Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

Socio-economic inequality: Relationship between Gini and Kolkata indices

Arnab Chatterjee [a], Asim Ghosh [b,*], Bikas K. Chakrabarti [a,c]

[a] Condensed Matter Physics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India
[b] Department of Computer Science, Aalto University School of Science, P.O. Box 15400, FI-0007 AALTO, Finland
[c] Economic Research Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India

HIGHLIGHTS

- Gini (g) index and Kolkata (k) indices of socioeconomic inequality are measured.
- Empirical data reveals a universal pattern for g–k relation.
- Approximate analytic solution is given to establish the relation between two indices.
- The study suggests an alternative way to find g index from k-index.

ARTICLE INFO

Article history:
Received 14 June 2016
Received in revised form 31 August 2016
Available online xxxx

Keywords:
Spatial inequality
Inequality indices
Data analysis

ABSTRACT

Socio-economic inequality is characterized from data using various indices. The Gini (g) index, giving the overall inequality is the most common one, while the recently introduced Kolkata (k) index gives a measure of $1-k$ fraction of population who possess top k fraction of wealth in the society. Here, we show the relationship between the two indices, using both empirical data and analytical estimates. The significance of their relationship has been discussed.

1. Introduction

Human social interactions often lead to complex dynamics. Repeated social interactions produce spontaneous variations which are manifested as inequalities at various levels. The availability of huge amount of empirical data for a plethora of measures of human social interactions has made it possible to uncover the patterns, analyze them and look for the reasons behind various socio-economic inequalities. Besides using tools of statistical physics, researchers are also bringing in knowledge and techniques from various other disciplines [1], e.g., statistics, applied mathematics, information theory and computer science to better the understanding of the precise nature (spatio-temporal) and origin of socio-economic inequalities prevalent in our society.

Socio-economic inequality [2–6] basically concerns the existence of unequal 'wealth' and 'fortunes' accumulated due to complex dynamics within the society. It usually contains structured and recurrent patterns of unequal distributions of goods, wealth, opportunities, and even rewards and punishments, and classically measured in terms of *inequality of conditions*, and *inequality of opportunities*. *Inequality of conditions* refers to the unequal distribution of income, wealth, assets and material goods. *Inequality of opportunities* refers to the unequal distribution of 'life chances'. This is reflected in measures like level

* Corresponding author.
E-mail addresses: arnabchat@gmail.com (A. Chatterjee), asim.ghosh@aalto.fi (A. Ghosh), bikask.chakrabarti@saha.ac.in (B.K. Chakrabarti).

http://dx.doi.org/10.1016/j.physa.2016.09.027
0378-4371/

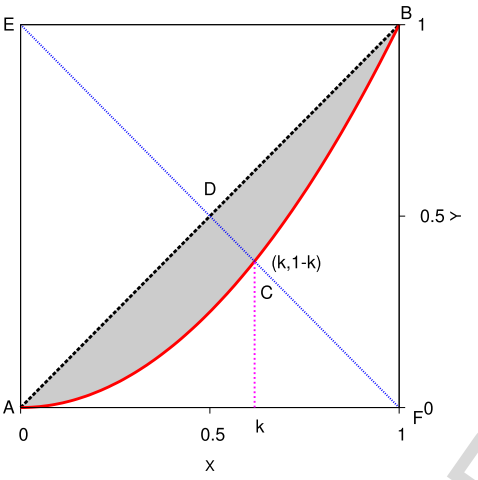

Fig. 1. Lorenz curve (in solid red) for a typical probability distribution function and the equality line (dotted black diagonal). Lorenz curve gives the cumulative fraction of 'wealth' possessed by the corresponding fraction of poorer population. g-index is given by area of the shaded region (normalized by the area of the triangle AFB). k-index is computed from the coordinate of the point of intersection C $(k, 1-k)$ of the Lorenz curve and the diagonal perpendicular to the equality line. Obviously, while g-index measures the overall inequality in the system, k-index gives the precise fraction k of wealth possessed by $1-k$ fraction of richer population. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of education, health status, treatment by the criminal justice system etc. Socio-economic inequalities are responsible for conflict, war, crisis, oppression, criminal activities, political instability and unrest, and that indirectly affects economic growth [7] of a region. Traditionally, economic inequalities have been extensively studied in the context of income and wealth [8–10], although it is also measured for many quantities like energy consumption [11]. The study of inequality in society [12–15] has always been very important, and is also a topic of current focus and immediate global interest, bringing together researchers across various disciplines—economics, sociology, mathematics, statistics, demography, geography, graph theory, computer science, and even theoretical physics.

Socio-economic inequalities are quantified in numerous ways. The most detailed measures are of course given by probability distributions of various quantities. What is usually observed is that most quantities display broad distributions—most common are log-normals, power-laws or their combinations. For example, the distribution of income is usually an exponential followed by a power law [16,9]. However, such distributions can widely differ in their forms and subtleties, and as such they are not quite convenient to handle. This leads to the introduction of various indices like the Gini [17], Theil [18], Pietra [19] and other socio-geometric indices [20,21], which try to characterize various geometric features of these distributions.

The degree of inequality is most commonly measured by the Gini index. One considers the Lorenz curve [22], representing the cumulative proportion X of ordered (from poorest to richest) individuals (entries) in terms of the cumulative Y of their wealth. Y can of course represent income or wealth of individuals but it can as well represent citation of articles, votes in favor of candidates, population of cities etc. The Gini index (g), defined as the ratio of the area enclosed between the Lorenz curve and the equality line, to that below the equality line, is the most common measure to quantify socio-economic inequality, taking values 0 for absolute equality and 1 for absolute inequality from a given statistical distribution. If the area between (i) the Lorenz curve and the equality line is represented as $\mathcal{A}$, and (ii) that below the Lorenz curve as $\mathcal{B}$ (See Fig. 1), the Gini index is $g = \mathcal{A}/(\mathcal{A}+\mathcal{B}) = 2\mathcal{A}$. Ghosh et al. [23] recently introduced the Kolkata index (symbolizing the extreme nature of social inequalities in Kolkata) or 'k-index', which is defined as the fraction k such that poorest $(1-k)$ fraction of people possess k fraction of income [24–26]. In fact, another recently proposed measure, the perpendicular width index I_{PW} [20] can be shown to be equal to $\sqrt{2}(2k-1)$.

2. Empirical findings on $g - k$ relationship

A large variety of socio-economic data suggest that there exists a simple relation between the two seemingly different inequality measures. We analyzed citations of papers published from academic institutions and journals (data from ISI Web of Science [27] and reported in Ref. [25]), consumption expenditure data of India [28], Brazil [29,30], Italy [31], income data from USA [32], voting data from open list proportional elections [33] of Italy, Netherlands and Sweden, *first past the post* election data for Indian Parliamentary elections and Legislative Assembly elections [34], United Kingdom [35], Canada [36], Bangladesh [37], Tanzania [38], and city population data from Ref. [39]. See Tables B.1–B.9 in Appendix B for details.

The relation is perfectly linear for smaller values while the curve becomes non-linear as g or k approaches unity, the limit of extreme inequality (Fig. 2). The most intriguing part is that the data from a variety of sources hardly depart from

Please cite this article in press as: A. Chatterjee, et al., Socio-economic inequality: Relationship between Gini and Kolkata indices, Physica A (2016), http://dx.doi.org/10.1016/j.physa.2016.09.027