Physica A 429 (2015) 184–204

Contents lists available at ScienceDirect

Physica A

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

Measuring social inequality with quantitative methodology: Analytical estimates and empirical data analysis by Gini and k indices

Jun-ichi Inoue [a], Asim Ghosh [b,d,*], Arnab Chatterjee [b], Bikas K. Chakrabarti [b,c]

[a] Graduate School of Information Science & Technology, Hokkaido University, N14-W9, Kita-ku, Sapporo 060-0814, Japan

[b] Condensed Matter Physics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India

[c] Economic Research Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India

[d] Department of Biomedical Engineering and Computational Science, Aalto University School of Science, P.O. Box 12200, FI-00076, Finland

HIGHLIGHTS

- Different social inequalities are measured using Lorenz curve, Gini index and k index analytically.
- Use of suitable fits for characterization, inequality measures and compare with the analytical results.
- We provide general formulas to calculate inequality measures.
- Special technique is used to specify the crossover point between distributions.

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form 3 November 2014
Available online 19 February 2015

Keywords:
Social inequality
Gini and k-indices
Empirical data analysis
Mixtures of distributions

ABSTRACT

Social inequality manifested across different strata of human existence can be quantified in several ways. Here we compute non-entropic measures of inequality such as Lorenz curve, Gini index and the recently introduced k index analytically from known distribution functions. We characterize the distribution functions of different quantities such as votes, journal citations, city size, etc. with suitable fits, compute their inequality measures and compare with the analytical results. A single analytic function is often not sufficient to fit the entire range of the probability distribution of the empirical data, and fit better to two distinct functions with a single crossover point. Here we provide general formulas to calculate these inequality measures for the above cases. We attempt to specify the crossover point by minimizing the gap between empirical and analytical evaluations of measures. Regarding the k index as an 'extra dimension', both the lower and upper bounds of the Gini index are obtained as a function of the k index. This type of inequality relations among inequality indices might help us to check the validity of empirical and analytical evaluations of those indices.

1. Introduction

Humans are social beings and our social interactions are often complex. Social interactions in many forms produce spontaneous variations manifested as inequalities while at times these inequalities result out of continued complex

* Corresponding author at: Department of Biomedical Engineering and Computational Science, Aalto University School of Science, P.O. Box 12200, FI-00076, Finland.

E-mail addresses: jinoue@cb4.so-net.ne.jp, j_inoue@complex.ist.hokudai.ac.jp (J.-i. Inoue), asim.ghosh@saha.ac.in (A. Ghosh), arnabchat@gmail.com (A. Chatterjee), bikask.chakrabarti@saha.ac.in (B.K. Chakrabarti).

http://dx.doi.org/10.1016/j.physa.2015.01.082
0378-4371/

Table 1
Table showing examples of what X and Y can represent.

X	Y
People	Income, wealth
Article/paper	Citation
Institution/university	Citations
Institution/university	Funding
Candidate	Vote
City	Population
Student	Marks
Company	Employee

interactions among the constituent human units. The availability of a large body of empirical data for a variety of measures from human social interactions has made it possible to uncover the patterns and investigate the reasons for socio-economic inequalities. With tools of statistical physics as a core, researchers are incorporating the knowledge and techniques from various disciplines [1] like statistics, applied mathematics, information theory and computer science for a better understanding of the nature and origin of socio-economic inequalities that shape the humankind. Socio-economic inequality [2–5] is the existence of unequal opportunities and rewards for various social positions or statuses within the society. It usually contains structured and recurrent patterns of unequal distributions of goods, wealth, opportunities, and even rewards and punishments, and mainly measured in terms of *inequality of conditions*, and *inequality of opportunities*. *Inequality of conditions* refers to the unequal distribution of income, wealth and material goods. *Inequality of opportunities* refers to the unequal distribution of 'life chances' across individuals. This is reflected in measures such as level of education, health status, and treatment by the criminal justice system. Socio-economic inequality is responsible for conflict, war, crisis, oppression, criminal activity, political unrest and instability, and indirectly affects economic growth [6]. Traditionally, economic inequalities have been studied in the context of income and wealth [7–9], although it is also measured for many quantities, including energy consumption [10]. The study of inequality in society [11–14] is a topic of current focus and global interest and brings together researchers from various disciplines—economics, sociology, mathematics, statistics, demography, geography, graph theory, computer science and even theoretical physics.

Socio-economic inequalities are quantified in various ways. The most popular measures are absolute, in terms of indices, e.g., Gini [15], Theil [16], Pietra [17] indices. The alternative approach is a relative measure, in terms of probability distributions of various quantities, but the most of the above mentioned indices can be computed from the distributions. Most quantities often display broad distributions, usually lognormals, power-laws or their combinations. For example, the distribution of income is usually an exponential followed by a power law [18] (see Ref. [8] for other examples).

The Lorenz curve [19] is function which represents the cumulative proportion X of ordered individuals (from lowest to highest) in terms of the cumulative proportion of their size Y. Here, X can represent income or wealth, citation, votes, city population etc. Table 1 shows the typical examples of X and the corresponding Y. The Gini index (g) is defined as the ratio between the area enclosed between the Lorenz curve and the equality line, and the area below the equality line. If the area between (i) the Lorenz curve and the equality line is A, and (ii) that below the Lorenz curve is B (see Fig. 1), the Gini index is given by $g = A/(A + B)$. It is an useful measure for quantifying socio-economic inequalities. Besides these well-established measures, Ghosh et al. [20] recently introduced a different measurement called 'k index' ('k' stands for the extreme nature of social inequalities in Kolkata) defined as the fraction k such that $(1 - k)$ fraction of people or papers possess k fraction of income or citations respectively. In fact, this was discussed in Ref. [10] in the context of global energy consumption, where the authors found that the top 1/3 of the world population consumes 2/3 of the energy, which they called "the law of 1/3".

One of the well studied data in this respect is that of income (see Fig. 2). For USA, incomes calculated from the income tax data tabulated in the Internal Revenue Service (IRS) [21] for 1996–2011 were used. The probability distribution $P(w)$ of income w was computed for each year. For Denmark, similar data were used from the years 2000–2012 [22]. The g index is around 0.54–0.60 for USA and 0.34–0.38 for Denmark, while the k-index is around 0.69–0.71 for USA and 0.65–0.69 for Denmark.

When the probability distribution is described using an appropriate parametric function, one can derive these inequality measures as a function of those parameters analytically. In fact, several empirical evidence have been reported to show that the distributions can be put into a finite number of types. Most of them turn out to be a of mixture of two distinct parametric distributions with a single crossover point.

In this paper, we have characterized empirical data and the fitting forms have been treated analytically for comparison. We show in this paper that the distributions of population in socio-economic sciences can be put into several categories. We specify each of the distributions by appropriate parameters. We present the general form of the inequality measures, namely, Lorenz curve, Gini index g and k index for a class of distributions which can be expressed as a mixture of two distributions with a single crossover point. We check the values obtained from empirical calculations with those from analytical expressions. Especially, by minimizing the empirical and analytical values of the inequality measures, one can find an estimate of the crossover point which is usually determined by eye estimates. As a use of k index, both the lower and upper bounds of the Gini index are obtained as a function of k index by considering the k index as an 'extra dimension'.