Chaos, Solitons and Fractals 92 (2016) 30–42

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Investigating equality: The poverty and riches indices

Iddo Eliazar

Smart Device Innovation Science Team, New Devices Group, Intel Corporation, Yakum, Israel

ARTICLE INFO

Article history:
Received 28 May 2016
Accepted 25 August 2016
Available online 17 September 2016

Keywords:
Equality and inequality
Relative entropy
Lorenz curves
Gini index
Pietra index

ABSTRACT

As noise is omnipresent, real-world quantities measured by scientists and engineers are commonly obtained in the form of statistical distributions. In turn, perhaps the most compact representation of a given statistical distribution is via the mean-variance approach: the mean manifesting the distribution's 'typical' value, and the variance manifesting the magnitude of the distribution's fluctuations about its mean. The mean-variance approach is based on an underlying Euclidean-geometry perspective. So very often real-world quantities of interest are non-negative sizes, and their measurements yield statistical size distributions. In this paper, and in the context of size distributions, we present an alternative to the Euclidean-based mean-variance approach: a mean-equality approach that is based on an underlying socioeconomic perspective. We establish two equality indices that score, on a unit-interval scale, the intrinsic 'egalitarianism' of size distributions: (i) the *poverty equality index* which is particularly sensitive to the existence of very small "poor" sizes; (ii) the *riches equality index* which is particularly sensitive to the existence of very large "rich" sizes. These equality indices, their properties, their computation, their application, and their connections to the mean-variance approach – are explored and described comprehensively.

1. Introduction

Statistical distributions are ubiquitous. Indeed, in essentially every field of science and engineering, empirical real-world data is commonly communicated and presented using the quantitative language of *Statistics*. The first-step approximation of an empirical real-valued quantity is via its *mean*: the quantity's "typical" value. The second-step approximation is via the quantity's *variance*: the magnitude of the quantity's fluctuations about its mean.

A most vivid illustration of the *mean-variance* approach is the well-known bell curve, the density function of perhaps the most fundamental statistical distribution – the "normal" Gauss law. Recall that the Gauss law emerges universally via the Central Limit Theorem [1], and that it is the only statistical distribution that is characterized by its mean and variance. In the context of the Gauss law the mean represents the location of the bell-curve's center, and the variance represents the bell-curve's width, i.e. the "peakness" of the curve about its center.

The mean and the variance follow from a Euclidean-geometry perspective. Indeed, given a random real-valued quantity, and with regard to the Euclidean metric: the mean is the best deterministic approximation of the random quantity; the variance is the mean square distance of the random quantity from its best deterministic approximation. Moreover, the quadratic structure of the Euclidean metric is induced to the variance, thus endowing the variance with handy mathematical properties.

Often, real-valued quantities are non-negative *sizes*, e.g. count, length, area, volume, mass, energy, and duration. *Size distributions* are omnipresent across science and engineering, and are of great importance. The distributions of *wealth* in human societies constitute a particular example of size distributions and of their significance [2–6]. In the context of the wealth distribution of a given society the abovementioned mean has a socioeconomic meaning: should the society be purely egalitarian, i.e. purely communist, the mean would represent the common wealth value of the society's members.

So, from a Euclidean perspective the mean manifests the best deterministic approximation. On the other hand, in the context of wealth distributions, and from a socioeconomic perspective, the mean manifests the state of *pure communism*. As noted above, in the Euclidean perspective the deviation from the best deterministic approximation is quantified by the variance. Analogously, in the socioeconomic perspective, measures of deviation from the state of pure communism will quantify the *inequality* of the wealth distribution under consideration.

The main portal to socioeconomic inequality is facilitated by the notion of *Lorenz curves* [7–11]. In turn, the Lorenz curves give rise to the notion of *inequality indices*: numerical scores of inequality that take values in the unit interval [12–16]. A zero inequality

E-mail addresses: iddo.eliazar@intel.com, eliazar@post.tau.ac.il

http://dx.doi.org/10.1016/j.chaos.2016.08.012
0960-0779/

score characterizes the state of pure communism, and the larger the score – the larger is the gap between the rich and poor of the wealth distribution under consideration. The most widely applied inequality index is the well-known *Gini index* [17–21].

While inequality is commonly introduced in the context of wealth distributions, it can be effectively applied in the contexts of size distributions at large. From an abstract mathematical perspective it makes no difference if the size considered represents wealth or not. Indeed, any non-negative quantity can be conceptually perceived as "wealth", and hence the notion of inequality indices can be applied to any size distribution of interest. The following references exemplify the application of the Gini index well beyond the realm of wealth distributions: [22–36].

In this paper we follow the line of thought described in the above paragraph: we consider general size distributions as "wealth" distributions, and analyze them from a socioeconomic perspective. Doing so we measure the *equality* – rather than the inequality – of general size distributions, and present an alternative to the aforementioned Euclidean-based mean-variance approach: a socioeconomic-based *mean-equality* approach.

Specifically, in this paper we establish two equality indices: the *poverty index* and the *riches index*. Both these indices take values in the unit interval, and yield a unit score if and only if the state of pure communism is attained – i.e. if and only if the size distribution is deterministic, with all its sizes being equal. Also, the lower the scores of the two indices – the larger the rich-poor gap, i.e. the gap between the large "rich" sizes and the small "poor" sizes.

As their names imply, the poverty and the riches indices set their focuses on the poor and on the rich, respectively. The poverty index assigns a zero score when *extreme poverty* is present – loosely speaking, when the size distribution includes an "impoverished" class of zero sizes. The riches index assigns a zero score when *extreme riches* is present – loosely speaking, when the size distribution includes an "oligarchic" class of large sizes.

The aforementioned Gini index is a *linear functional* of Lorenz curves, and this linear structure renders the Gini index insensitive to neither extreme poverty nor extreme riches. On the other hand, the poverty and the riches indices are *nonlinear functionals* of Lorenz curves, and their nonlinear structure is key in enabling the detection of extreme poverty and extreme riches, respectively. Also, this nonlinear structure equips the poverty and the riches indices with a multiplicative feature that the Gini index does not exhibit.

Researches and practitioners tackle an abundance of size distributions on an everyday basis. In the context of size distributions, the goal of this paper is to offer a socioeconomic-based mean-equality toolbox that is both alternative and parallel to the commonly applied Euclidean-based mean-variance toolbox. To that end we construct the poverty and the riches indices step-by-step and in a fully self-contained fashion, analyze these indices comprehensively, and explore their properties in detail. In addition, illustrative examples and insightful discussions are intertwined throughout the manuscript.

A note about notation: Throughout the manuscript $\mathbb{E}[\cdot]$ denotes the operation of mathematical expectation. Namely, $\mathbb{E}[Z]$ is the mathematical expectation, i.e. the mean, of a real-valued random variable Z.

2. Foundation

Consider a population that is partitioned into n groups labeled $i = 1, \ldots, n$. The groups are non-empty and non-overlapping, and their union is the entire population. The proportion of group i, with respect to the entire population, is p_i. Note that the proportions are positive and that they sum up to one; hence $\mathbf{p} = (p_i)_{i=1}^n$ is a probability vector. The partitioning is general, and it can be determined by arbitrary factors. For example, the partitioning can be: (i) gender-based – grouping the population members according to their sex; (ii) age-based – grouping the population members into age segments; (iii) geography-based – grouping the population members according to their areas of residence; (iv) occupation-based – grouping the population members according to their professions; (v) income-based – grouping the population members according to their wages; etc.

Here and hereinafter X denotes a random variable that represents the wealth of a randomly sampled member of the population. Similarly, X_i denotes a random variable that represents the wealth of a randomly sampled member of group i $(i = 1, \ldots, n)$. The random variable X is considered to be non-negative valued, yet not identically zero $(\Pr(X = 0) < 1)$, and hence its mean is positive: $\mu := \mathbb{E}[X] > 0$. Consequently, the random variables $\{X_i\}_{i=1}^n$ are non-negative valued, and so are their means: $\mu_i := \mathbb{E}[X_i] \geq 0$ $(i = 1, \ldots, n)$. Conditional expectation implies that the connection between the population mean μ and the groups" means $\{\mu_i\}_{i=1}^n$ is given by

$$\mu = \sum_{i=1}^{n} p_i \mu_i. \tag{1}$$

Dividing both sides of Eq. (1) by the positive population mean μ we introduce the quantities $q_i = p_i\mu_i/\mu$ $(i = 1, \ldots, n)$. Note that these quantities are non-negative and sum up to one; hence $\mathbf{q} = (q_i)_{i=1}^n$ is a probability vector. On the one hand, the probability vector $\mathbf{p}$ manifests the partition's *population-distribution*. On the other hand, the probability vector $\mathbf{q}$ manifests the partition's *wealth-distribution*. Consequently, the divergence of the probability vector $\mathbf{q}$ from the probability vector $\mathbf{p}$ can be used to quantify the degree of the partition's *socioeconomic inequality*. Introduced by Kullback and Leibler, the most common measure of divergence of one probability vector from another is *relative entropy* [37,38]. Specifically, the relative entropy of the probability vector $\mathbf{q}$, with respect to probability vector $\mathbf{p}$, is given by

$$\mathcal{H}(\mathbf{q}|\mathbf{p}) = -\sum_{i=1}^{n} \ln\left(\frac{q_i}{p_i}\right) p_i. \tag{2}$$

The *Gibbs inequality* asserts that the relative entropy is always non-negative, $\mathcal{H}(\mathbf{q}|\mathbf{p}) \geq 0$, and that it vanishes if and only if its two probability vectors coincide: $\mathcal{H}(\mathbf{q}|\mathbf{p}) = 0 \Leftrightarrow \mathbf{q} = \mathbf{p}$ [39,40].

In this paper we set

$$\begin{aligned}\mathcal{E}(\mathbf{q}|\mathbf{p}) &:= \exp\left[-\mathcal{H}(\mathbf{q}|\mathbf{p})\right] \\ &= \prod_{i=1}^{n}\left(\frac{q_i}{p_i}\right)^{p_i} = \prod_{i=1}^{n}\left(\frac{\mu_i}{\mu}\right)^{p_i} \\ &= \frac{1}{\mu}\prod_{i=1}^{n}\mu_i^{p_i} = \frac{\prod_{i=1}^{n}\mu_i^{p_i}}{\sum_{i=1}^{n}p_i\mu_i}\end{aligned} \tag{3}$$

to be the partition's *equality index*. As its name suggests, the equality index $\mathcal{E}(\mathbf{q}|\mathbf{p})$ is a quantitative score of the partition's socioeconomic equality. Indeed, the Gibbs inequality and Eq. (3) imply that the equality index $\mathcal{E}(\mathbf{q}|\mathbf{p})$ takes values in the unit interval, $0 \leq \mathcal{E}(\mathbf{q}|\mathbf{p}) \leq 1$, and that it exhibits the three following properties:

1. It attains its unit upper bound if and only if all the groups means coincide: $\mathcal{E}(\mathbf{q}|\mathbf{p}) = 1 \Leftrightarrow \mu_i = \mu$ for all $i = 1, \ldots, n$.
2. It attains its zero lower bound if and only if the mean of at least one group vanishes: $\mathcal{E}(\mathbf{q}|\mathbf{p}) = 0 \Leftrightarrow \mu_i = 0$ for some $i = 1, \ldots, n$.
3. It is invariant with respect to linear transformations of the measurement of wealth: applying on the means the linear transformation $x \mapsto ax$, where a is a positive parameter, does not affect $\mathcal{E}(\mathbf{q}|\mathbf{p})$.

Among all n-dimensional probability vectors $\mathbf{p}$, the vector that *maximizes entropy* is the *uniform* one: $p_i = 1/n$ $(i = 1, \ldots, n)$ [41,42].