Gintropy: Gini Index Based Generalization of Entropy

Tamás S. Biró [1,2] and Zoltán Néda [3,*]

[1] Wigner Research Centre for Physics, 1121 Budapest, Hungary; biro.tamas@wigner.hu
[2] Complexity Science Hub, 1080 Vienna, Austria
[3] Department of Physics, Babeş-Bolyai University, 400084 Cluj-Napoca, Romania
[*] Correspondence: zneda@phys.ubbcluj.ro

Abstract: Entropy is being used in physics, mathematics, informatics and in related areas to describe equilibration, dissipation, maximal probability states and optimal compression of information. The Gini index, on the other hand, is an established measure for social and economical inequalities in a society. In this paper, we explore the mathematical similarities and connections in these two quantities and introduce a new measure that is capable of connecting these two at an interesting analogy level. This supports the idea that a generalization of the Gibbs–Boltzmann–Shannon entropy, based on a transformation of the Lorenz curve, can properly serve in quantifying different aspects of complexity in socio- and econo-physics.

Keywords: entropy; Gini index; socio-economic inequalities; econophysics

1. Introduction

1.1. Motivation

Many researchers use entropy as an appropriate measure for quantifying complexity or the inequality level in a complex system. There is an overwhelming choice in generalized entropy formulas, some of them satisfying more of the basic axioms than the others [1]. The classical Boltzmann–Gibbs–Shannon formula is often used in economic and social studies without elaborating too much on the conditions under which it is an appropriate thermodynamic function. Most prominently, the additivity of entropy upon the factorization of probabilities is, as a rule, not tested and therefore the use of entropy remains at the level of a crude analogy. Using the Tsallis- or Rényi entropy formula [2] is also not a sufficient choice. Although a free parameter in this entropy provides more flexibility in processing and interpreting statistical data and generalizing the additivity, there is no basic reason to not use yet another formula that satisfies the basic physical requirements for the entropy.

On the other hand, the most popular way for quantifying the inequality level in a socio-economic system is to use the Gini index, introduced for the first time by the economist Corrado Gini [3]. This measure provides a simple method of quantifying the deviation from a uniform distribution, and it is not a quantity borrowed by a simple analogy from thermodynamics. It also has the advantage that its value is a number in the $[0, 1]$ interval, like an order parameter. The Gini index is 0 when all members of the investigated society are equal in the relevant quantity, and it is 1 if one member is monopolizing the whole of the available resources. The Gini index can be determined experimentally either graphically by constructing the Lorentz curve [4], or by the simple formula

$$G = \frac{1}{\langle x \rangle} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} |x_i - x_j|}{2N^2}, \tag{1}$$

where x_i is the relevant quantity for element i, and $\langle x \rangle$ is its average value for the whole system with N elements. While the Gini index is traditionally used to measure wealth-, income- or other inequality, the entropy is a concept stemming from physics and mathematics and is applied to understand, describe and construct optimal or equilibrium distributions. At first glance, these two termini show no reason to be connected. However, in recent publications, it has been observed that the Gini index and the total Shannon entropy of socio-economical models and data show a synergic behavior [5].

In this paper, we shall demonstrate that the mathematical construction formulas of the Gini measure of inequality in a society on the one hand and the entropy–probability trace formula on the other hand bring intriguing similarities at a certain step of their derivation. Both quantities are integrated quantities, in the sense of summing over alternative values of a basic variable, x. We propose the usage of the phrase *"gintropy"* in order to express the combination of the Gini index [3,6,7] and the entropy, both associated with a probability density distribution (PDF).

1.2. Basics

Let us consider the relevant quantity of the investigated system as a continuous variable x. This could be, for example, salary, wealth, population, etc. The occurrence frequency of this given value in a huge set of data are described by the normalized PDF:

$$\int_0^\infty \rho(x)\,dx \ = \ 1. \tag{2}$$

An approximation to such mathematical PDFs is given in the praxis by observing the number of occurrences of values in a short bin $[x, x + dx]$ and dividing these by their sum, the total number:

$$\rho(x) \ = \ \lim_{\Delta x \to 0} \frac{N(x, x + \Delta x)}{N_{\text{tot}} \cdot \Delta x} \tag{3}$$

with N_{tot} the total number of observed data. In income distributions, for example, $N(x, x + \Delta x)$ is the number of persons having an income in the Δx interval starting at x. The total income is then obtained as

$$X_{\text{tot}} \ = \ N_{tot} \int_0^\infty x\,\rho(x)\,dx, \tag{4}$$

and the average income is given by

$$\langle x \rangle \ = \ \int_0^\infty x\,\rho(x)\,dx \ = \ \frac{X_{\text{tot}}}{N_{\text{tot}}}. \tag{5}$$

Both the entropy and the Gini index can be expressed as expectation values of some functions of x over the PDF $\rho(x)$, and we are going to demonstrate the latter in the present paper.

Not only the PDFs, but frequently the cumulative distributions are in our light-spot. The first reason for this is that the experimental shape of the cumulative functions are smoother even in the case of a poorer statistics. The second reason is that, especially for income distribution and inequality, the total body of "rich" is better contrasted to the "poor".

It is straightforward to construct the quantity "the population fraction of richer than x" as the tail-cumulative integral of the PDF:

$$\overline{C}(x) \ = \ \int_x^\infty \rho(y)\,dy. \tag{6}$$

A similar cumulative quantity is the wealth accumulated by this richer class, divided by the average income:

$$\overline{F}(x) \;=\; \frac{1}{\langle x \rangle} \int\limits_{x}^{\infty} y \, \rho(y) \, dy. \tag{7}$$

Trivially, one obtains $\overline{C}(0) = 1$ and $\overline{F}(0) = 1$.

The famous Pareto-law expresses that a p fraction of the population possesses a $(1 - p)$ fraction of the wealth. In the original statement about the economy at the end of 19th century, it was $p = 0.2$, formulated as the "80/20" rule: 20 percent of the population having 80 percent of the total wealth [8–10]. Later, a "90/20" rule has also been suggested by Dunford [11]; this loses, however, the elegant definition of the *Pareto point* (see the next paragraph). Analyses of national GDP comparisons and wealth distribution in certain countries often use in the wealthy region a power-law fit, $\rho(x) = cx^{-(1+\alpha)}$, calling the parameter α the Pareto-index [12–15]. It is however largely debated where should one consider the cut-off in the distribution curve, over which the tail is of the power-law type. For part of the PDF, exponential fits can also be done [16]. As an overall fit to the whole income distribution curve recently, it has been shown that a Tsallis–Pareto cut power-law or some special beta prime distribution works well [17].

For a simple division of the system in an upper and lower class the x_P Pareto-point is used, satisfying:

$$\overline{C}(x_P) \;=\; p, \qquad \text{while} \qquad \overline{F}(x_P) \;=\; 1 - p. \tag{8}$$

The implicit relation, $x_P(p)$, depends on the underlying PDF, $\rho(x)$. Since $\overline{C}(0) + \overline{F}(0) = 2$ and the general sum is monotonically decreasing, due to

$$\frac{d}{dx} \left(\overline{C}(x) + \overline{F}(x) \right) \;=\; - \left(1 + \frac{x}{\langle x \rangle} \right) \rho(x) \;\leq\; 0, \tag{9}$$

there is always a point $x = x_P$, where $\overline{C}(x_P) + \overline{F}(x_P) = 1$. However, the value p cannot be arbitrary.

As we shall discuss in the next section, the Gini index, G, can be expressed in several alternative ways: i) as the average of big differences in the data set, ii) as a construction using the above cumulative quantities or iii) as an expectation value of the cumulative of the cumulative. G expressed as an integral over $\overline{C}$ contains an integrand $\sigma(\overline{C})$. For some PDFs, this function turns out to be formally identical with the terms in entropy—a probability trace formula known from elsewhere. These formulas define the *gintropy*, as a function of the cumulative measure of being "richer than", $\sigma(\overline{C})$—and this function coincides with the classical entropy for an exponential PDF, like the Gibbs–Boltzmann distribution of energy in thermodynamics. For some other, frequently considered distributions in complex systems, the *gintropy* resembles terms of various generalizations of the Gibbs–Boltzmann-Shannon entropy. Among others, we arrive at the Tsallis entropy for the original Pareto distribution, and some further interesting cases. By construction, as we shall demonstrate later, the *gintropy* curve is the difference between the Lorenz curve and the diagonal in the $\overline{F}$ vs. $\overline{C}$ maps.

In the sequel of this paper, we explore these formulas as several facets of the Gini index and its calculation. After the mathematical definitions and equivalent forms, we present certain analytically given PDFs, each reflecting a theoretical possibility about income inequalities: extreme communism giving every person the same income; a divided society defining two classes of the previous case with a fixed share; eco-window, providing equal probability to any income in a fixed, but possibly even an infinite interval; the exponentially distributed income taken as an analogy to the nature of atomic physics; and finally the Pareto-distribution characteristic to capitalism. A different Gini index, G, and also a different *gintropy*, $\sigma(\overline{C})$ belong to each model. Finally, we collect a few ideas about what laws the Gini index and *gintropy* may follow: is there a trend akin to the second law of thermodynamics? Are societies closed systems or not? Can or must inflation distort our analysis?

2. Gross Inequality in General

Let $\rho(x)$ be a normalized PDF. The Gini index in the continuous x case is defined as:

$$G \equiv \frac{1}{2\langle x \rangle} \int_0^\infty dx \int_0^\infty dy \, |x - y| \, \rho(x)\rho(y) = \frac{1}{\langle x \rangle} \int_0^\infty dx \int_x^\infty dy \, (y - x)\rho(x)\rho(y). \tag{10}$$

It can easily be proven that its value is always between zero and one, and is used to quantify the gross inequality in the distribution $\rho(x)$. The original definition (10) can be expressed by using the cumulatives as

$$G = \int_0^\infty \rho(x) \left[\overline{F}(x) - \frac{x}{\langle x \rangle} \overline{C}(x) \right] dx. \tag{11}$$

This expression can be further comprised by considering the cumulative of the cumulative:

$$\overline{h}(x) \equiv \int_x^\infty dy \, \overline{C}(y) = \int_x^\infty dy \int_y^\infty dz \, \rho(z) = \int_x^\infty dz \int_x^z dy \, \rho(z) = \int_x^\infty (z - x)\rho(z) \, dz = \langle x \rangle \overline{F}(x) - x\overline{C}(x). \tag{12}$$

Finally, from here, the Gini index is then expressed as a ratio of two expectation values:

$$G = \frac{1}{\langle x \rangle} \int_0^\infty \rho(x) \, \overline{h}(x) \, dx = \frac{\langle \overline{h}(x) \rangle}{\langle x \rangle}. \tag{13}$$

Alternatively, it can be solely expressed via the cumulative population. From the corresponding definitions, we have the derivatives: $\rho(x) = -d\overline{C}/dx$, $x\rho(x) = -\langle x \rangle \, d\overline{F}/dx$ and therefore $x = \langle x \rangle \, d\overline{F}/d\overline{C}$. Using the immediate equations, $\rho(x) = \frac{d^2\overline{h}(x)}{dx^2}$ and $\overline{C}(x) = -\frac{d\overline{h}(x)}{dx}$, from Equation (13), and, integrating by parts, we get:

$$\langle x \rangle \, G = \int_0^\infty \overline{h} \frac{d^2\overline{h}}{dx^2} \, dx = \overline{h}(0)\overline{C}(0) - \int_0^\infty \overline{C}^2(x) \, dx. \tag{14}$$

Using the boundary conditions $\overline{C}(0) = 1$ and $\overline{h}(0) = \langle x \rangle$, we arrive at

$$G = 1 - \frac{1}{\langle x \rangle} \int_0^\infty \overline{C}^2(x) \, dx = \frac{1}{\langle x \rangle} \int_0^\infty \overline{C} \, (1 - \overline{C}) \, dx. \tag{15}$$

This form is reminiscent of the quantum impurity measure, $Tr(\rho - \rho^2)$, which is zero only for pure states. In the theory of searching trees in informatics, the expression $I_G = \sum_i (p_i - p_i^2)$ is called Gini impurity measure [18]. Let us also note here that for scaling PDFs, i.e $\rho(x) = \frac{1}{\langle x \rangle} f\left(\frac{x}{\langle x \rangle}\right)$, the cumulative functions, $\overline{C}$, and the Gini index, G, do not depend directly on $\langle x \rangle$, it depends only on the form of the $f(z)$ function. This is important when studying the history (time evolution) of G and the related constructions: an overall inflation increasing $\langle x \rangle$ in time will not influence this inequality measure.

Finally, we arrive now at the construction of the quantity *gintropy*. A fashionable representation of the Gini index is realized by plotting the cumulative wealth percentage in terms of the cumulative population possessing that wealth like in Figure 1. It can be shown that the half-moon area between

the Lorenz curve [4,19–22] and the diagonal of the unit square (known as *equality line*) in such an $\overline{F}(x)$ vs. $\overline{C}(x)$ plot,

$$\Sigma \equiv \int\limits_0^1 \sigma(\overline{C}) \, d\overline{C}, \tag{16}$$

is exactly $G/2$. The integrand, $\sigma(\overline{C})$, under the integral over $\overline{C}$—which runs between zero and one—behaves alike an entropy-density. (The original and nowadays used Lorenz curve actually maps the low-cumulatives, integrated from zero to x. However, $\sigma(C)$ instead of $\sigma(\overline{C})$ does not remind to entropy formulas). We call this quantity **gintropy**, and define as the difference between the rich-end-cumulative Lorenz curve and the diagonal:

$$\sigma(x) \equiv \overline{F}(x) - \overline{C}(x) = \int\limits_x^\infty \left(\frac{y}{\langle x \rangle} - 1 \right) \rho(y) dy. \tag{17}$$
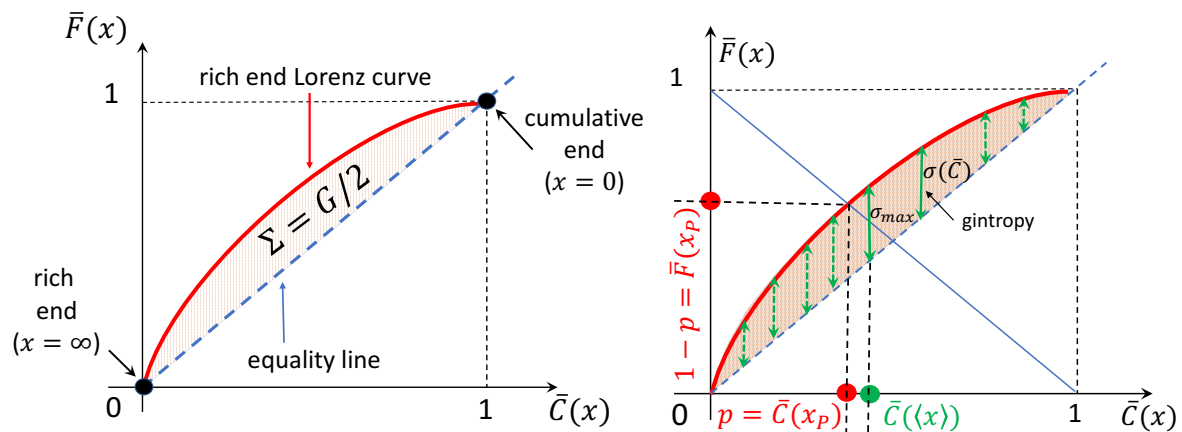
Figure 1. (**left**) The rich end Lorenz curve, and connection with the Gini index. (**right**) Visual illustration of the gintropy, the Pareto Point and the maximal gintropy.

From the above definition, $\sigma(\overline{C})$ remains to be reconstructed with the help of $\overline{C}(x)$. We note that, using the relations $\overline{C}(x) = 1 - C(x)$ and $\overline{F}(x) = 1 - F(x)$, this quantity equivalently can be expressed by the poor-end-cumulative Lorenz curve (the generally used form of the Lorenz curve) too:

$$\sigma(x) = C(x) - F(x) = \int\limits_0^x \left(1 - \frac{y}{\langle x \rangle} \right) \rho(y) dy. \tag{18}$$

The Gini index is expressed from the gintropy as a simple integral

$$G = 2\Sigma = 2 \int\limits_0^1 \sigma(\overline{C}) \, d\overline{C} = 2 \int\limits_0^\infty \sigma(x) \rho(x) \, dx = 2 \langle \sigma(x) \rangle. \tag{19}$$

We note here that, for any integral, one substitutes $\int\limits_0^1 f(\overline{C}) \, d\overline{C} = \int\limits_0^\infty f(\overline{C}(x)) \rho(x) \, dx$.

It is interesting to summarize the proof of this statement here because it is a central motivation of thinking in terms of gintropy. Using the respective definitions of the tail-cumulative quantities, the half-moon area (16) is calculated as the following double integral:

$$\Sigma \;=\; \int\limits_0^\infty dx\, \rho(x) \int\limits_x^\infty dy \left(\frac{y}{\langle x \rangle} - 1 \right) \rho(y). \tag{20}$$

Changing the order of integration leads to

$$\Sigma \;=\; \int\limits_0^\infty dy \int\limits_0^y dx\, \rho(x) \left(\frac{y}{\langle x \rangle} - 1 \right) \rho(y) \;=\; \int\limits_0^\infty dy\, (1 - \overline{C}(y)) \left(\frac{y}{\langle x \rangle} - 1 \right) \rho(y). \tag{21}$$

Here, the term with 1 in the first parenthesis integrates to zero due to the definition of the expectation value, $\langle x \rangle$. Then, we replace $-\overline{C}(y)\rho(y) = \frac{1}{2} \frac{d}{dy} \overline{C}^2$, integrate by parts and compare the result to (15) to conclude:

$$\Sigma \;=\; \frac{1}{2} \overline{C}^2(0) - \frac{1}{2\langle x \rangle} \int\limits_0^\infty dy\, \overline{C}^2(y) \;=\; \frac{1}{2} G. \tag{22}$$

Now, we explore some basic properties of gintropy. Some of these provide further evidence to consider gintropy like a generalized entropy density.

1. The gintropy is never negative: $\sigma = \overline{F} - \overline{C} \geq 0$ is proven by inspecting the integral

$$\sigma(x) \;=\; \int\limits_x^\infty (y/\langle x \rangle - 1)\rho(y)dy \;=\; \int\limits_0^x (1 - y/\langle x \rangle)\rho(y)dy \;\geq\; 0,$$

   and taking the first form for $x \geq \langle x \rangle$, the second form for the opposite case. This implies that the rich-end wealth fraction is always bigger or equal to the population fraction possessing it.
2. The gintropy is maximal at $x = \langle x \rangle$, $\sigma_{max} = \sigma(\langle x \rangle)$, since $d\sigma/dx = (1 - x/\langle x \rangle)\rho(x)$ changes its sign exactly there and only there.
3. According to eq. (8) at the Pareto-point, the gintropy equals $\sigma(x_P) = 1 - 2p$, and therefore, for the Pareto point, $p \leq 1/2$ holds for the rich fraction. Since $\sigma_{max} \geq \sigma(x_P)$, in order to get a Pareto point: $\sigma(\langle x \rangle) \geq 1 - 2p$, i.e., the maximum of the gintropy has to be bigger than this difference value. As a consequence for the Pareto Point, we have a restriction imposed by the maximal gintropy $(1 - \sigma(\langle x \rangle))/2 \leq p \leq 1/2$.
4. The expectation value of gintropy is the half of the gini index: $\int\limits_0^\infty \sigma(x)\,\rho(x)\,dx \;=\; \int\limits_0^1 \sigma(\overline{C})\,d\overline{C} \;=\; \Sigma \;=\; G/2$.
5. The integral of gintropy over the base value x is the non-Poissonity index, $\int\limits_0^\infty \sigma(x)\,dx \;=\; \frac{\text{Var}(x)}{\langle x \rangle}$, with $\text{Var}(x) = \langle x^2 \rangle - \langle x \rangle^2$ being the variance of x. The proof of this statement uses the same mathematical trick as the one in eq. (12).
6. For some particular PDFs, $\sigma(\overline{C})$ looks like an entropy density formula, $s(p_i)$. We present important examples in the next section.

3. Important Examples

In this section, we list some important examples of the gintropy, $\sigma(\overline{C})$. We go through primitive models of income/wealth distributions, labelled as communism, communism++, eco-window, natural, or capitalism. Starting from model PDFs, the gintropy expression and the Gini index are calculated.

3.1. Communism

Our first example is communism: all incomes are equal, the PDF is simply a singular delta-distribution, peaked at the single value a: $\rho(x) = \delta(x - a)$ leading to $\langle x \rangle = a$, $\overline{C}(x) = \Theta(a - x)$ and $\overline{h}(x) = (a - x)\Theta(a - x)$, with $\Theta(x)$ the Heaviside step function defined as:

$$\Theta(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & (x > 0) \end{cases} \tag{23}$$

This leads to $\langle x \rangle \, \overline{F} = \overline{h} + x\overline{C} = a\Theta(a - x)$ and by that

$$\sigma(x) = \overline{F}(x) - \overline{C}(x) = 0, \tag{24}$$

i.e., to an identically vanishing gintropy. As a consequence, also $G = 0$. Here, the Pareto-point belongs to a 50/50 division.

3.2. Communism++

The next example we present is a slight variation of the previous: now, two peaks in a given ratio constitute the PDF. This belongs to a two-class-society where *all are equal but some of them are more equal*. The two-peak-PDF, $\rho(x) = w\,\delta(x - a) + (1 - w)\,\delta(x - b)\,(b > a)$, delivers $\langle x \rangle = w\,a + (1 - w)\,b$. The w fraction of the population has an income a and the $(1 - w)$ fraction b. The cumulative rich population graph shows two steps, at a and b, respectively:

$$\overline{C}(x) = w\,\Theta(a - x) + (1 - w)\,\Theta(b - x), \tag{25}$$

having the value 1 for $x \le a$, $(1 - w)$ for $x \in [a, b]$, and 0, otherwise. Therefore, $C(x) = 1 - \overline{C}(x)$ is zero for $x \le a$, equals w in the mid interval, and has the value 1 otherwise. The Gini index is obtained from this as:

$$G = \frac{1}{\langle x \rangle} \int_0^\infty \overline{C}(1 - \overline{C})dx = \frac{1}{\langle x \rangle}(b - a)\,w\,(1 - w). \tag{26}$$

Expressing the weights, $w = \frac{b - \langle x \rangle}{b - a}$ and $1 - w = \frac{\langle x \rangle - a}{b - a}$, we obtain the alternative form

$$G = \frac{(\langle x \rangle - a)(b - \langle x \rangle)}{(b - a)\,\langle x \rangle}. \tag{27}$$

It is worth noting that, for $a \to 0$, i.e., when the lower class has (almost) zero income, the Gini index, cf. (26) tends to $G \to w$, exactly the share of the people earning $a \to 0$ in the population. This result is independent of b, the income in the upper class.

The gintropy, following its definition, first is expressed as a function of x:

$$\sigma(x) = \overline{F}(x) - \overline{C}(x) = w\left(\frac{a}{\langle x \rangle} - 1\right)\Theta(a - x) + (1 - w)\left(\frac{b}{\langle x \rangle} - 1\right)\Theta(b - x). \tag{28}$$

It is easy to see that, outside the interval $[a, b]$, the gintropy is zero. Inside the interval, only the second term survives giving
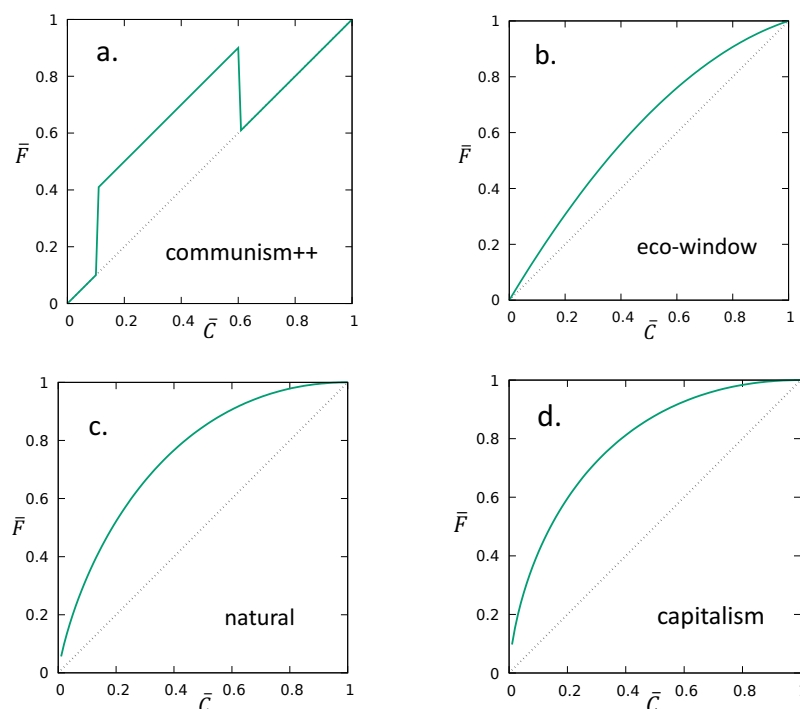
$$\sigma(x) = G\left[\Theta(b-x) - \Theta(a-x)\right]. \tag{29}$$

In conclusion, $\sigma(\overline{C})$ shows a plateau at $\overline{C} = 1 - w$ with the value G and its jumps are at $\overline{C}(a) = 1 - w/2$ and $\overline{C}(b) = (1-w)/2$:

$$\sigma(\overline{C}) = G\left[\Theta(\overline{C}(a) - \overline{C}) - \Theta(\overline{C}(b) - \overline{C})\right]. \tag{30}$$

It is easy to check that indeed

$$\Sigma = \int_0^1 \sigma(\overline{C}) \, d\overline{C} = G\left[\overline{C}(a) - \overline{C}(b)\right] = G/2. \tag{31}$$

The corresponding Lorenz curve is illustrated in Figure 2a.

Figure 2. The $\overline{F}$ vs. $\overline{C}$ cumulative maps (Lorenz curves) for the (**a**) communism++ ($a = 1$, $b = 4$ and $w = 0.8$), (**b**) eco-window ($a = 1, b = 5$), (**c**) natural exponential ($\langle x \rangle = 1$) and for the capitalism (**d**) ($A = 1, B = 3 \to q = 3/4$) distributions. The corresponding Gini index are $G = 0.3$, $G = 2/9$, $G = 1/2$ and $G = 4/7$, respectively.

3.3. Eco-Window

The next example is still mathematically simple with a window-form PDF. We label this as *eco-window*: here, everyone has the same chance for all of possible incomes between a and b. Eventually, $a = 0$ and/or $b = \infty$ may be considered, as special cases. For the PDF $\rho(x) = \frac{1}{b-a}\left[\Theta(b-x) - \Theta(a-x)\right]$, one obtains the following cumulative rich distribution:

$$\overline{C}(x) = \frac{b-x}{b-a}\Theta(b-x) - \frac{a-x}{b-a}\Theta(a-x) = \begin{cases} 1 & (x < a) \\ \frac{b-x}{b-a} & x \in [a, b] \\ 0 & (x > b) \end{cases} \tag{32}$$

Obviously, $\langle x \rangle = (a+b)/2$ and, according to eq. 15, the Gini index becomes:

$$G = \frac{1}{\langle x \rangle} \int_a^b \frac{(b-x)(x-a)}{(b-a)^2} dx = \frac{1}{3}\frac{b-a}{b+a}. \tag{33}$$

After some tedious but straightforward calculation, the *gintropy* is obtained as a function of $\overline{C}$:

$$\sigma(\overline{C}) = 3\,G\,\overline{C}(1-\overline{C}). \tag{34}$$

For a specific choice of a and b, the corresponding Lorenz curve is illustrated in Figure 2b.

3.4. Natural Distribution

Our next example is the natural distribution, mimicking the Boltzmann–Gibbs exponential energy distribution, known from statistical physics. This is not necessarily an equilibrium distribution, it may also be the stationary limit of "growth and resetting" type processes with quantity-independent rates [23]. The PDF is a scaling one: $\rho(x) = \frac{1}{\langle x \rangle} e^{-x/\langle x \rangle}$. The corresponding tail-cumulative probability, the rich population is given by

$$\overline{C}(x) = e^{-x/\langle x \rangle}, \tag{35}$$

and the Gini index becomes

$$G = 1 - \frac{1}{\langle x \rangle} \int_0^\infty e^{-2x/\langle x \rangle} dx = \frac{1}{2}. \tag{36}$$

Our gintropy formula is constructed as follows: first, we obtain the cumulative of the cumulative,

$$\overline{h} = \int_x^\infty e^{-y/\langle x \rangle}\, dy = \langle x \rangle\, e^{-x/\langle x \rangle}. \tag{37}$$

From this, it is easy to obtain the wealth share of the rich classes, $\langle x \rangle\, \overline{F} = \overline{h} + x\overline{C} = (x + \langle x \rangle)e^{-x/\langle x \rangle}$, and based on this the gintropy

$$\sigma(x) = \frac{x}{\langle x \rangle} e^{-x/\langle x \rangle}. \tag{38}$$

In order to express it as a function of $\overline{C}$, we invert (35) to have

$$x(\overline{C}) = -\langle x \rangle \ln \overline{C}. \tag{39}$$

Finally, it leads to

$$\sigma(\overline{C}) = -\overline{C} \ln \overline{C}. \tag{40}$$

Apart from a constant proportionality factor, this formula formally coincides with the terms in the sum of the Boltzmann–Gibbs–Shannon entropy:

$$S = -k \sum_i p_i \ln(p_i) \tag{41}$$

To continue the analogy, also $\overline{C} \in [0,1]$. Indeed, in this case, gintropy is like the entropy density, with the caveat that the cumulative values $\overline{C}(x)$ are never disjunct for different x-s; instead, they overlap and show a definite hierarchy. The Lorenz curve for $\langle x \rangle = 1$ is illustrated in Figure 2c.

3.5. Capitalism

Our last example is capitalism, conjecturing the base PDF being the cut Pareto (known also as Tsallis–Pareto or Lomax II) distribution [24]:

$$\rho(x) = A(B+1)(1+Ax)^{-B-2} \tag{42}$$

This distribution can also be obtained as the canonical equilibrium optimizer of the Tsallis entropy [25]. The tail-cumulative integral is

$$\overline{C}(x) = (1+Ax)^{-B-1}, \tag{43}$$

which upon integration leads to the following cumulative of the cumulative:

$$\overline{h}(x) = \frac{1}{AB}(1+Ax)^{-B}. \tag{44}$$

This result also delivers the expectation value, $\langle x \rangle = \overline{h}(0) = 1/AB$. The Gini index is calculated in the (15) form, and it becomes

$$G = 1 - AB \int_0^\infty (1+Ax)^{-2B-2}dx = \frac{B+1}{2B+1}. \tag{45}$$

The gintropy as a function of the income, x, follows the form

$$\sigma(x) = A(B+1)\,x\,(1+Ax)^{-B-1}. \tag{46}$$

In order to express this result akin to the entropy, we write σ as a function of $\overline{C}$ using the inversion of eq. (43)

$$x(\overline{C}) = \frac{1}{A}\left(\overline{C}^{-\frac{1}{B+1}} - 1\right), \tag{47}$$

and we obtain

$$\sigma(\overline{C}) = (B+1)\left(\overline{C}^{\frac{B}{B+1}} - \overline{C}\right). \tag{48}$$

Finally, using the Tsallis parameter, $q = B/(B+1)$, we arrive at the formula:

$$\sigma(\overline{C}) = \frac{1}{1-q}(\overline{C}^q - \overline{C}), \tag{49}$$

One immediately makes an analogy with the terms in the Tsallis entropy formula:

$$S_q = \frac{k}{1-q}\sum_i (p_i^q - p_i), \tag{50}$$

The Gini index is simply

$$G = \frac{1}{q+1}. \tag{51}$$

Similar to the previously considered cases, we illustrate the Lorenz curve for this distributions as well. For $A = 1$ and $B = 3$, the corresponding Lorenz curve is plotted in Figure 2d.

Finally, we summarize the lesson of the considered theoretical examples in Table 1 and Figure 3.
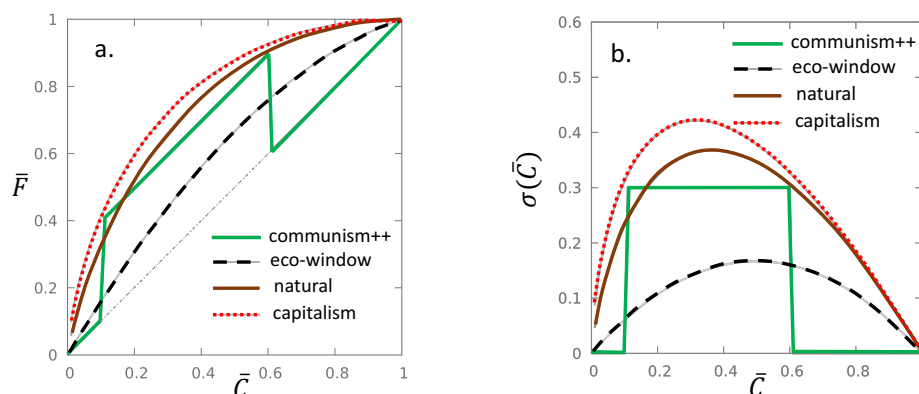
Figure 3. (**a**) $\overline{F} - \overline{C}$ Lorenz curves in one comparison and (**b**) the corresponding **gintropy** curves, $\sigma(\overline{C})$, for the communism++, eco-window, natural and capitalism models. The Gini indices are $G = 0.3$, $G = 2/9$, $G = 1/2$ and $G = 4/7$, respectively.

Table 1. Summary of PDFs, the gintropy formulas and Gini index values for some ideal income/wealth distribution schemes.

	$\rho(x)$	$\sigma(\overline{C})$	G
natural	$\frac{1}{\langle x \rangle} e^{-x/\langle x \rangle}$	$-\overline{C} \ln \overline{C}$	$\frac{1}{2}$
capitalism	$\frac{A}{1-q}(1 + Ax)^{\frac{-1}{1-q}}$	$\frac{1}{1-q}\left(\overline{C}^q - \overline{C}\right)$	$\frac{1}{q+1} \geq \frac{1}{2}$
eco-window	$\frac{1}{b-a}[\Theta(b-x) - \Theta(a-x)]$	$3G\,\overline{C}(1 - \overline{C})$	$\frac{1}{3}\frac{b-a}{b+a} \leq \frac{1}{3}$
communism++	$w\,\delta(x-a) + (1-w)\,\delta(x-b)$	$G\left[\Theta(\overline{C}(a) - \overline{C}) - \Theta(\overline{C}(b) - \overline{C})\right]$	$\frac{(b-a)q\,(1-w)}{w\,a+(1-w)\,b}$
communism	$\delta(x-a)$	0	0

4. Conclusions

In this work, we explored a density-like quantity called *gintropy* which occurs in calculating the Gini index, G, for a given relevant socio-economic distribution, $\rho(x)$. This gintropy can be deduced from two cumulative functions, the rich population fraction and the corresponding richness fraction, $\overline{C}(x)$ and $\overline{F}(x)$, respectively. The proposed "gintropy" name is meant to suggest a connection between the inequality measure quantified by the Gini index and the entropy. Its dependence on the rich population fraction cumulative function is reminiscent of terms in entropy formulas, known from physics, statistics, and informatics. More precisely, we found that, for the the natural, exponential PDF, the gintropy is reminiscent of the classical Boltzmann–Gibbs–Shannon formula, $\sigma(\overline{C}) = -\overline{C} \ln \overline{C}$. The Gini index is then the expectation value of the gintropy function; for the exponential PDF, its value is 1/2. For the Tsallis–Pareto distribution, the Gini index must be always over this value.

Several other PDFs have been suggested to describe income or wealth distributions in terms of time [17,26–29]. Many of them are not treatable analytically, so the $\sigma(\overline{C})$ relation can only be explored numerically.

Beyond igniting the theoretical phantasy, the gintropy—reminscent of generalized entropy formulas—is also the one-variable density, which lays under the Gini index, originally defined for measuring inequality. Turning this statement around, should we look for such generalizations of the classical entropy formula that are inequality or impurity measures at the same time? We believe that this criterion selects a subclass of possible statistical theories among all possible approaches to the origin, behavior and future of social and economical inequalities. Even generalizations of the Gini index formula have been suggested a few times, cf. [30,31]. We do not expect that a corresponding gintropy ("Lorenz curve minus the diagonal") would resemble any known entropy formula, but this question needs further study.

Finally, it seems that the "correct" entropy measure for economical and social theories can hardly be a simple copy of the classical formula known from physics, mathematics and informatics. Our procedure, described above, is more promising: a recipe for constructing gintropy from cumulative functions of the underlying PDF whose expectation value is the half Gini index and whose dependency on the cumulative rich population coincides with various generalizations of the entropy–probability formula.

Author Contributions: Conceptualization, T.B.; Methodology, T.B.; Validation, Z.N., Formal Analysis, T.B. and Z.N.; Investigation, T.B. and Z.N; Writing—Original Draft Preparation, T.B.; Writing—Review & Editing, T.B. and Z.N.; Visualization, T.B. and Z.N.; Funding Acquisition, T.B. and Z.N. All authors have read and agreed to the published version of the manuscript.

References

1. Thurner, S.; Hanel, R.; Corominas-Murtra, B. The three faces of entropy for complex systems- information, thermodynamics and the maxent principle. *Phys. Rev. E* **2017**, *96*, 032124

2. Amigo, J.; Balogh, S.; Hernandez, S. A Brief review of Generalized Entropies. *Entropy* **2020**, *22*, e20110813.

3. Gini, C. Sulla misura della concentrazione e della variabilitá dei caratteri. *Lettere e Arti* **1914**, *73*, 1203–1248.

4. Lorenz, M.O. Methods of measuring the concentration of wealth. *Publ. Am. Stat. Assoc.* **1905**, *9*. pp. 209-219.

5. Marmani, S.; Ficcadenti, V.; Kaur, P.; Dhesi, G. Entropic Analysis of Votes Expressed in Italian Elections between 1948 and 2018. *Entropy* **2020**, *22*, 523.

6. Atkinson, A.B. On the measurement of inequality. *J. Econ. Theory* **1970**, *2*, 244–263.

7. Shorrocks, A.F. The Class of Additivity Decomposable Inequality Measures. *Econometrica* **1980**, *48*, 613–625.

8. Pareto, V. Cours d' economie politique. *Political Sci. Q.* **1896**, *11*, pp.750–751.

9. Pareto, V. The New Theories of economics. *J. Political Econ.* **1897**, *5*, pp 485–502.

10. Schumpeter, J.A. Vilfredo Pareto (1848–1923). *Q. J. Econ.* **1949**, *63*, 147–173.

11. Dunford, R.; Su, Q.; Tamang, E. The Pareto Principle. *Plymouth Stud. Sci.* **2014**, *7*, 140–148.

12. Levy, M.; Solomon, S. New evidence for the power-law distribution of wealth. *Physica A* **1997**, *242*, 90–94.

13. Piketty, I. *Capital in the Twenty-First, Century*. Harvard University Press: Cambridge, MA, USA, 2014.

14. Sinha, S. Evidence for the Power-law tail of the wealth-distribution in India. *Physica A* **2006**, *359*, 555–562.

15. Dragulescu, A.; Yakovenko, V.M. Statistical mechanics of money. *Eur. Phys. J. B* **2000**, *17*, 723–729.

16. Dragulescu, A.; Yakovenko, V.M. Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A* **2001**, *299*, 213–221.

17. Neda, Z.; Gere, I.; Biro, T.S.; Toth, G.; Derzsy, N. Scaling in income inequalities and its dynamical origin. *Physica A* **2020**, *549*, 124491.

18. Molala, R. Entropy, Information Gain, Gini Index – The Crux of a Decision Tree. Available online: https://blog.clairvoyantsoft.com (accessed on 12 May 2020).

19. Iritani, J.; Kuga, K. Duality between the Lorenz curves and the income distribution functions. *Econ. Stud. Q.* **1983**, *34*, 9–21.

20. Thistle, P.D. Duality between generalized Lorenz curves and distribution functions. *Econ. Stud. Q.* **1989**, *40*, 183–187.

21. Aaberge, R. Characterizations of Lorenz curves and income distributions. *Social Choice Welfare* **2000**, *17*, 639–653.

22. Kleiber, C. The Lorenz curve in economics and econometrics. Gini-Lorenz Centennial Conference, Siena, May 23–26, 2005, and in Gianni Betti and Achille Lemmi (eds.): Advances on Income Inequality and Concentration Measures. Collected Papers in Memory of Corrado Gini and Max O. Lorenz (London Routledge, 2008)

23. Bíró, T.S.; Néda, Z. Unidirectional random growth with resetting. *Physica A* **2018**, *499*, 335-361

24. Dogum, C. A new model of personal income distributiuons: Specification and estimation. *Economie Appliquée* **1977**, *30*, 413–437.

25. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.

26. McDonald, J.B. Some generalized functions for the size distribution of income. *Econometrica* **1984**, *52*, 647–663.

27. Taillie, C. Lorenz ordering within the generalized gamma family of income distributions. *Stat. Distrib. Sci. Work* **1981**, *6*, 181–192.

28. Wilfling, B.; Krämer, W. Lorenz ordering of Singh-Maddala income distributions. *Econ. Lett.*, **1993**, *43*, 53–57.

29. Bouchaud, J.P.; Mezard, M. Wealth condensation in a simple model of economy. *Physica A* **2000**, *282*, 536–545.

30. Kleiber, C.; Kotz, S. A characterization of income distributions in terms of generalized Gini coefficients. *Social Choice Welfare* **2002**, *19*, 789–794.

31. Yitzhaki, S. On an extension of the Gini inequality index. *Int. Econ. Rev.* **1983**, *24*, 617–628