Physica A 410 (2014) 30–34

Contents lists available at ScienceDirect

Physica A

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

Inequality in societies, academic institutions and science journals: Gini and k-indices

Asim Ghosh [a,*], Nachiketa Chattopadhyay [b], Bikas K. Chakrabarti [a,c]

[a] Condensed Matter Physics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700 064, India
[b] Sampling & Official Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700 108, India
[c] Economic Research Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700 108, India

HIGHLIGHTS

- Science journals and academic institutions are shown to have high inequality in citation distributions.
- The inequalities are much higher than that for wealth distributions.
- A new measure of inequality is proposed apart from the Gini index.
- Both of these measures suggest a universal nature of academic inequalities in terms of citations.

ARTICLE INFO

Article history:
Received 13 February 2014
Received in revised form 5 May 2014
Available online 13 May 2014

Keywords:
Social inequality
Lorenz curves
Inequality indices
Power-laws

ABSTRACT

Social inequality is traditionally measured by the Gini-index (g). The g-index takes values from 0 to 1 where $g = 0$ represents complete equality and $g = 1$ represents complete inequality. Most of the estimates of the income or wealth data indicate the g value to be widely dispersed across the countries of the world: g values typically range from 0.30 to 0.65 at a particular time (year). We estimated similarly the Gini-index for the citations earned by the yearly publications of various academic institutions and the science journals. The ISI web of science data suggests remarkably strong inequality and universality ($g = 0.70 \pm 0.07$) across all the universities and institutions of the world, while for the journals we find $g = 0.65 \pm 0.15$ for any typical year. We define a new inequality measure, namely the k-index, saying that the cumulative income or citations of $(1 - k)$ fraction of people or papers exceed those earned by the fraction (k) of the people or publications respectively. We find, while the k-index value for income ranges from 0.60 to 0.75 for income distributions across the world, it has a value around 0.75 ± 0.05 for different universities and institutions across the world and around 0.77 ± 0.10 for the science journals. Apart from above indices, we also analyze the same institution and journal citation data by measuring Pietra index and median index.

1. Introduction

Social inequality is often measured by the Gini-index or Gini coefficient (g) [1–6] obtained from the area between the diagonal (equality line) and the Lorenz curve, given by the plot of cumulative fraction (w) of income or wealth (when ordered

* Corresponding author.
E-mail addresses: asimghosh066@gmail.com, asim.ghosh@saha.ac.in (A. Ghosh), nachiketa@isical.ac.in (N. Chattopadhyay), bikask.chakrabarti@saha.ac.in (B.K. Chakrabarti).

http://dx.doi.org/10.1016/j.physa.2014.05.026
0378-4371/

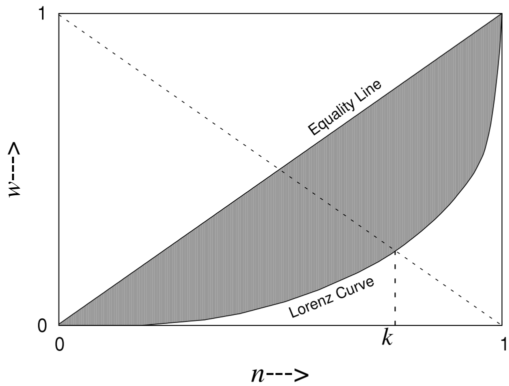

Fig. 1. The Lorenz curve plots the cumulative fraction w of the (a) income or wealth or (b) citations when ordered from the lowest to the highest income, wealth (for (a)) or citations (for (b)) in any particular year for any country (for (a)) or any institution (for (b)), against the cumulative fraction (n) of the people in the country (for (a)) or the fraction of papers published in that year (for (b)) sharing that income or wealth (for (a)) or citations (for (b)) respectively. The diagonal starting from the origin then gives the equality line (corresponding to uniform income or wealth that year per head of the population of that country or citations per paper published that year from the institution). The (normalized) ratio of the area of the shaded region (between the equality line and the Lorenz curve) and of the triangle formed by equality line (effectively twice the area value of the shaded region) gives the Gini-index value (g). The ordinate k (on the n axis) of the intersection point of the Lorenz curve and the other diagonal (perpendicular to the equality line) gives the k-index. This k-index value gives another measurement (similar to the h-index for individual scientists): k-value of a society says that $1 - k$ fraction of people (for (a)) or papers (for (b)) possess more income, wealth (for (a)) or citations (for (b)) than those earned by the rest k fraction of people (for (a)) or papers (for (b)).

from lowest to highest income or wealth) against the corresponding cumulative fraction (n) of people sharing them in any society (at any particular time or year). In a similar way, we measure here the inequalities in the output of the various academic institutions and universities by determining g-values obtained from Lorenz curves of the institutions obtained by plotting the cumulative fraction of the citations of the papers (when ordered from lowest to highest citations) published in any year from that institution, against the corresponding cumulation fraction of papers sharing those citations (see Fig. 1). The Gini index or g-value is again given by twice the (normalized) area of the region (shaded in Fig. 1) between the equality line or diagonal through the origin and the Lorenz curve. We introduce then a new inequality measure, the k-index (k for Kolkata; in view of the extreme nature of social inequalities in Kolkata [7]) which is given by the coordinate value k in the n-axis in Fig. 1 of the cutting point of the Lorenz line with the diagonal orthogonal to the equality line. As one can see in the case of income inequality, it says the fraction $(1 - k)$ of people earn more than that earned by fraction k of people in the country or society. In the case of an academic institution the k-value says that the fraction $1-k$ of their papers published (in a year) from that institution have more citations than those earned by k fraction of papers. As is obvious from Fig. 1, $g = 0$ corresponds to complete equality (Lorenz curve merges with the diagonal) while $g = 1$ corresponds to extreme inequity. The corresponding values of the k-index are $k = 1/2$ for $g = 0$ for complete equality and $k = 1$ for $g = 1$ for limiting (extreme) value of inequality. In the income or wealth inequality context, Pareto had already observed [8] (see also Ref. [5]) that a tiny fraction (typically less than 20%) of (rich) people possess 80% of the total wealth of the nations. The k-index defined here gives a more quantitative measure of this inequality. Also in the context of academic institutions or universities, the k-index gives a (normalized) complementary measure of the h-index [9] equivalent for the respective institution for that year; h-index of a scientist gives the number h of his or her papers, each of which has at least h citations.

Apart from k-index and g-index, we analyze these citation data for institutions and journals by measuring two other inequality indices introduced recently: Pietra index or p-index [10–12] and median index or m-index [6]. p-index is defined as the maximal vertical distance between the Lorenz curve and the line of perfect equality in Fig. 1. m-index is given by $2m' - 1$ where $L(m') = 1/2$ ($L(x)$ denoting the Lorentz curve). Values for p-index and m-index both range from 0 to 1; the value 0 represents complete equality while value 1 represents extreme inequality in the society. In this article, we measure inequality in income, citations for papers published in institutions and journals as examples for inequality in society, by using some newly defined quantities.

2. Analysis and results

We give here a few estimated values of g and k indices of different countries in different years in Table 1 for income inequalities across the countries of the world, with the data taken from Refs. [13,14]. In Tables 2 and 3 we give the estimates of the values of g, k, p and m indices for different institutions and universities across the world in different times (years). These estimates are made from the corresponding Lorenz curves drawn numerically from the respective data sets. For citations of the papers, the data are taken from ISI web of science [15] and are counted up to the year 2013, while the publications (corresponding to the institutions of origin or to the journal where published) are for the individual years of publication (see e.g., Tables 2–4). Assuming that the Lorenz curve can be approximated by two discontinuous straight lines forming a triangle with vertex opposite to the equality line given by intersection point of the Lorenz curve and the diagonal perpendicular to