivariate generalized gaussian distributions. *IEEE Signal Process. Lett.* **2019**, *26*, 1021–1025. [CrossRef]
11. Gochhayat, S.P.; Shetty, S.; Mukkamala, R.; Foytik, P.; Kamhoua, G.A.; Njilla, L. Measuring Decentrality in Blockchain Based Systems. *IEEE Access* **2020**, *8*, 178372–178390. [CrossRef]
12. Hecksher, M.D.; do Nascimento Silva, P.L.; Corseuil, C.H.L. Dominance of the Richest in Brazilian Income Inequality Measured with J-divergence (1981–2015). In Proceedings of the 2nd Regional Statistics Conference on "Enhancing Statistics, Prospering Human Life", Bali, Indonesia, 20–24 March 2017; Bank Indonesia: Jakarta, Indonesia 2017; pp. 464–468.
13. Yu, X. A Unified Entropic Pricing Framework of Option: Using Cressie-Read Family of Divergences. *N. Am. J. Econ. Financ.* **2021**, *58* 101495. [CrossRef]
14. Yamano, T. A generalization of the Kullback–Leibler divergence and its properties. *J. Math. Phys.* **2009**, *50*, 043302. [CrossRef]
15. Vigelis, R.F.; de Andrade, L.H.; Cavalcante, C.C. Conditions for the existence of a generalization of Rényi divergence. *Phys. A Stat. Mech. Its Appl.* **2020**, *558*, 124953. [CrossRef]
16. Esposito, A.R.; Gastpar, M.; Issa, I. Robust Generalization via α-Mutual Information. *arXiv* **2020**, arXiv:2001.06399.
17. Kim, T.; Oh, J.; Kim, N.; Cho, S.; Yun, S.Y. Comparing Kullback-Leibler Divergence and Mean Squared Error Loss in Knowledge Distillation. *arXiv* **2021**, arXiv:2105.08919.
18. Kimura, M.; Hino, H. α-Geodesical Skew Divergence. *Entropy* **2021**, *23*, 528. [CrossRef]
19. Ghosh, A.; Basu, A. A Scale-Invariant Generalization of the Rényi Entropy, Associated Divergences and Their Optimizations Under Tsallis' Nonextensive Framework. *IEEE Trans. Inf. Theory* **2021**, *67*, 2141–2161. [CrossRef]

20. Singhal, A.; Sharma, D.K. Generalization of F-Divergence Measures for Probability Distributions with Associated Utilities. *Solid State Technol.* **2021**, *64*, 5525–5531.
21. Giorgi, G.M.; Gigliarano, C. The Gini concentration in dex: A review of the inference literature. *J. Econ. Surv.* **2016**, *31*, 1130–1148. [CrossRef]
22. Inoue, J.; Ghosh, A.; Chatterjee, A.; Chakrabarti, B.K. Measuring social inequality with quantitative methodology: Analytical estimates and empirical data analysis by Gini and k-indices. *Phys. A* **2015**, *429*, 184–204. [CrossRef]
23. Biró, T.; Néda, Z. Gintropy: A Gini Index Based Generalization of Entropy. *Entropy* **2020**, *22*, 879–892. [CrossRef] [PubMed]
24. Csiszár, I. Information-type measures of difference of probability distributions and indirect observation. *Stud. Sci. Math. Hung.* **1967**, *2*, 229–318.
25. Csiszár, I. Information measures: A critical survey. In Proceedings of the Transactions of the Seventh Prague Conference on Information Theory. Statistical Decision Functions, Random Processes, Prague Czech Republic, 18–27 August 1974; pp. 73–86.
26. Bhatia, P.K.; Singh, S. On a new Csiszar's f-divergence measure. *Cybern. Inf. Technol.* **2013**, *13*, 43–57. [CrossRef]
27. Anastassiou, G.A. Generalized Hilfer Fractional Approximation of Csiszar's f-Divergence. In *Unification of Fractional Calculi with Applications*; Springer: Cham, Switzerland, 2022; pp. 115–131.
28. Néda, Z.; Gere, I.; Biró, T.S.; Tóth, G.; Derzsy, N. Scaling in income inequalities and its dynamical origin. *Phys. A Stat. Mech. Its Appl.* **2020**, *549*, 124491. [CrossRef]
29. Derzsy, N.; Néda, Z.; Santos, M.A. Income distribution patterns from a complete social security database. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 5611–5619. [CrossRef]
30. Hubbell, S.P. *The Unified Neutral Theory of Biodiversity and Biogeography*; Princeton University Press: Princeton, NJ, USA, 2001.
31. Preston, F.W. The Commonness and Rarity of Species. *Ecology* **1948**, *29*, 254–283. [CrossRef]
32. Yakovenko V.M.; Rosser, J.B., Jr. Colloqium: Statistical mechanics of money, wealth, and income. *Rev. Mod. Phys.* **2009**, *81*, 1703. [CrossRef]
33. Chakraborti, A.; Chatterjee, A.; Chakrabarti, B.; Chakravarty, S.R. *Econophysics of Income and Wealth Distributions*; Cambridge University Press: Cambridge, UK, 2013.
34. Dewit, A.; Steinmo, S. The Political Economy of Taxes and Redistribution in Japan. *Soc. Sci. Jpn. J.* **2002**, *5*, 159–178. [CrossRef]