Physica A 454 (2016) 66–80

Contents lists available at ScienceDirect

Physica A

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

Visualizing inequality

Iddo Eliazar

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

HIGHLIGHTS

- An alternative to Lorenz curves is presented: Hill curves.
- Hill curves quantify and visualize socioeconomic inequality.
- Hill curves are a 'hyperspectral measure' for the statistical variability of size distributions at large.

ARTICLE INFO

Article history:
Received 21 August 2015
Received in revised form 22 February 2016
Available online 2 March 2016

Keywords:
Inequality
Gini index
Lorenz curves
Hill curves
Statistical variability
Infographics

ABSTRACT

The study of socioeconomic inequality is of substantial importance, scientific and general alike. The graphic visualization of inequality is commonly conveyed by Lorenz curves. While Lorenz curves are a highly effective statistical tool for quantifying the distribution of wealth in human societies, they are less effective a tool for the visual depiction of socioeconomic inequality. This paper introduces an alternative to Lorenz curves—the *hill curves*. On the one hand, the hill curves are a potent scientific tool: they provide detailed scans of the rich–poor gaps in human societies under consideration, and are capable of accommodating infinitely many degrees of freedom. On the other hand, the hill curves are a powerful infographic tool: they visualize inequality in a most vivid and tangible way, with no quantitative skills that are required in order to grasp the visualization. The application of hill curves extends far beyond socioeconomic inequality. Indeed, the hill curves are highly effective 'hyperspectral' measures of statistical variability that are applicable in the context of size distributions at large. This paper establishes the notion of hill curves, analyzes them, and describes their application in the context of general size distributions.

1. Introduction

One of the few topics that draws great interest both among the scientific community and among the general public is *socioeconomic inequality*. Scientific interest in the study of inequality was ignited at the close of the nineteenth century by the trailblazing work of the Italian polymath Vilfredo Pareto [1]. Following Pareto, economists, social scientists, and nowadays econophysicists, are vigorously and continuously studying inequality in human societies [2–11]. General interest in inequality dates back to biblical times, and is peaking in the recent years due to the global social trend of increasing gaps between the rich and the poor [12–21]. Two illuminating examples underscoring the wide and cross-disciplinary interest in inequality are the magnum opus of the political philosopher John Rawls [22], and a recent issue of the journal Nature that is all dedicated to this topic [23].

The most popular measure of inequality is the *Gini index*, which takes values in the unit interval [24–26]. Applied to the distribution of wealth in a given human society, a low Gini index manifests socioeconomic egalitarianism, whereas a high Gini index manifests a significant rich–poor disparity. The Gini index is derived from the *Lorenz curve*, which is a function that

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

http://dx.doi.org/10.1016/j.physa.2016.02.062

quantifies in full detail the distribution of wealth in the human society under consideration [27–29]. The Lorenz curve has two important advantages over the Gini index: it can code infinitely more information, and it can be visualized graphically.

The visualization of quantitative data, "*infographics*" [30–32], is of the utmost importance. On the one hand, visual pattern recognition is hard-wired into our brains by a very long evolution process of the human species. On the other hand, in the modern era we represent data by numbers, and number systems are a relatively new human invention which is not naturally intuitive to our brains. Thus, to comprehend a given set of numerical data, a visually meaningful representation of this data is indeed paramount. As said, "a picture is worth a thousand words", and this adage is doubly true when we are dealing with numbers rather than words.

The Lorenz curve has however a visualization drawback: the socioeconomic inequality conveyed by a given Lorenz curve is not straightforwardly evident from its graph. For example, given two Lorenz curves – each representing a different human society – it is not always obvious which of the societies is more egalitarian. Clearly, one can resort to the Gini index and check out which society has a smaller index. Yet this comes at the expense of collapsing a full-detail graphic visualization (Lorenz curves) to a pair of numbers (Gini indices).

This paper presets an alternative to the Lorenz curve: *hill curves*. As in the case of the Lorenz curve, the hill curves code an infinite amount of information regarding the distribution of wealth in the human society under consideration. Contrary to the case of Lorenz curves, the hill curves provide a most obvious and straightforward graphic visualization of socioeconomic inequality. The hill curves combine the best of the two worlds. For scientists knowledgeable of their mathematical construction, the hill curves are detailed quantitative scans of the rich–poor gaps in the human society under consideration. For the general public – even people with no quantitative skills whatsoever – the hill curves provide a most vivid picture of inequality.

And the applications of hill curves go far beyond the distribution of wealth in human societies. While the Gini index and the Lorenz curves were devised in order to measure socioeconomic inequality, they are actually measures of *statistical variability* that are applicable to *size distributions* at large. Indeed, be it a distribution of energies, a distribution of masses, a distribution of temperatures, a distribution of pollution levels, a distribution of city sizes, or a degree distribution of a graph—any distribution of non-negative sizes with a finite mean has its hill curves, and these curves are infinitely more informative than the distribution's *variance*.

Just as the Gini index is the most popular measure of socioeconomic inequality, the variance is the most widely applied measure of statistical variability. The 'benchmark' for both the Gini index and the variance is *determinism*. In the context of wealth distributions determinism means a fixed wealth for all, i.e. a purely communist socioeconomic structure. In the context of general size distributions determinism means a fixed size for all, i.e. no statistical variability. Both the Gini index and the variance are measures of the divergence from the benchmark of determinism—with the value zero characterizing determinism. Consequently, both the Gini index and the variance are crude measures—as they collapse a potentially infinite-dimensional information coded in an underlying size distribution into a number, a one-dimensional quantity.

Contrary to the Gini index and the variance, the hill curves can accommodate infinitely many degrees of freedom, and can represent them in a highly tangible graphical visualization. This paper, in the contemporary spirit of the big-data approach, and in the context of general size distributions, is a call for transcending from the one-dimensional Gini index and variance to the infinitely more informative hill curves.

In what follows we introduce and analyze the hill curves, and then discuss their application. Throughout the paper, $x = \varphi^{-1}(y)$ denotes the inverse function of a monotone real-valued function $y = \varphi(x)$ that is defined over a real range.

2. Hill curves

In this section we describe the construction, the computation, and the properties of hill curves. Based on the notion of hill curves, we further show how to 'normalize' the popular Gini index.

2.1. Construction

Consider a large human society whose distribution of wealth is represented by the *Lorenz curves* $y = L(x)$ and $y = \bar{L}(x)$ [27–29]. The Lorenz representation $y = L(x)$ of the wealth distribution has the following meaning: the low (poor) $100x\%$ of the society's population hold $100y\%$ of the society's overall wealth ($0 \le x, y \le 1$). Similarly, the Lorenz representation $y = \bar{L}(x)$ of the wealth distribution has the following meaning: the top (rich) $100x\%$ of the society's population hold $100y\%$ of the society's overall wealth ($0 \le x, y \le 1$). The two Lorenz curves initiate from the level zero ($L(0) = \bar{L}(0) = 0$), increase monotonically to the level one ($L(1) = \bar{L}(1) = 1$), and are coupled by the connection $L(1-x) + \bar{L}(x) = 1$ ($0 \le x \le 1$). Also, the Lorenz curve $y = L(x)$ is convex, the Lorenz curve $y = \bar{L}(x)$ is concave, and the two curves satisfy the ordering $0 \le L(x) \le x \le \bar{L}(x) \le 1$ ($0 \le x \le 1$). Schematic illustrations of the Lorenz curves are depicted in Fig. 1.

The *Lorenz set* is defined as the closure of the collection of the points, in the unit square, that are captured between the two Lorenz curves: $\{0 \le x, y \le 1 \mid L(x) \le y \le \bar{L}(x)\}$ [33]. The Lorenz set is contained in the unit square, it contains the diagonal line $y = x$ ($0 \le x, y \le 1$), and it is a convex set whose shape provides a geometric representation of the distribution of wealth. The 'smallest' Lorenz set – the diagonal line $y = x$ ($0 \le x, y \le 1$) – characterizes the socioeconomic extreme of *pure communism*: a society whose overall wealth is equally distributed among all its members. The 'largest' the Lorenz set