

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

Jun-ichi Inoue,^{1,*} Asim Ghosh,^{2,†} Arnab Chatterjee,^{2,‡} and Bikas K. Chakrabarti^{2,3,§}

¹Graduate School of Information Science & Technology, Hokkaido University, N14-W9, Kita-ku, Sapporo 060-0814, Japan.

²Condensed Matter Physics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India.

³Economic Research Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India.

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.

I. 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 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 I 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

* Email: jinoue@cb4.so-net.ne.jp, j'inoue@complex.ist.hokudai.ac.jp

† Email: asim.ghosh@saha.ac.in

‡ Email: arnabchat@gmail.com

§ Email: bikask.chakrabarti@saha.ac.in

TABLE I. 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

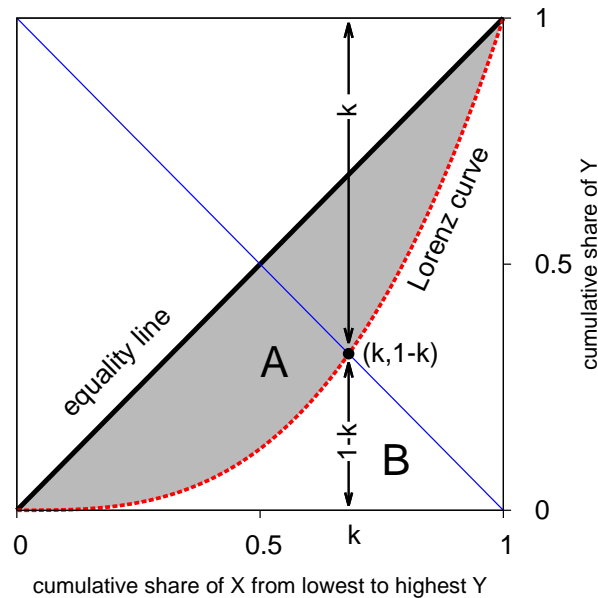


FIG. 1. Schematic representation of Lorenz curve, Gini index g and k index. The dashed red line stands for the Lorenz curve and the black solid line represents perfect equality. The area enclosed by the equality line and the Lorenz curve is A and that below the Lorenz curve is B . The Gini index is given by $g = A/(A + B)$. The k index is given by the abscissa of the intersection point of the Lorenz curve and $Y = 1 - X$.

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

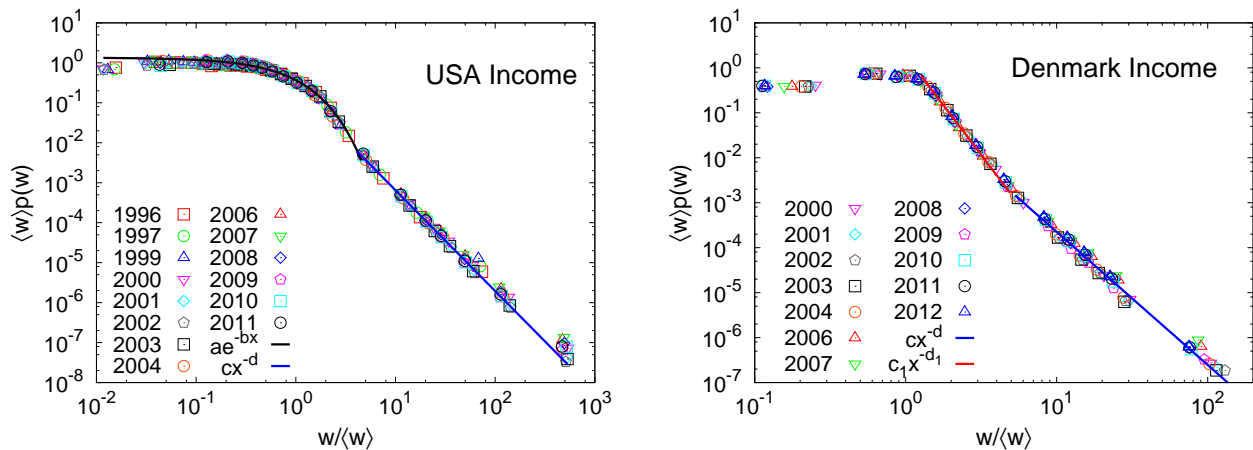


FIG. 2. Left: Income distribution $p(w)$ for USA [21] for several years, rescaled by the average income $\langle w \rangle$. The lower income range fits to an exponential form $a \exp(-bx)$ with $b = 1.3$ while the high income range fits to a power law decay cx^{-d} with exponent $d = 2.52$. $x = w/\langle w \rangle$ is the rescaled dimensionless income. Right: Same for Denmark [22]. The middle and upper ranges fit to different power laws cx^{-d} with $d = 2.96$ and $c_1x^{-d_1}$ with $d_1 = 4.41$.

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’. This type of inequality relation among the inequality indices might help us to check the validity of empirical and analytical estimates of these indices.

This paper is organized as follows. In Sec. II, we introduce the basics and some generic properties of our measures – Lorenz curve, Gini and k indices. In Sec. III, we provide the general formulas of the inequality measures for the empirically observed distributions. In Sec. IV, regarding the k index as an ‘extra dimension’, both the lower and upper bounds of the Gini index are obtained as a function of k index. In Sec. V, we report our results. Here, we provide some empirical findings. Out of the data we considered, we found six categories of distributions. We observed that most of the data can be described by a mixture of two distinct distributions with a crossover point. Here we give numerical evaluations of our measures. Then, we compare the empirical and analytical evaluations of inequality measures. Minimizing the gap between two results obtained by different ways, we infer the best possible crossover point for a given data set. We conclude with a summary and discussions.

II. BASICS AND GENERIC PROPERTIES OF INEQUALITY MEASURES

In this section, we introduce the measures to quantify the degree of social inequality, namely, Lorenz curve, Gini index and k index. Then, the generic properties are explained.

A. Lorenz curve

The Lorenz curve is given as a relationship between the cumulative distribution and the cumulative first moment of $P(m)$. Namely, for the mean and the normalized first moment of $P(m)$,

$$X(r) = \int_{m_0}^r P(m) dm, \quad Y(r) = \frac{\int_{m_0}^r mP(m) dm}{\int_{m_0}^{\infty} mP(m) dm}. \quad (1)$$

The Lorenz curve is given as a set of $(X(r), Y(r))$, where we assume that the $P(m)$ is defined in $[m_0, \infty)$. In Fig. 1, we show the typical behavior of Lorenz curve by dashed line.

The intuitive meaning of the Lorenz curve is as follows: the cumulative proportion X of ordered (from lowest to highest) individuals hold the cumulative proportion Y of wealth. For sake of simplicity, we will use ‘individuals’ for

attributing X and ‘wealth’ for attributing Y , for the simple reason that Lorenz curve, Gini index etc. were historically introduced in the context of income/wealth, but in principle the attributes X and Y can be any of the combinations mentioned in Table. I. Hence, when all individuals take the same amount of wealth, say m_* , we have

$$P(m) = \delta(m - m_*), \quad m_0 < m_* < \infty, \quad (2)$$

and one obtains

$$X(r) = \int_{m_0}^r \delta(m - m_*) dm = \Theta(r - m_*), \quad (3)$$

$$Y(r) = \frac{\int_{m_0}^r m \delta(m - m_*) dm}{\int_{m_0}^{\infty} m \delta(m - m_*) dm} = \frac{m_* \Theta(r - m_*)}{m_*} = X(r). \quad (4)$$

where $\Theta(x)$ is a unit step function defined by

$$\Theta(x) = \begin{cases} 1, & x \geq 1, \\ 0, & x < 1. \end{cases} \quad (5)$$

Thus we have $Y = X$ as the ‘perfect equality line’ (see thick line in Fig. 1). This means that X fraction of people takes X fraction of total wealth in society.

On the other hand, when the total wealth in the society consisting of N persons is concentrated to a few persons, namely,

$$P(m) = (1 - \varepsilon)\delta_{m,0} + \varepsilon\delta_{m,1}, \quad (6)$$

where $\varepsilon \sim \mathcal{O}(1/N)$ and we assume that the total amount of wealth is normalized as 1, we obtain $X(r) = 1 - \varepsilon + \varepsilon\delta_{r,1}$ and $Y(r) = \delta_{r,1}$. Hence, $Y = 1$ iff $X = r = 1$ and $Y = 0$ otherwise, and the Lorenz curve is given as ‘perfect inequality line’ by $Y = \delta_{X,1}$ where $\delta_{x,y}$ is a Kronecker’s delta (see Fig. 1).

B. Gini index

For a given Lorenz curve, the Gini index is evaluated as twice of area between the curve $(X(r), Y(r))$ and perfect equality line $Y = X$. The area is shown in the shaded part (named ‘A’) in Fig. 1. Namely, it reads

$$g = 2 \int_0^1 (X - Y) dX = 2 \int_{r_0}^{\infty} (X(r) - Y(r)) \frac{dX}{dr} dr, \quad (7)$$

where we should keep in mind that $X^{-1}(0) = r_0, X^{-1}(1) = \infty$ should hold. In a graphical way, the Gini index is given as a ratio of two areas (‘A’ and ‘B’) by $g = A/(A + B)$. From the definition, the Gini index g is zero for perfect equality and unity for perfect inequality. It should be noted that the Gini index may be evaluated analytically when the distribution of population is obtained in a parametric way.

In fact, in the references [23, 24], in the context of analysis of waiting time (duration) of time series, the Gini index was analytically calculated for a parametric distribution. In [23, 24], the so-called Weibull distribution was selected to quantify the inequality of duration t of financial time series. The Weibull distribution is described by

$$P_{\mu,\eta}(t) = \frac{\mu t^{\mu-1}}{\eta} \exp\left(-\frac{t^\mu}{\eta}\right). \quad (8)$$

The resulting Gini index for the Weibull distribution is given as

$$g = 1 - \left(\frac{1}{2}\right)^\mu. \quad (9)$$

The Weibull distribution $P_{\mu,\eta}(t)$ is identical to exponential distribution $\sim e^{-t/\eta}$ for $\mu = 1$, which means that the point process specified by exponentially distributed duration t between events follows a Poisson process. Hence, we are confirmed that $g = 1/2$ for $\mu = 1$, and the deviation of the Gini index from $1/2$ for arbitrary process shows to what extent the resulting time series is different from randomly generated events. This result was also reported in Ref. [25].

C. k index

The k index which was recently introduced is defined as the value of X -axis for the intersection between the Lorenz curve and a straight line $Y = 1 - X$. Namely, for the solution of equation

$$X(r) + Y(r) = 1, \quad (10)$$

say $r_* = Z^{-1}(1)$, $Z(r) \equiv X(r) + Y(r)$, the k index is given by

$$k = X(r_*). \quad (11)$$

From the definition, the k index denotes the situation in which k fraction of people shares totally $(1 - k)$ fraction of the wealth [20]. Obviously, the k index takes 1/2 for perfectly equal society, whereas it takes 1 for perfectly unequal society. This measure was used in Ref. [10] in the context of global energy consumption.

The k index is obviously easier to estimate by eyes in comparison with the Gini index (shaded area A in Fig. 1). We will also discuss another use of the k index by regarding the k as an extra dimension in Sec. IV.

Besides g and k indices, Pietra's p index [17] and median index or m index [26] has been used as inequality measures derived from the Lorenz curve. The p index is defined as the maximal vertical distance between the Lorenz curve and the line of perfect equality $Y = X$ (in Fig. 1), whereas m index is given by $2m - 1$ for the solution of $Y(m) = 1/2$, where we assumed that the Lorenz curve is given as $(X(r), Y(r))$ using a parameter r . It might be important for us to discuss these two indices, however, in this paper we limit ourselves to g and k indices.

III. GENERAL FORMULA

In this section, we describe the general formula for the Lorenz curve, Gini index and k index for the distribution $P(m)$. In this paper, we calculate the g and k indices using different theoretical distribution functions, and sometimes combinations of two of them. It is very common to find that the probability distributions of many quantities (wealth, income, votes, citations etc.) fit to more than one theoretical function depending on the range:

$$P(m) = F_1(m)\theta(m, m_\times) + F_2(m)\Theta(m - m_\times), \quad (12)$$

$\theta(m, m_\times) \equiv \Theta(m) - \Theta(m - m_\times)$, where m_\times is the crossover point.

The functions $F_1(m)$ and $F_2(m)$ are suitably normalized and computed for their continuity at m_\times . We will compute the functional fits to the empirical distributions of several quantities, compute the g and k indices, and compare with the theoretical values computed from the fitting distributions.

A. Lorenz curve

Our main purpose here is to derive a general form of the Lorenz curve for the distribution having the form Eq. (12). The resulting form of the Lorenz curve is given by

$$Y = \begin{cases} \frac{R_1[Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}X)]}{R_1(m_\times) + R_2(m_\times)}, & 0 \leq X \leq \frac{Q_1(m_\times)}{Q_1(m_\times) + Q_2(m_\times)}, \\ 1 - \frac{R_2[Q_2^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}(1-X))]}{R_1(m_\times) + R_2(m_\times)}, & \frac{Q_1(m_\times)}{Q_1(m_\times) + Q_2(m_\times)} \leq X \leq 1, \end{cases} \quad (13)$$

where we defined the cumulative 'persons' and 'wealth' of the distributions $F_1(m)$ and $F_2(m)$ as

$$Q_1(r) = \int_{m_0}^r F_1(m)dm, \quad Q_2(r) = \int_r^\infty F_2(m)dm, \quad (14)$$

$$R_1(r) = \int_{m_0}^r mF_1(m)dm, \quad R_2(r) = \int_r^\infty mF_2(m)dm. \quad (15)$$

The derivation is given in Appendix A.

It should be noticed that when the mean and the first moment of $F_1(m)$ and $F_2(m)$ are identical in such a way as $Q_1(r) = R_1(r)$ and $Q_2(r) = R_2(r)$; for instance, $F_1(m)$ and $F_2(m)$ are both $P(m) = \delta(m - m_*)$, $0 < m_* < \infty$, we have

$$Y = \frac{R_1[Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}X)]}{R_1(m_\times) + R_2(m_\times)} = 1 - \frac{R_2[Q_2^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}(1-X))]}{R_1(m_\times) + R_2(m_\times)} = X, \quad (16)$$

which is nothing but the Lorenz curve for perfect equal society.

The above argument is very general and independent of specific choice of the distributions. However, it is important for us to check the validity of the above general form Eq. (13) for well-known limiting cases without crossover, namely, $m_\times \gg 1$ or $m_\times = m_0$.

1. Uniform distribution

For this purpose, we first examine a single uniform distribution $F_1(m) = 1/a$, $m_\times = a \gg 1$, $m_0 = 0$. For this case, we find $Q_1(m_\times) = 1$, $Q_1(m_0) = Q_2(\infty) = Q_2(m_\times) = 0$ and

$$Q_1(r) = \frac{r}{a}, \quad R_1(r) = \frac{r^2}{2a}. \quad (17)$$

Those lead to $Q_1^{-1}[\{Q_1(m_\times) + Q_2(m_\times)\}X] = aX$, $R_1(Q_1^{-1}[\{Q_1(m_\times) + Q_2(m_\times)\}X]) = R_1(aX) = aX^2/2$ and $R_1(m_\times) = a/2 \gg 1$, $R_2(m_\times) = 0$. Hence, we obtain the Lorenz curve from the first branch of Eq. (13) as

$$Y = X^2. \quad (18)$$

2. Power law distribution

We next consider the case of $m_\times \gg 1$ and $F_1(m) = (\alpha - 1)m^{-\alpha}$, $m_0 = 1$, namely, for a single power law distribution. Then, the first branch in Eq. (13) survives and we have $Q_1(m_\times) = 1$, $R_1(m_\times) = (\alpha - 1)/(\alpha - 2)$, $Q_2(m_\times) = R_2(m_\times) = 0$, and $Q_1(r) = 1 - r^{1-\alpha} = (Q_1(m_\times) + Q_2(m_\times))X = X$, namely,

$$Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}X) = (1 - X)^{\frac{1}{1-\alpha}}. \quad (19)$$

Therefore, using $R_1(r) = (\alpha - 1)(1 - r^{2-\alpha})/(\alpha - 2)$, we have

$$R_1[Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}X)] = \left(\frac{\alpha - 1}{\alpha - 2}\right) \{1 - (1 - X)^{\frac{2-\alpha}{1-\alpha}}\}. \quad (20)$$

Inserting these staffs into the first branch of Eq. (13), we finally obtain

$$Y = 1 - (1 - X)^{\frac{2-\alpha}{1-\alpha}}. \quad (21)$$

3. Lognormal distribution

We next consider the case of $m_\times \gg 1$, $m_0 = 0$ and $F_1(m)$ is a lognormal distribution given by

$$F_1(m) = \frac{1}{\sqrt{2\pi}\sigma m} \exp\left[-\frac{(\log m - \mu)^2}{2\sigma^2}\right]. \quad (22)$$

For this case, the first branch in Eq. (13) is selected and $Q_1(m_\times) = 1$, $R_1(m_\times) = e^{\mu + \sigma^2/2}$, $Q_2(m_\times) = R_2(m_\times) = 0$. We also have

$$Q_1(r) = H\left(\frac{\mu - \log r}{\sigma}\right), \quad R_1(r) = e^{\mu + \frac{\sigma^2}{2}} H\left(\frac{\mu + \sigma^2 - \log r}{\sigma}\right). \quad (23)$$

where we defined

$$H(x) = \int_x^\infty Dz, \quad Dz \equiv \frac{dz}{\sqrt{2\pi}} e^{-z^2/2}. \quad (24)$$

This reads

$$Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}X) = \exp[\mu - \sigma H^{-1}(X)], \quad (25)$$

$$R_1[Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}X)] = e^{\mu + \frac{\sigma^2}{2}} H(\sigma + H^{-1}(X)). \quad (26)$$

Substituting these all staffs to the first branch of Eq. (13), finally we obtain

$$Y = H(\sigma + H^{-1}(X)). \quad (27)$$

This reads $H^{-1}(Y) = \sigma + H^{-1}(X)$. Hence, we recover the result for perfect equality $Y = X$ in the limit of $\sigma \rightarrow 0$.

B. Gini index

For the general distribution with a crossover Eq. (12), we can derive the general form of the Gini index as follows.

$$g = \frac{Q_1(m_\times)^2 - Q_1(m_0)^2 + Q_2(\infty)^2 - Q_2(m_\times)^2}{\{Q_1(m_\times) + Q_2(m_\times)\}^2} - \frac{2(S_1(m_0, m_\times) + T_2(m_\times))}{\{Q_1(m_\times) + Q_2(m_\times)\}\{R_1(m_\times) + R_2(m_\times)\}}, \quad (28)$$

where we defined

$$S_1(m_0, m_\times) = \int_{m_0}^{m_\times} R_1(r) \frac{dQ_1(r)}{dr} dr; \quad T_2(m_0) = \int_{m_\times}^{\infty} R_2(r) \frac{dQ_2(r)}{dr} dr. \quad (29)$$

We should keep in mind that we replace ∞ in the upper bound of integral in $T_2(m_0)$ and $Q_2(\infty)$ by $M < \infty$ when the distribution function $F_2(m)$ possesses a cut-off M . The detail of the derivation is explained in Appendix A.

To check the validity of the general form Eq. (28), we examine the case of a single power law distribution and a single lognormal distribution as we did for checking the Lorenz curve.

1. Uniform distribution

We first examine a single uniform distribution $F_1(m) = 1/a, m_\times = a \gg 1, m_0 = 0$. Taking into account the result $S_1(m_0, m_\times) = 2a/3, T_2(m_\times) = 0$, we obtain from the general form Eq. (28) as

$$g = \frac{1}{3}. \quad (30)$$

2. Power law distribution

For a power law distribution, we should set $m_\times \gg 1, m_0 = 1$ and we have $S_1(m_0, m_\times) = (\alpha - 2)/(2\alpha - 3), T_2(m_\times) = 0$ Taking into account the result and $Q_1(m_0) = Q_2(\infty) = 0$, we have

$$g = \frac{1}{2\alpha - 3}. \quad (31)$$

Therefore, the g for $\alpha = 3$ is identical to the result of uniform distribution $g = 1/3$.

3. Lognormal distribution

We next consider the case of a single lognormal distribution Eq. (22). Here we notice

$$S_1(m_0, m_\times) = e^{\mu + \frac{\sigma^2}{2}} \int_{-\infty}^{\infty} Dz H(z + \sigma), \quad T_2(m_\times) = 0 \quad (32)$$

and $Q_1(m_0) = Q_2(\infty) = 0$. Thus, we obtain

$$g = 1 - 2 \int_{-\infty}^{\infty} Dz H(z + \sigma). \quad (33)$$

Using the fact $\int_{-\infty}^{\infty} Dx H(x) = \{H(-\infty)\}^2/2 = 1/2$, the above expression is rewritten in terms of the integral of difference between two complementary error functions as

$$g = 2 \int_{-\infty}^{\infty} Dx \{H(x) - H(x + \sigma)\}, \quad (34)$$

and one can confirm that we recover the result for perfect equality case in the limit of $\sigma \rightarrow 0$ as $g = 0$. On the other hand, in the limit of $\sigma \rightarrow \infty$, the mode of the lognormal behaves as $e^{\mu - \sigma^2} \rightarrow 0$. Hence, the distribution might possess the form $P(m) = (1 - \varepsilon)\delta_{m,0} + \varepsilon\delta_{m,1}$, where ε is a small fraction of people in the society, and here we suppose the maximum value of wealth is normalized as 1 (see also Eq. (6)). Therefore, the Gini index should be identical to the value for perfectly unequal society, and actually we have

$$g = 2 \int_{-\infty}^{\infty} Dx H(x) = 2 \times \frac{1}{2} \{H(-\infty)\}^2 = 1. \quad (35)$$

C. k index

The general form of the k index for the distribution $P(m)$ (Eq. (12) is given as a solution of the following equation:

$$\begin{cases} Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}k) = R_1^{-1}(\{R_1(m_\times) + R_2(m_\times)\}(1-k)), & 0 \leq k \leq \frac{Q_1(m_\times)}{Q_1(m_\times) + Q_2(m_\times)}, \\ Q_2^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}(1-k)) = R_2^{-1}(\{R_1(m_\times) + R_2(m_\times)\}k), & \frac{Q_1(m_\times)}{Q_1(m_\times) + Q_2(m_\times)} < k \leq 1. \end{cases} \quad (36)$$

The derivation is given in Appendix A.

It should be noticed that for $Q_1(r) = R_1(r)$ and $Q_2(r) = R_2(r)$; for instance, $F_1(m)$ and $F_2(m)$ are both $P(m) = \delta(m - m_*)$, $0 < m_* < \infty$, we have $k = 1 - k$. It reads $k = \frac{1}{2}$ which is the k index for perfectly equal society.

1. Uniform distribution

To check the validity, we next examine the case of $F_1(m)$ is a uniform distribution $F_1(m) = 1/a$, $m_\times = a \gg 1$, $m_0 = 0$. From Eq. (17), we have $Q_1^{-1}(k) = ak$, $R_1^{-1}(a(1-k)/2) = a\sqrt{1-k}$. Thus, from the first branch of Eq. (36), we obtain $k = \sqrt{1-k}$. That is

$$k = \frac{-1 + \sqrt{5}}{2} \sim 0.62. \quad (37)$$

2. Power law distribution

We next consider the case in which $F_1(m)$ follows a power law distribution under the condition $m_\times \gg 1$, $m_0 = 1$. From Eq. (19), we have $Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}k) = (1-k)^{\frac{1}{1-\alpha}}$. Using $R_1^{-1}(\{R_1(m_\times) + R_2(m_\times)\}(1-k)) = k^{\frac{1}{2-\alpha}}$, the k index is determined as a solution of

$$k = (1-k)^{\frac{2-\alpha}{1-\alpha}}. \quad (38)$$

In particular, we obtain for the choice of $\alpha = 3$ as $k = \frac{-1 + \sqrt{5}}{2} \sim 0.62$. We should notice that the value is exactly the same as that of uniform distribution. Hence, the k index for a power law distribution with exponent $\alpha = 3$ is identical to that of a uniform distribution as we saw it for g index.

3. Lognormal distribution

On the other hand, for a lognormal as $F_1(m)$ and $m_\times \gg 1$, $m_0 = 0$, we obtain from Eq. (25) as $Q_1^{-1}(\{Q_1(m_\times) + Q_2(m_\times)\}k) = \exp[\mu - \sigma H^{-1}(k)]$. Using the relation $R_1^{-1}(\{R_1(m_\times) + R_2(m_\times)\}(1-k)) = \exp[\mu + \sigma^2 - \sigma H^{-1}(1-k)]$, we obtain the k index as a solution of

$$H^{-1}(1-k) - H^{-1}(k) = \sigma. \quad (39)$$

It should be noticed that from the definition of the lognormal distribution, we find $P(m) = \delta(m - e^\mu)$ in the limit of $\sigma \rightarrow 0$ (in this limit, the median, mean and mode of the lognormal take the same value e^μ). Namely, all person possesses the same wealth. In this limit, Eq. (39) leads to $H^{-1}(1-k) = H^{-1}(k)$, and this gives the k index for perfectly equal society $k = 1/2$. On the other hand, in the limit of $\sigma \rightarrow \infty$, the solution of Eq. (39) is $k = 1$. We confirm these two limits in Fig. 3 (right).

IV. UPPER AND LOWER BOUNDS FOR g IN TERMS OF k

In the previous sections, we have introduced the Lorenz curve and g and k indices, and discussed several analytic forms and their properties. These two indices g and k are both derived from the Lorenz curve and we might use some geometrical (graphical) interpretations for them. In particular, it might be very useful for us to derive the inequality for both inequality measures, that is, ‘inequality of inequalities’.

Here we derive the upper and lower bounds for the Gini index g in terms of the k index. The use of this lower and upper bounds is one of the advantage of the k index. For human eyes, it is very difficult to estimate the ‘area’

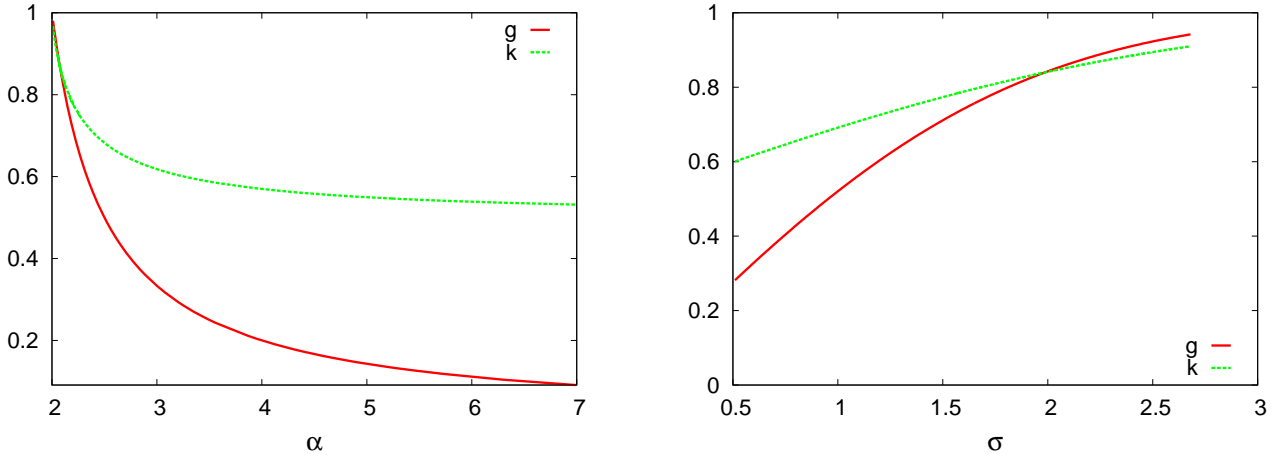


FIG. 3. The g and k indexes for a single power law distribution (left) and a single lognormal distribution with $\mu = 1$ (right: the case (a) in Fig. 6) as a function of α and σ , respectively. For a single lognormal distribution, the g goes to 0 in the limit of $\sigma \rightarrow 0$, whereas the k goes to $1/2$, each of which is a limit of ‘perfect equality’. On the other hand, both g and k go to 1 in the limit of $\sigma \rightarrow \infty$.

surrounded by $Y = X$ and the Lorenz curve, that is, a half of Gini index $g/2$, whereas the k index, it is relatively easier for us to estimate the value by eyes because each axis X and Y is calibrated in $[0, 1]$. In this sense, once we obtain both the bounds as a function of k , we can easily estimate the g from both the bounds through the k index. Additionally, if we can make the bound a tighter one, the estimation will be closer to the exact value. The argument that follows is just an application of basic Euclidean geometry.

Let us denote the origin $(0, 0)$ as ‘O’, $(1, 1)$ as ‘A’, $(k, 1 - k)$ as ‘B’, and the intersection of perfect quality line $Y = X$ and $Y = 1 - X$, that is, $(1/2, 1/2)$ as ‘C’ in Fig. 1 (or Fig. 4 for uniform distribution as a special case). Then, we compare the area of the triangle OAB and the shaded area which gives a half of the Gini index $g/2$. Obviously, as long as the Lorenz curve is convex, the area of the triangle OAB, $k - 1/2$, is smaller than that of the shaded area $g/2$. Hence, we have

$$g \geq 2k - 1. \quad (40)$$

The above equality is valid for $k = 1$ (perfect unequal) and $k = 1/2$ (perfect equal).

The convexity of the Lorenz curve is proved as follows. From the definition of the Lorenz curve Eq. (1), we immediately have

$$\frac{dY}{dX} = \frac{\frac{dY}{dr}}{\frac{dX}{dr}} = \frac{rP(r)}{Y_0P(r)} = \frac{r}{Y_0}, \quad (41)$$

and we conclude

$$\frac{d^2Y}{dX^2} = \frac{d}{dX} \left(\frac{dY}{dX} \right) = \frac{d}{dr} \left(\frac{r}{Y_0} \right) \cdot \frac{1}{\frac{dX}{dr}} = \frac{1}{Y_0P(r)} > 0, \quad (42)$$

where we defined $Y_0 \equiv \int_{m_0}^{\infty} mP(m)dm$ as a positive constant. Therefore, the Lorenz curve is convex at any point of $X(r)$, and the inequality Eq. (40) is actually satisfied for any set of inequality measures g and k for a given set of data sets or parameters which specify the probability density $P(m)$.

We next derive the upper bound of the Gini index by means of the k index. To derive the bound, we consider the tangential line of the Lorenz curve at $(k, 1 - k)$ in Fig. 1 (or Fig. 4 for uniform distribution as a special case), that is,

$$Y = \xi(k)(X - k) + 1 - k, \quad (43)$$

$$\xi(k) \equiv \left. \frac{dY}{dX} \right|_{X=k} = \frac{X^{-1}(k)}{Y_0}. \quad (44)$$

Then, let us define the intersections of this tangential line and $X = 1$, namely, $(1, (1 - k)(1 + \xi(k)))$ as ‘D’ and $Y = 0$, $(k - (1/\xi(k))(1 - k), 0)$ as ‘E’, respectively. Then, the area of a quadrilateral OADEO is larger than or equal to a half

of the Gini index. Hence, using this fact, we can derive another inequality. The area is easily calculated and we have

$$g \leq 2k(2-k) - 1 - (1-k)^2 \left(\xi(k) + \frac{1}{\xi(k)} \right). \quad (45)$$

We should notice that the upper bound of the Gini index gives 1 for $k = 1$, whereas for $k = 1/2$, we have

$$g \leq \frac{1}{2} - \frac{1}{4} \left(\xi(1/2) + \frac{1}{\xi(1/2)} \right). \quad (46)$$

From the definition Eq. (44), $\xi(1/2) = dY/dX|_{X=k} = 1$ because $k = 1/2$ means the perfect equality line $Y = X$. Thus, we conclude $g \leq 0$ (which means $g = 0$ from the definition of g) for $k = 1/2$. Therefore, the equality in Eq. (45) should hold if and only if $k = 1$ (perfect unequal) and $k = 1/2$ (perfect equal).

From the argument above, we finally obtain the following inequality

$$\phi(k) \equiv 2k - 1 \leq g \leq 2k(2-k) - 1 - (1-k)^2 \left(\xi(k) + \frac{1}{\xi(k)} \right) \equiv \psi(k, \xi(k)), \quad (47)$$

where we should notice that the lower bound $\phi(k)$ is dependent on the detail of the distribution $P(m)$ through k index itself, whereas the upper bound $\psi(k, \xi(k))$ depends on the wealth distribution $P(m)$ through k and the slope $\xi(k)$.

To check the validity of the inequality Eq. (47), we first consider a uniform distribution. The situation is shown in Fig. 4. As we already saw, the Lorenz curve is given by $Y = X^2$ (see Eq. (18)) which gives $\xi(k) = 2k$. Thus, the

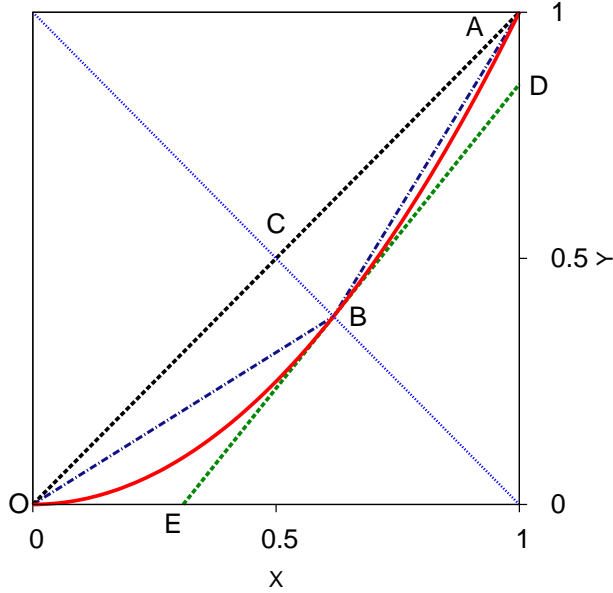


FIG. 4. The case of uniform distribution. The Lorenz curve is simply given by $Y = X^2$ (see Eq. (18)). As the k index is given by $k = (-1 + \sqrt{5})/2$ (see Eq. (37)), the tangential line is obtained as $Y = (-1 + \sqrt{5})(X - (\sqrt{5} - 1)/2) + (3 - \sqrt{5})/2$, which touches with the Lorenz curve at point B. We compare the area of the quadrilateral OADEO and the half of the Gini index g . Apparently the former is bigger than the latter, which gives the upper bound $\psi(k, \xi(k))$ of g . On the other hand, when we compare the area of triangle OAB and $g/2$, the former is smaller than the latter, which gives the lower bound $\phi(k)$ of g .

upper bound is simply given by

$$\psi(k, \xi(k)) = -2k^3 + 2k^2 + 2k - 1 - \frac{(1-k)^2}{2k}. \quad (48)$$

Substituting the k index for a uniform distribution $k = (-1 + \sqrt{5})/2$ (see Eq. (37)) into the bounds in Eq. (47) $\phi(k)$ and $\psi(k, 1/2)$, we obtain

$$2k - 1 = \sqrt{5} - 2 = 0.2360 < g < \psi(k, 1/2) = \frac{12 - 5\sqrt{5}}{2} = 0.4098. \quad (49)$$

We should notice that the exact value of $g = 1/3 = 0.3333$ for a uniform distribution is lying on the interval suggested by inequality Eq. (47).

We next check the bounds for a single exponential distribution: $P(m) = \beta e^{-\beta m}$. It is easy to derive the Lorenz curve and we obtain

$$Y = X + (1 - X) \log(1 - X). \quad (50)$$

Therefore, it is independent of the parameter β . For the Lorenz curve, the k index is obtained as a solution of equation $2k = 1 - (1 - k) \log(1 - k)$, and it leads to $k = 0.6822$. This was also reported in Ref. [10], where the authors argued that this fraction is close to $2/3$, so that $1/3$ of the world population consumes $2/3$ of produced energy.

Now, taking into account $\xi(k) = -\log(1 - k)$, we obtain

$$\psi(k, -\log(1 - k)) = -2k^2 + 4k - 1 + (1 - k)^2 \left\{ \log(1 - k) + \frac{1}{\log(1 - k)} \right\} \quad (51)$$

and the inequality for g in terms of k is given by

$$\phi(k) = 2k - 1 = 0.3644 < g < \psi(k, -\log(1 - k)) = 0.5940. \quad (52)$$

Exact value of the Gini index for a single exponential distribution $\beta e^{-\beta m}$ is evaluated as $g = 1/2$, which is of course independent on β and a special case of Weibull distribution Eqs. (8), (9) with $\mu = 1$. Hence, we are confirmed that the above inequality on g actually works for a single exponential distribution.

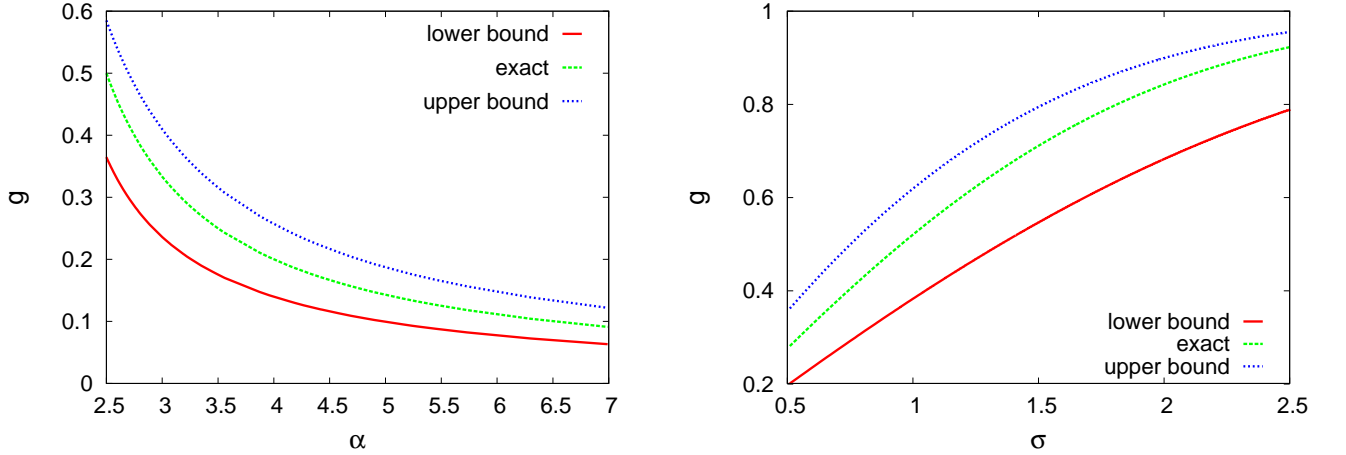


FIG. 5. The upper and lower bounds of the Gini index for a single power law distribution (left) and a single lognormal distribution (right) with $\mu = 1$ as a function of parameters α and σ respectively. The exact values of g are the same as shown in Fig. 3.

For the above two cases, both bounds $\phi(k), \psi(k, \xi(k))$ are independent of the parameters appearing in the wealth distribution $P(m)$. It should be stressed that the bounds are also evaluated even for parameter-dependent cases and $\phi(k), \psi(k, \xi(k))$ are obtained as a function of parameters. As we shall show in the next section, most of the distributions of population in social science is categorized into several classes of distribution. Especially, those are described by a mixture of lognormal and power law distributions. Hence, it might be worthwhile for us to discuss the bounds for each of the distributions independently. We first consider the case of power law distribution: $P(m) = (\alpha - 1) m^{-\alpha}$. As we already obtain the Lorenz curve, we have the slope $\xi(k) = dY/dX|_{X=k}$ as

$$\xi(k) = \left(\frac{2 - \alpha}{1 - \alpha} \right) (1 - k)^{\frac{1}{1 - \alpha}}. \quad (53)$$

Hence, we have

$$\psi \left(k, \left(\frac{2 - \alpha}{1 - \alpha} \right) (1 - k)^{\frac{1}{1 - \alpha}} \right) = 2k(2 - k) - 1 - (1 - k)^2 \left\{ \left(\frac{2 - \alpha}{1 - \alpha} \right) (1 - k)^{\frac{1}{1 - \alpha}} + \left(\frac{1 - \alpha}{2 - \alpha} \right) (1 - k)^{-\frac{1}{1 - \alpha}} \right\}. \quad (54)$$

On the other hand, for a single lognormal distribution, we have

$$\xi(k) = \frac{X^{-1}(k)}{e^{\mu + \frac{\sigma^2}{2}}}. \quad (55)$$

Therefore, we have the upper bound as a function of k as

$$\psi\left(k, \frac{X^{-1}(k)}{e^{\mu + \frac{\sigma^2}{2}}}\right) = 2k(2-k) - 1 - (1-k)^2 \left\{ \frac{e^{\mu + \frac{\sigma^2}{2}}}{X^{-1}(k)} + \frac{X^{-1}(k)}{e^{\mu + \frac{\sigma^2}{2}}} \right\}. \quad (56)$$

In Fig. 5, we draw the upper and lower bounds for a single power law distribution (left) and a single lognormal distribution (right) with $\mu = 1$ as a function of parameters α and σ , respectively. From this figure, we find that the inequality Eq. (47) detect the exact evaluation of the Gini index.

We should notice that both upper and lower bounds can be improved easily by considering a ‘polygon’ surrounding or including the area of a half Gini index $g/2$. Using the procedure, one can improve the bounds recursively and systematically. Eventually, the both bounds are expected to be very closed to the true g .

V. RESULTS FOR MIXTURE OF DISTRIBUTIONS

In the previous section, we introduced Gini and k indices and discuss the generic properties. From the definition of these measures, we always evaluated the values empirically from a finite number of data set. However, as we showed in Sec. III, it is easy to calculate the value analytically when the distribution of population is described by parametric distribution such as a uniform, power law and lognormal distributions. In fact, in the previous section, we derived the measures for these distribution functions. Turning now to the situation of reality, the distribution of population like wealth, number of citation, etc. is well fitted to a mixture of uniform, power-law and lognormal distributions.

A. Empirical data

In this paper, we focus on the following types of socio-economic data: (i) voting (ii) citations of different science journals, and (iii) population of cities and municipalities.

We use the voting data for open-list proportional elections from several countries of Europe (data taken from Ref. [27]). The number of votes v_i of a candidate is divided by the average number of votes v_0 of all candidates in his/her party list. We focus on the probability distribution $P(v/v_0)$ of the quantity v/v_0 , known as the ‘performance’ of a candidate. We use data for Italy, Netherlands and Sweden.

For citations to journals, we collected data from ISI Web of Science [28], citations gathered until a certain date by all articles/papers published in a particular year, for a (i) few scientific journals (PRL = Physical Review Letters, CPL = Chemical Physics Letters, PRA = Physical Review A, PNAS = Proceedings of the National Academy of Sciences USA, Lancet, BMJ = British Medical Journal, NEJM = New England Journal of Medicine), and (ii) universities/institutions (University of Oxford, University of Cambridge, University of Tokyo, University of Melbourne). We computed the probability distribution of citations $p(c)$, and found the corresponding scaling collapses for similar categories, by rescaling with the average number of citations $\langle c \rangle$.

We also collected data for city sizes for Brasil [29], municipalities of Spain [30] and Japan [31]. We computed the probability distribution of city/ municipality population $p(s)$, and rescaled them with the average population $\langle s \rangle$.

In Fig. 6, we show broad distributions of the above quantities and their fitting. The approach we take is simple: (i) we plot the data in log-log scale; (ii) if a part of the data falls on a straight line, we declare it a power law; (iii) if a portion of the data falls on a curved line, we try to fit it to a parabola, and if that succeeds, declare it a lognormal law. For the data we have considered, this procedure works reasonably well.

From this Figure, we are confirmed that the distribution which is well-fitted to these empirical evidences falls into six categories, namely, (a) a single lognormal, (b) a single lognormal with a power law tail, (c) uniform with a power law tail, (d) uniform with a lognormal tail, (e) a mixture of power laws, (f) a single power law with a lognormal tail.

Therefore, it is worth while for us to prepare the formula to calculate the measures analytically for those six cases.

B. Analytic formulas for six categories and relationship between g and k

We plot the typical behavior of Lorentz curve, g and k indexes by the explicit formulas for the cases of (b)-(f) in Fig. 7. The formulas and the details are given in Appendix A.

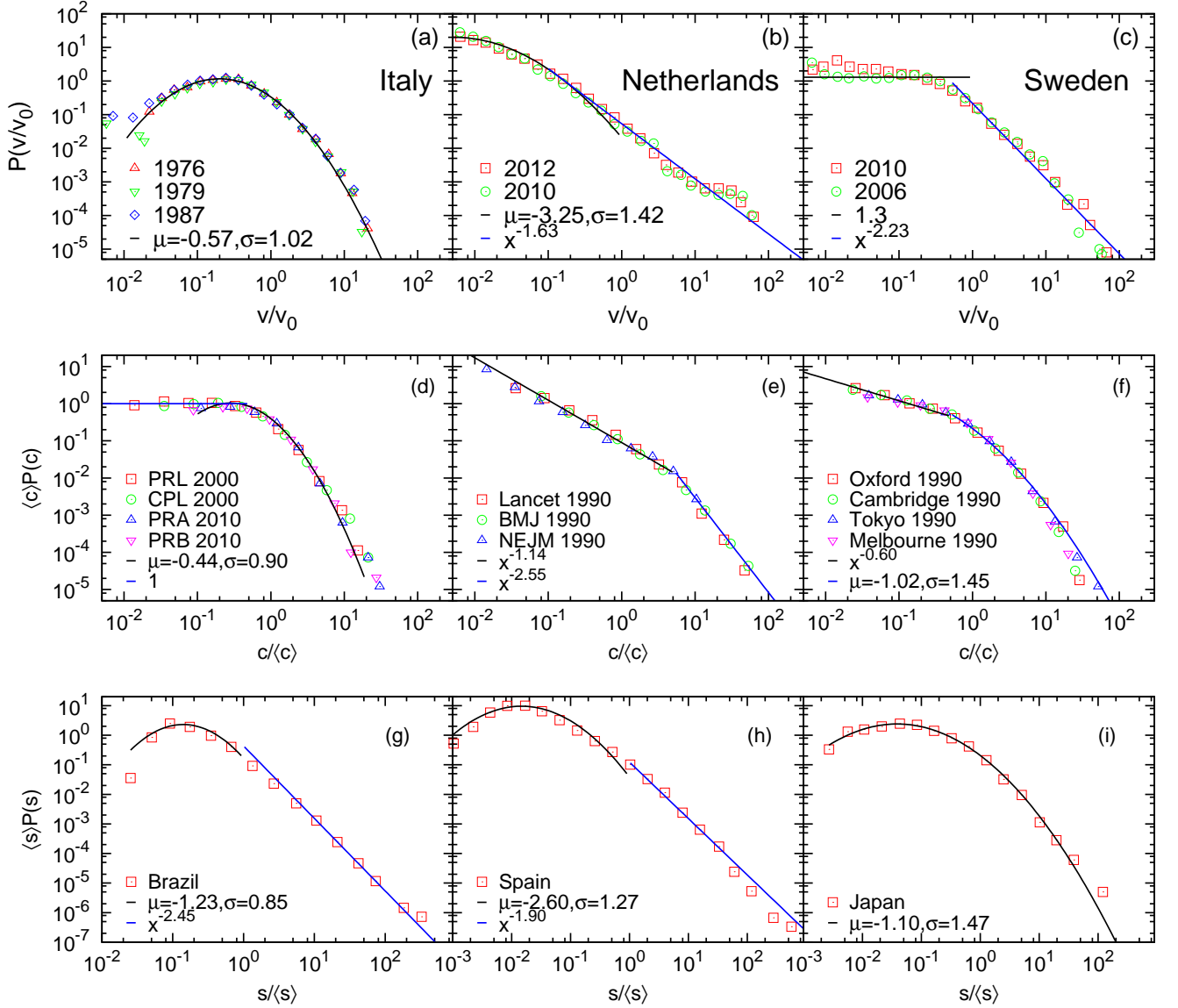


FIG. 6. Figure showing broad distributions of different quantities and their fitting. (a) Distribution of performance (v/v_0) for candidates in open list proportional elections in Italy for several years, fitted to lognormal function. Data taken from Ref. [27]. (b) for proportional elections with semi-open lists in Netherlands, fitted to lognormal followed by power law distribution. (c) for proportional elections with semi-open lists in Sweden, fitted to uniform followed by power law distribution. (d) Distribution of citations for journals, fitted to uniform distribution followed by lognormal [28]. (e) double power law. (f) power law, followed by lognormal. Population of (g) cities in Brazil [29], the same as case (b), (h) municipalities of Spain [30], the same as case (b), (i) municipalities of Japan [31], the same as (a).

In practice, it is convenient for us to clarify the relationship between g and k indices. As we already showed, these two indices are both dependent on the crossover point m_\times as $g(m_\times)$ and $k(m_\times)$. It is difficult for us to obtain the relation analytically. However, one can obtain it numerically when we consider m_\times as ‘time’ and also consider the ‘trajectory’ in the g - k space. In Fig. 7 (lower right), we show the trajectories for five cases. From this panel, we confirmed that g as a function of k is single-valued function only for the case (e) for the parameters taken in these plots. We should also notice that the Gini index g changes almost linearly as a function of k . Hence, one can infer the value of g by using the relationship when the parameter sets including m_\times are known beforehand. On the other hand, for the other cases, g are piecewise multi-valued functions. Therefore, we should keep in mind that for a given value of k , there are several candidates for the corresponding g value depending on the crossover point m_\times .

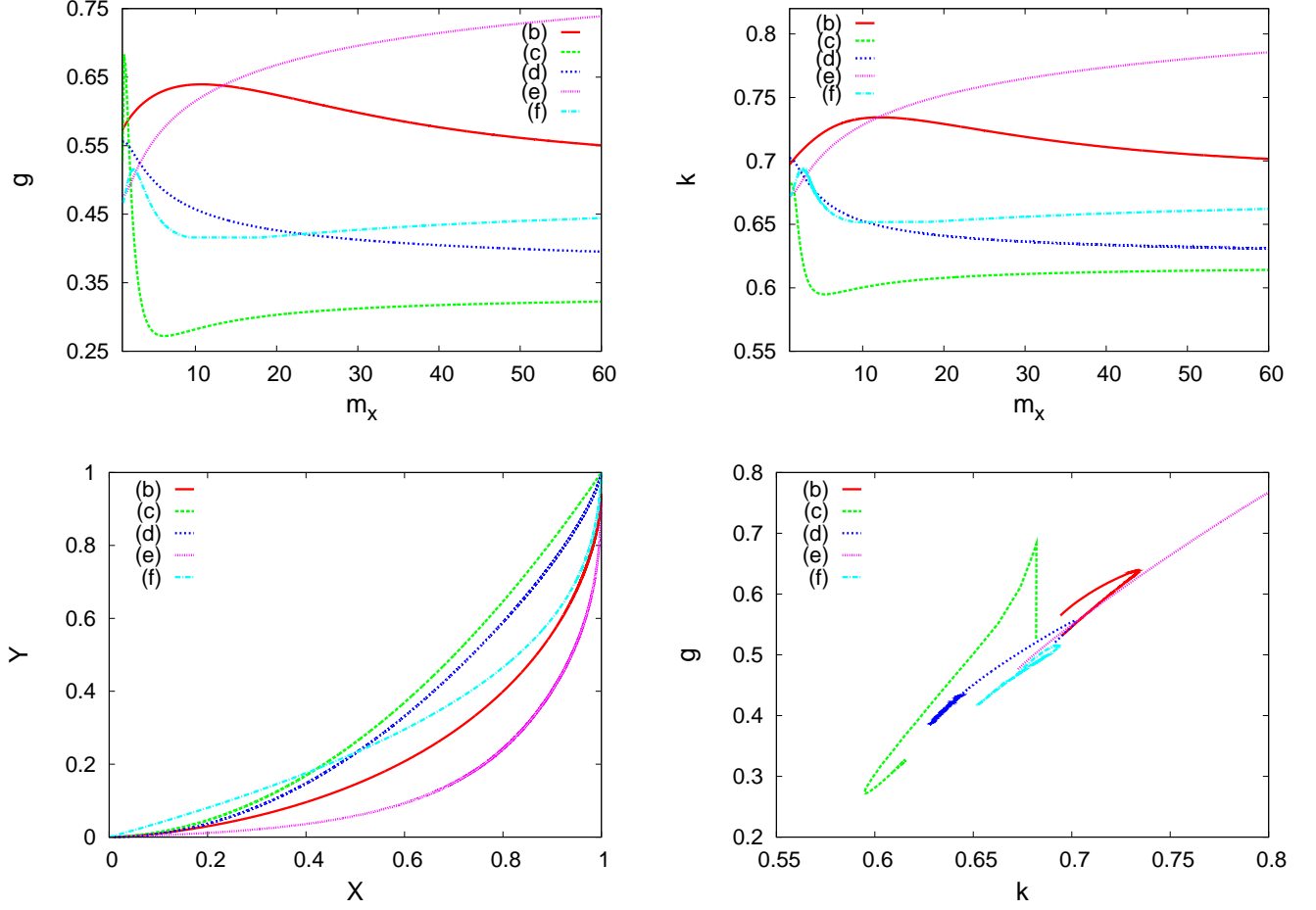


FIG. 7. Clockwise from the upper left, we show the m_x -dependence of the g and k indices, the relationship between g and k indices for five cases (b)-(f) (refer to Fig. 6), and the Lorenz curve. We set $\sigma = \mu = 1$ for lognormal distributions appearing in (b), (d), (f) and $\alpha = 2.5$ for power law distribution in (b), (c). In the case (c), we set $a = 1$ as a constant for uniform distribution. In the case (e), the two exponents for power law distributions are set to $\alpha = 1.14$ and $\beta = 2.55$.

C. Gap between empirical and analytical values of measures

To compare the empirical and analytical estimates for g and k indices whose distribution functions are shown in Fig. 6, we need to calculate those measures from empirical data sets. It is important for us to bear in mind that the measures could be calculated independently from the analytic formulas from data sets.

Let us consider the data set of wealth for N persons: $m_1 \leq m_2 \leq \dots \leq m_N$. Then, the Lorenz curve is given by

$$X(r) = \frac{r}{N}, \quad Y(r) = \frac{\sum_{i=1}^r m_i}{\sum_{i=1}^N m_i} = \frac{1}{\mu N} \sum_{i=1}^r m_i, \quad (57)$$

where we defined the empirical mean $\mu = (1/N) \sum_{i=1}^N m_i$. Using the general definition of the Gini index Eq. (7), one can obtain the g using the following non-parametric way as

$$\hat{g} = \frac{2}{\mu N^2} \sum_{r=1}^N r m_r - \frac{(N+1)}{N}. \quad (58)$$

The detail of derivation is explained in Appendix B.

On the other hand, empirical k index, say \hat{k} is given by

$$\hat{k} = \frac{r_*}{N}, \quad (59)$$

TABLE II. Estimates of Gini (g) and k index for distributions of different quantities: voting data (Taken from Ref. [27]), citation data for journals and institutions [28] and sizes of cities and municipalities [29–31]. The functional fits and their ranges are mentioned. where \hat{g}, \hat{k} are empirical estimates while g and k are computed from the analytical functions. N is the number of data and $\langle \cdot \rangle$ denotes the empirical mean $\mu = (1/N) \sum_{i=1}^N m_i$.

Country	Year	\hat{g}	g	\hat{k}	k	m_\times	$m < m_\times$	$m > m_\times$	N	v_0
Italy	1976	0.5593		0.7077			log-normal		5839	1.0
	1979	0.5463	0.5292	0.7014	0.6948	-	$\mu = -0.57; \sigma = 1.02$	-	7153	1.0
	1987	0.5720		0.7144					8620	1.0
Netherlands	2010	0.9406		0.9214			log-normal	power law	7229	1.0
	2012	0.9250	0.8038	0.9071	0.8935	0.20	$\mu = -3.25; \sigma = 1.42$	$\alpha = 1.64$	8889	1.0
Sweden	2006	0.6903		0.7650			uniform	power law	5150	1.0
	2010	0.7374	0.6825	0.7842	0.7315	0.50	$1/a = 1.3$	$\alpha = 2.24$	9053	1.0

Journals/ Institutions	Year	\hat{g}	g	\hat{k}	k	m_\times	$m < m_\times$	$m > m_\times$	N	$\langle c \rangle$
PRL	2000	0.5859		0.7154					3124	72.21
CPL	2000	0.5788		0.7123			uniform	log-normal	1512	28.44
PRA	2010	0.5271	0.5214	0.6895	0.6880	0.30	$1/a = 0.95$	$\mu = -0.44; \sigma = 0.90$	1410	27.62
PNAS	2010	0.4616		0.6641					2698	117.01
Lancet	1990	0.8448		0.8410			power law	power law	3232	27.71
BMJ	1990	0.8840	0.8808	0.8662	0.8660	5.0	$\alpha = 1.14$	$\beta = 2.55$	2847	12.48
NEJM	1990	0.8536		0.8498					1684	69.28
Oxford	1990	0.7276		0.7769					2147	39.10
Cambridge	1990	0.7366		0.7791			power law	log-normal	2616	42.74
Tokyo	1990	0.6834	0.4755	0.7564	0.6664	0.50	$\alpha = 0.60$	$\mu = -1.02; \sigma = 1.46$	4196	25.77
Melbourne	1990	0.6772		0.7515					1131	26.83

Country	Year	\hat{g}	g	\hat{k}	k	m_\times	$m < m_\times$	$m > m_\times$	N	$\langle s \rangle$
Brazil	2012	0.7270	0.6253	0.7795	0.7275	1.0	log-normal $\mu = -1.23; \sigma = 0.86$	power law $\alpha = 2.45$	5570	34825.23
Spain	2011	0.8661	0.8451	0.8560	0.8469	1.0	log-normal $\mu = -2.60; \sigma = 1.27$	power law $\alpha = 1.90$	8116	5814.50
Japan	2010	0.7192	0.7014	0.7738	0.7689	-	log-normal $\mu = -1.10; \sigma = 1.47$	-	1720	74451.95

where r_* is the solution of the following equation

$$\frac{r_*}{N} = 1 - \frac{\sum_{i=1}^{r_*} m_i}{\mu N}. \quad (60)$$

Of course, it is hard to find the solution with precision for finite N , we might use in practice

$$r_* = \arg \min_r \left| \frac{r}{N} - 1 + \frac{\sum_{i=1}^r m_i}{\mu N} \right|. \quad (61)$$

From the Table II, we found that there is a finite gap between empirical and analytical values of measures. When we assume that the exponent such as $\alpha, \beta, \mu, \sigma$ are precisely determined by means of maximum likelihood estimate, one can estimate the crossover point m_\times by minimizing the gaps $\Delta g(m_\times), \Delta k(m_\times)$. Namely, the cost function for the estimation could be constructed by

$$\Delta_g(m_\times) = (\hat{g} - g(m_\times))^2, \quad \Delta_k(m_\times) = (\hat{k} - k(m_\times))^2, \quad (62)$$

where \hat{g}, \hat{k} are empirical estimates, whereas $g(m_\times), k(m_\times)$ are the analytical expressions as a function of the crossover point m_\times . We show the result in Fig. 8 for the case of (e) (left) and (g) (right). Then minimum point is estimate by means of minimization of the gap. For the case of (e), we find $m_\times = 5.16$ for minimization of Δg and $m_\times = 3.74$ for minimization of Δk , and the resulting values of g and k are $g = 0.8844$ and $k = 0.8416$, respectively. On the other hand, for the case of (g), we have $m_\times = 0.765$ for minimization of Δg and $m_\times = 0.767$ for minimization of Δk , which lead to the corrected estimates $g = 0.6297$ and $k = 0.7281$.

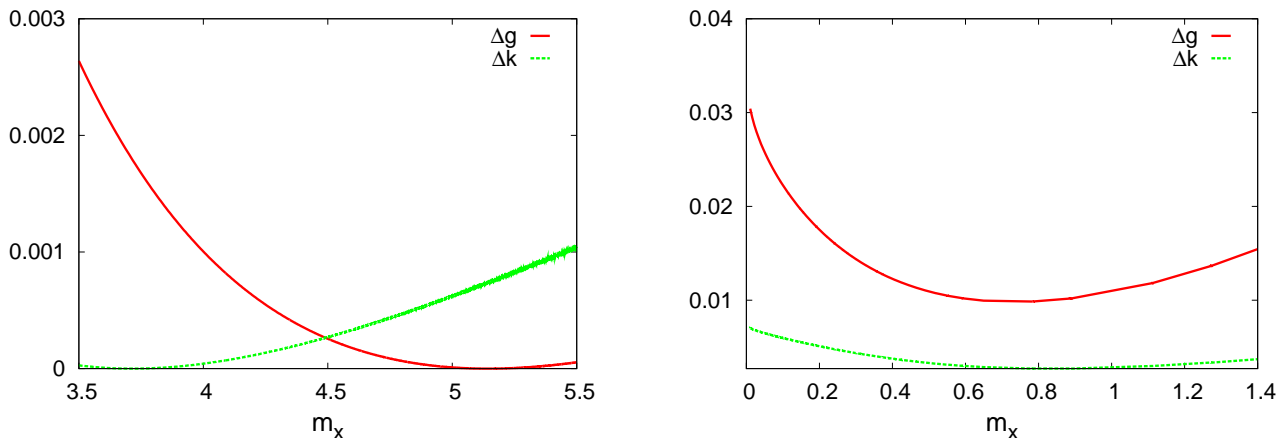


FIG. 8. The gaps $\Delta g, \Delta k$ as a function of the crossover point m_x for the case of (e) (left) and (g) (right). For (e), we chose \hat{g}, \hat{k} of ‘Lancet’ because N is largest among three of (Lancet, BMJ, NEJM).

VI. SUMMARY AND DISCUSSIONS

The probability distributions of several socio-economic quantities showing inequality have broad distributions. In Sec. II, we presented the general form of the inequality measures, by computing the Lorenz curve, and using that to compute the Gini and k indices for a class of distributions. From the empirical data we analyzed, we showed that the distributions can be put into several categories and each of these can be specified by an appropriate parametric distribution. In fact, we found six categories of distributions, most of which are a mixture of two distinct distributions with a crossover point. In Sec. III, we computed the general formulas of the inequality measures for the combinations of functions as observed from empirical distributions, and compared them with those from analytical calculations. In Sec. IV, we considered 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 relation between inequality indices might help us to check the validity of empirical and analytical evaluations of these indices. In Sec. V, we reported our results. We provided numerical evaluations of our measures, and compared their empirical and analytical values. By minimizing the gap between two results obtained in different ways, we provide the estimates of the best possible crossover point for a given data set.

Socio-economic inequality is a topic of major concern [32], drawing attention of researchers across various disciplines. Researchers have always concentrated on (i) characterizing empirical data and thereby computing inequality measures like Gini index, and (ii) modeling the origins of broad distributions. Our paper focuses on extensively computing the Gini index g and the newly introduced k index from single analytical distributions or combinations of them, which fit well to empirical data, and also compare the results with those calculated directly from the empirical data sets themselves. Our proposed quantitative methodology to estimate the crossover point between two functional fits to empirical data could also prove to be useful, beyond the realm of inequality research. While the much studied Gini index gives an overall measure of the inequality, the k index tells us that the cumulative wealth of $(1 - k)$ fraction of individuals are held by k fraction of individuals. The k index extends the observations done more than a century ago by Pareto who reported that 80% of land in Italy was owned by 20% of the people [33].

Acknowledgement

J.I. was financially supported by Grant-in-Aid for Scientific Research (C) of Japan Society for the Promotion of Science (JSPS) No. 2533027803 and Grant-in-Aid for Scientific Research (B) of 26282089, Grant-in-Aid for Scientific Research on Innovative Area No. 2512001313. He also thanks Saha Institute of Nuclear Physics for their hospitality during his stay in Kolkata. B.K.C. and A.C. acknowledges support from B.K.C.’s J. C. Bose Fellowship and Research Grant.

Appendix A: Derivation of the general forms of inequality measures

Here we drive the general form Eq. (13)(28) and Eq. (36). for the mixture of two normalized distribution Eq. (12). From the definition of Lorenz curve, we have

$$X(r) = \frac{Q_1(r)}{Q_1(m_\times) + Q_2(m_\times)} \theta(r, m_\times) + \left\{ 1 - \frac{Q_2(r)}{Q_1(m_\times) + Q_2(m_\times)} \right\} \Theta(r - m_\times) \quad (\text{A1})$$

$$Y(r) = \frac{R_2(r)}{R_1(m_\times) + R_2(m_\times)} \theta(r, m_\times) + \left\{ 1 - \frac{R_2(r)}{R_1(m_\times) + R_2(m_\times)} \right\} \Theta(r - m_\times) \quad (\text{A2})$$

For r of active $\theta(r, m_\times) = 1$, we have $(Q_1(m_\times) + Q_2(m_\times))X = Q_1(r)$, namely, $r = Q_1^{-1}[(Q_1(m_\times) + Q_2(m_\times))X]$, and this reads

$$Y = \frac{R_1(Q_1^{-1}[(Q_1(m_\times) + Q_2(m_\times))X])}{R_1(m_\times) + R_2(m_\times)}, \quad 0 \leq X \leq \frac{Q_1(m_\times)}{Q_1(m_\times) + Q_2(m_\times)}. \quad (\text{A3})$$

On the other hand, for r of active $\Theta(r - m_\times) = 1$, we have $(1 - X)(Q_1(m_\times) + Q_2(m_\times)) = Q_2(r)$, namely, $r = Q_2^{-1}[(Q_1(m_\times) + Q_2(m_\times))(1 - X)]$ and Y is given by

$$Y = 1 - \frac{R_2^{-1}[Q_2^{-1}[(Q_1(m_\times) + Q_2(m_\times))(1 - X)]]}{R_1(m_\times) + R_2(m_\times)}, \quad \frac{Q_1(m_\times)}{Q_1(m_\times) + Q_2(m_\times)} < X \leq 1. \quad (\text{A4})$$

Therefore, Eq. (A3) and Eq. (A4) are the general form of the Lorenz curve Eq. (13).

The Gini index is calculated by

$$\begin{aligned} G &= 2 \int_{m_0}^{\infty} (X(r) - Y(r)) \frac{dX(r)}{dr} dr \\ &= \frac{2 \int_{m_0}^{m_\times} Q_1(r) \frac{dQ_1(r)}{dr} dr}{\{Q_1(m_\times) + Q_2(m_\times)\}^2} - \frac{2 \int_{m_0}^{m_\times} R_1(r) \frac{dQ_1(r)}{dr} dr}{\{Q_1(m_\times) + Q_2(m_\times)\} \{R_1(m_\times) + R_2(m_\times)\}} \\ &\quad + \frac{2 \int_{m_\times}^{\infty} Q_2(r) \frac{dQ_2(r)}{dr} dr}{\{Q_1(m_\times) + Q_2(m_\times)\}^2} - \frac{2 \int_{m_\times}^{\infty} R_2(r) \frac{dQ_2(r)}{dr} dr}{\{Q_1(m_\times) + Q_2(m_\times)\} \{R_1(m_\times) + R_2(m_\times)\}}. \end{aligned} \quad (\text{A5})$$

When we notice

$$\int_{m_0}^{m_\times} Q_1(r) \frac{dQ_1(r)}{dr} dr = \int_{m_\times}^{\infty} \frac{1}{2} \frac{d}{dr} \{Q_1(r)^2\} dr = \frac{1}{2} \{Q_1(m_\times)^2 - Q_1(m_0)^2\} \quad (\text{A6})$$

$$\text{and } \int_{m_\times}^{\infty} Q_2(r) \frac{dQ_2(r)}{dr} dr = \int_{m_\times}^{\infty} \frac{1}{2} \frac{d}{dr} \{Q_2(r)^2\} dr = \frac{1}{2} \{Q_2(\infty)^2 - Q_2(m_\times)^2\}, \quad (\text{A7})$$

and using the definition of $S_1(m_0, m_\times)$ and $T_2(m_\times)$ (see Eq. (29)), we obtain the general form Eq. (28).

The general form of the k index Eq. (36) is simply obtain by setting $X = k$ and $Y = 1 - k$, which means $Y = 1 - X$, in Eq. (A3) and Eq. (A4).

Appendix B: Derivation of empirical Gini index

Here we show the derivation of empirical form of the Gini index Eq. (58). From the discrete expressions Eq. (57) with the relation $dX = (r + 1)/N - r/N = 1/N$, the Gini index \hat{g} is written by

$$\begin{aligned}
\hat{g} &= 2 \sum_{r=1}^N (X_r - Y_r) \frac{1}{N} \\
&= \frac{2}{N^2} \sum_{r=1}^N r - \frac{2}{\mu N^2} \sum_{r=1}^N \sum_{i=1}^r m_i \\
&= \frac{(N+1)}{N} - \frac{2}{\mu N^2} \sum_{r=1}^N (N-r+1)m_r \\
&= \frac{(N+1)}{N} - \frac{2}{\mu N^2} \left\{ \mu N(N+1) - \sum_{r=1}^N r m_r \right\} \\
&= \frac{2}{\mu N^2} \sum_{r=1}^N r m_r - \frac{(N+1)}{N}.
\end{aligned} \tag{B1}$$

This is nothing but Eq. (58).

Appendix C: Explicit forms of measures for mixture of distributions

1. (b) Lognormal with a power-law

We consider the case of $F_1(m) = e^{-\frac{(\log m - \mu)^2}{2\sigma^2}} / \sqrt{2\pi}\sigma m$ and $F_2(m) = (\alpha - 1)m^{-\alpha}$ with $m_0 = 0$. We have

$$Q_1(r) = H\left(\frac{\mu - \log r}{\sigma}\right), \quad Q_2(r) = r^{1-\alpha}; \tag{C1}$$

$$R_1(r) = e^{\mu + \frac{\sigma^2}{2}} H\left(\frac{\mu + \sigma^2 - \log r}{\sigma}\right), \quad R_2(r) = \frac{(\alpha - 1)r^{2-\alpha}}{(\alpha - 2)}. \tag{C2}$$

Then, using the following staffs $Q_1(m_\times) = H(\frac{\mu - \log m_\times}{\sigma})$, $Q_1(m_0) = Q_2(\infty) = 0$, $Q_2(m_\times) = m_\times^{1-\alpha}$, and $R_1(m_\times) = e^{\mu + \frac{\sigma^2}{2}} H(\frac{\mu + \sigma^2 - \log m_\times}{\sigma})$, $R_2(m_\times) = (\alpha - 1)m_\times / (\alpha - 2)$, and accompanying

$$S_1(m_0, m_\times) = e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\mu + \sigma^2 - \log m_\times}{\sigma}}^{\infty} Dx H(x + \sigma), \tag{C3}$$

$$T_2(m_\times) = -\frac{(\alpha - 1)^2 m_\times^{3-2\alpha}}{(\alpha - 2)(2\alpha - 3)}, \tag{C4}$$

the Lorenz curve is given as

$$Y = \begin{cases} \frac{e^{\mu + \frac{\sigma^2}{2}} H(\sigma - H^{-1}[\{H(\frac{\mu - \log m_\times}{\sigma}) + m_\times^{1-\alpha}\}X])}{e^{\mu + \frac{\sigma^2}{2}} H(\frac{\mu + \sigma^2 - \log m_\times}{\sigma}) + (\frac{\alpha-1}{\alpha-2}) m_\times^{1-\alpha}}, & 0 \leq X \leq \frac{H(\frac{\mu - \log m_\times}{\sigma})}{H(\frac{\mu - \log m_\times}{\sigma}) + m_\times^{1-\alpha}}, \\ 1 - \frac{(\frac{\alpha-1}{\alpha-2}) [H(\frac{\mu - \log m_\times}{\sigma}) + m_\times^{1-\alpha}]^{\frac{2-\alpha}{1-\alpha}}}{e^{\mu + \frac{\sigma^2}{2}} H(\frac{\mu + \sigma^2 - \log m_\times}{\sigma}) + (\frac{\alpha-1}{\alpha-2}) m_\times^{2-\alpha}} (1 - X)^{\frac{2-\alpha}{1-\alpha}}, & \frac{H(\frac{\mu - \log m_\times}{\sigma})}{H(\frac{\mu - \log m_\times}{\sigma}) + m_\times^{1-\alpha}} < X \leq 1, \end{cases} \tag{C5}$$

and the g index is

$$g = \frac{H(\frac{\mu - \log m_\times}{\sigma})^2 - m_\times^{2-2\alpha}}{(H(\frac{\mu - \log m_\times}{\sigma}) + m_\times^{1-\alpha})^2} - \frac{2(e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\mu + \sigma^2 - \log m_\times}{\sigma}}^{\infty} Dx H(x + \sigma) - \frac{(\alpha-1)^2 m_\times^{3-2\alpha}}{(\alpha-2)(2\alpha-3)})}{(H(\frac{\mu - \log m_\times}{\sigma}) + m_\times^{1-\alpha})(e^{\mu + \frac{\sigma^2}{2}} H(\frac{\mu + \sigma^2 - \log m_\times}{\sigma}) + (\frac{\alpha-1}{\alpha-2}) m_\times^{2-\alpha})} \tag{C6}$$

and k index is given as a solution of

$$H^{-1} \left[\left\{ H \left(\frac{\mu + \sigma^2 - \log m_\times}{\sigma} \right) + \left(\frac{\alpha - 1}{\alpha - 2} \right) e^{-\mu - \frac{\sigma^2}{2}} m_\times^{1-\alpha} \right\} (1 - k) \right] - H^{-1} \left[\left\{ H \left(\frac{\mu - \log m_\times}{\sigma} \right) + m_\times^{1-\alpha} \right\} k \right] = \sigma, \quad 0 \leq X \leq \frac{H \left(\frac{\mu - \log m_\times}{\sigma} \right)}{H \left(\frac{\mu - \log m_\times}{\sigma} \right) + m_\times^{1-\alpha}}, \quad (\text{C7})$$

and,

$$k = \frac{\left(\frac{\alpha-1}{\alpha-2} \right) \left(H \left(\frac{\mu - \log m_\times}{\sigma} \right) + m_\times^{1-\alpha} \right)^{\frac{2-\alpha}{1-\alpha}}}{e^{\mu + \frac{\sigma^2}{2}} H \left(\frac{\mu + \sigma^2 - \log m_\times}{\sigma} \right) + \left(\frac{\alpha-1}{\alpha-2} \right) m_\times^{2-\alpha}} (1 - k)^{\frac{2-\alpha}{1-\alpha}}, \quad \frac{H \left(\frac{\mu - \log m_\times}{\sigma} \right)}{H \left(\frac{\mu - \log m_\times}{\sigma} \right) + m_\times^{1-\alpha}} < X \leq 1. \quad (\text{C8})$$

2. (c) Uniform distribution follows a power law distribution

We next consider the case $F_1(m) = 1/a$ and $F_2(m) = (\alpha - 1)m^{-\alpha}$ with $m_0 = 0$. For this case, we have

$$Q_1(r) = \frac{r}{a}, \quad Q_2(r) = r^{1-\alpha}; \quad (\text{C9})$$

$$R_1(r) = \frac{r^2}{2a}, \quad R_2(r) = \frac{(\alpha - 2)r^{2-\alpha}}{(\alpha - 2)}, \quad (\text{C10})$$

and using the staffs $Q_1(m_\times) = m_\times/a$, $Q_1(m_0) = Q_2(\infty) = 0$, $Q_2(m_\times) = m_\times^{1-\alpha}$ and $R_1(m_\times) = m_\times^2/2a$, $R_2(m_\times) = (\alpha - 1)m_\times/(\alpha - 2)$, and accompanying

$$S_1(m_0, m_\times) = \frac{m_\times^3}{6a^2}, \quad (\text{C11})$$

$$T_2(m_\times) = -\frac{(\alpha - 1)^2 m_\times^{3-2\alpha}}{(\alpha - 2)(\alpha - 3)}, \quad (\text{C12})$$

the Lorenz curve is explicitly given as

$$Y = \begin{cases} \frac{(m_\times + am_\times^{1-\alpha})^2}{m_\times^2 + \frac{2a(\alpha-1)}{(\alpha-2)} m_\times^{2-\alpha}} X^2, & 0 \leq X \leq \frac{m_\times}{m_\times + am_\times^{1-\alpha}}, \\ 1 - \frac{\frac{1}{(\alpha-2)} \frac{2a(\alpha-1)}{(\alpha-2)} (m_\times + am_\times^{1-\alpha})^{\frac{2-\alpha}{1-\alpha}}}{m_\times^2 + \frac{2a(\alpha-1)}{(\alpha-2)} m_\times^{2-\alpha}} (1 - X)^{\frac{2-\alpha}{1-\alpha}}, & \frac{m_\times}{m_\times + am_\times^{1-\alpha}} < X \leq 1. \end{cases} \quad (\text{C13})$$

The g index is calculated as

$$g = \frac{m_\times^2 - am_\times^{2-2\alpha}}{(m_\times + am_\times^{1-\alpha})^2} - \frac{2(m_\times^3 - \frac{6a^2(\alpha-1)^2 m_\times^{3-2\alpha}}{(\alpha-2)(2\alpha-3)})}{3(m_\times + am_\times^{1-\alpha})(m_\times^2 + \frac{2a(\alpha-1)}{(\alpha-2)} m_\times^{2-\alpha})}, \quad (\text{C14})$$

and k index is determined by the solution of

$$k = \begin{cases} \frac{(m_\times^2 + \frac{2a(\alpha-1)}{(\alpha-2)} m_\times^{2-\alpha}) \left(\sqrt{1 + \frac{4(m_\times + am_\times^{1-\alpha})^2}{m_\times^2 + \frac{2a(\alpha-1)}{(\alpha-2)} m_\times^{2-\alpha}} - 1} \right)}{2(m_\times + am_\times^{1-\alpha})^2}, & 0 \leq k \leq \frac{m_\times}{m_\times + am_\times^{1-\alpha}}, \\ \frac{\frac{1}{(\alpha-2)} \frac{2a(\alpha-1)}{(\alpha-2)} (m_\times + am_\times^{1-\alpha})^{\frac{2-\alpha}{1-\alpha}}}{m_\times^2 + \frac{2a(\alpha-1)}{(\alpha-2)} m_\times^{2-\alpha}} (1 - k)^{\frac{2-\alpha}{1-\alpha}}, & \frac{m_\times}{m_\times + am_\times^{1-\alpha}} < k \leq 1. \end{cases} \quad (\text{C15})$$

3. (d) Uniform distribution with a lognormal tail

Here we choose $F_1(m) = 1/a$, $F_2(m) = e^{-\frac{(\log m - \mu)^2}{2\sigma^2}} / \sqrt{2\pi\sigma} m$ with $m_0 = 0$. We have

$$Q_1(r) = \frac{r}{a}, \quad Q_2(r) = H \left(\frac{\log r - \mu}{\sigma} \right); \quad (\text{C16})$$

$$R_1(r) = \frac{r^2}{2a}, \quad R_2(r) = e^{\mu + \frac{\sigma^2}{2}} H \left(\frac{\log r - \mu - \sigma^2}{\sigma} \right), \quad (\text{C17})$$

and by making use of the staffs $Q_1(m_\times) = m_\times/a$, $Q_1(m_0) = Q_2(\infty) = 0$, $Q_2(m_\times) = H(\frac{\log m_\times - \mu}{\sigma})$ and $R_1(m_\times) = m_\times^2/2a$, $R_2(m_\times) = e^{\mu + \frac{\sigma^2}{2}} H(\frac{\log m_\times - \mu - \sigma^2}{\sigma})$, and accompanying

$$S_1(m_0, m_\times) = \frac{m_\times^3}{6a^2}, \quad (C18)$$

$$T_2(m_\times) = -e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\log m_\times - \mu}{\sigma}}^{\infty} DxH(x - \sigma), \quad (C19)$$

the Lorenz curve is given by

$$Y = \begin{cases} \frac{(m_\times + aH(\frac{\log m_\times - \mu}{\sigma}))^2}{m_\times^2 + 2ae^{\mu + \frac{\sigma^2}{2}} H(\frac{\log m_\times - \mu - \sigma^2}{\sigma})} X^2, & 0 \leq X \leq \frac{m_\times}{m_\times + aH(\frac{\log m_\times - \mu}{\sigma})}, \\ 1 - \frac{2ae^{\mu + \frac{\sigma^2}{2}} H(H^{-1}[\frac{1}{a}\{m_\times + aH(\frac{\log m_\times - \mu}{\sigma})\}](1-X)] - \sigma)}{m_\times^2 + 2ae^{\mu + \frac{\sigma^2}{2}} H(\frac{\log m_\times - \mu - \sigma^2}{\sigma})}, & \frac{m_\times}{m_\times + aH(\frac{\log m_\times - \mu}{\sigma})} < X \leq 1. \end{cases} \quad (C20)$$

The g index is obtained as

$$g = \frac{m_\times^2 - \{2aH(\frac{\log m_\times - \mu}{\sigma})\}^2}{(m_\times + 2aH(\frac{\log m_\times - \mu}{\sigma}))^2} - \frac{2(m_\times^3 - 6a^2e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\log m_\times - \mu}{\sigma}}^{\infty} DxH(x - \sigma))}{3(m_\times + 2aH(\frac{\log m_\times - \mu}{\sigma}))(m_\times^2 + 2ae^{\mu + \frac{\sigma^2}{2}} H(\frac{\log m_\times - \mu - \sigma^2}{\sigma}))}, \quad (C21)$$

and k index becomes

$$k = \frac{(m_\times^2 + 2ae^{\mu + \frac{\sigma^2}{2}} H(\frac{\log m_\times - \mu - \sigma^2}{\sigma})) \sqrt{1 + \frac{4(m_\times + aH(\frac{\log m_\times - \mu}{\sigma}))^2}{(m_\times^2 + 2ae^{\mu + \frac{\sigma^2}{2}} H(\frac{\log m_\times - \mu - \sigma^2}{\sigma}))} - 1}}{2(m_\times + aH(\frac{\log m_\times - \mu}{\sigma}))^2}, \quad 0 \leq k \leq \frac{m_\times}{m_\times + aH(\frac{\log m_\times - \mu}{\sigma})}. \quad (C22)$$

and

$$k = H^{-1} \left[\frac{1}{a} \left\{ m_\times + aH \left(\frac{\log m_\times - \mu}{\sigma} \right) \right\} (1 - k) \right] - H^{-1} \left[\frac{1}{2ae^{\mu + \frac{\sigma^2}{2}}} \left\{ m_\times^2 + 2ae^{\mu + \frac{\sigma^2}{2}} H \left(\frac{\log m_\times - \mu - \sigma^2}{\sigma} \right) \right\} k \right] \\ = \sigma, \quad \frac{m_\times}{m_\times + aH(\frac{\log m_\times - \mu}{\sigma})} < k \leq 1. \quad (C23)$$

4. (e) Double power laws

We consider the case $F_1(m) = (\alpha - 1)m^{-\alpha}$ and $F_2(m) = (\beta - 1)m^{-\beta}$ with $m_0 = 1$. Then, we have

$$Q_1(r) = 1 - r^{1-\alpha}, \quad Q_2(r) = r^{1-\beta}; \quad (C24)$$

$$R_1(r) = \left(\frac{\alpha - 1}{\alpha - 2} \right) (1 - r^{2-\alpha}), \quad R_2(r) = \left(\frac{\beta - 1}{\beta - 2} \right) r^{2-\beta}. \quad (C25)$$

Using the staffs $Q_1(m_\times) = 1 - m_\times^{1-\alpha}$, $Q_1(m_0) = Q_2(\infty) = 0$, $Q_2(m_\times) = m_\times^{1-\beta}$ and $R_1(m_\times) = (\alpha - 1)(1 - m_\times^{2-\alpha})/(\alpha - 2)$, $R_2(m_\times) = (\beta - 1)m_\times^{2-\beta}/(\beta - 2)$, accompanying

$$S_1(m_0, m_\times) = \frac{\alpha - 1}{2\alpha - 3} - \frac{(\alpha - 1)(2\alpha - 3)m_\times^{1-\alpha} - (\alpha - 1)^2 m_\times^{3-2\alpha}}{(\alpha - 2)(2\alpha - 3)}, \quad (C26)$$

$$T_2(m_\times) = -\frac{(\beta - 1)^2 m_\times^{3-2\beta}}{(\beta - 2)(2\beta - 3)}, \quad (C27)$$

the Lorenz curve is obtained as

$$Y = \begin{cases} \frac{1 - [1 - (1 - m_\times^{1-\alpha} + m_\times^{1-\beta})X]^{\frac{2-\alpha}{1-\alpha}}}{1 - m_\times^{2-\alpha} + \frac{(\alpha-2)(\beta-1)}{(\alpha-1)(\beta-2)} m_\times^{2-\beta}}, & 0 \leq X \leq \frac{1 - m_\times^{1-\alpha}}{1 - m_\times^{1-\alpha} + m_\times^{1-\beta}}, \\ 1 - \frac{\frac{(\alpha-2)(\beta-1)}{(\alpha-1)(\beta-2)} [(1 - m_\times^{1-\alpha} + m_\times^{1-\beta})(1-X)]^{\frac{2-\beta}{1-\beta}}}{1 - m_\times^{2-\alpha} + \frac{(\alpha-2)(\beta-1)}{(\alpha-1)(\beta-2)} m_\times^{2-\beta}}, & \frac{1 - m_\times^{1-\alpha}}{1 - m_\times^{1-\alpha} + m_\times^{1-\beta}} < X \leq 1. \end{cases} \quad (C28)$$

The g index is given by

$$g = \frac{(1 - m_x^{1-\alpha})^2 - m_x^{2-2\beta}}{(1 - m_x^{1-\alpha} + m_x^{1-\beta})^2} - \frac{2\left(\frac{\alpha-2}{\alpha-1}\right)\left\{\frac{\alpha-1}{2\alpha-3} - \frac{(\alpha-1)(2\alpha-3)m_x^{1-\alpha} - (\alpha-1)^2m_x^{3-2\alpha}}{(\alpha-2)(2\alpha-3)} - \frac{(\beta-1)^2m_x^{3-2\beta}}{(\beta-2)(2\beta-3)}\right\}}{(1 - m_x^{1-\alpha} + m_x^{1-\beta})(1 - m_x^{2-\alpha} + \frac{(\alpha-2)(\beta-1)}{(\alpha-1)(\beta-2)}m_x^{2-\beta})}, \quad (C29)$$

and the k index as

$$k = \begin{cases} \frac{1 - [1 - \{1 - m_x^{2-\alpha} + \frac{(\alpha-2)(\beta-1)}{(\alpha-1)(\beta-2)}m_x^{2-\beta}\}(1-k)]^{\frac{1-\alpha}{2-\alpha}}}{1 - m_x^{1-\alpha} + m_x^{1-\beta}}, & 0 \leq X \leq \frac{1 - m_x^{1-\alpha}}{1 - m_x^{1-\alpha} + m_x^{1-\beta}}, \\ \frac{\left(\frac{\alpha-2}{\alpha-1}\right)\left(\frac{\beta-1}{\beta-2}\right)\left[\{1 - m_x^{1-\alpha} + m_x^{1-\beta}\}(1-k)\right]^{\frac{2-\beta}{1-\beta}}}{1 - m_x^{2-\alpha} + m_x^{2-\beta}}, & \frac{1 - m_x^{1-\alpha}}{1 - m_x^{1-\alpha} + m_x^{1-\beta}} < X \leq 1. \end{cases} \quad (C30)$$

5. (f) Power law distribution with a lognormal tail

Finally we consider the case $F_1(m) = (\alpha - 1)m^{-\alpha}$ and $F_2(m) = e^{-\frac{(\log m - \mu)^2}{2\sigma^2}}/\sqrt{2\pi}\sigma m$ with $m_0 = 1$. Then, we have

$$Q_1(r) = 1 - r^{1-\alpha}, \quad Q_2(r) = H\left(\frac{\log r - \mu}{\sigma}\right); \quad (C31)$$

$$R_1(r) = \left(\frac{\alpha - 1}{\alpha - 2}\right)(1 - r^{2-\alpha}), \quad R_2(r) = e^{\mu + \frac{\sigma^2}{2}} H\left(\frac{\log r - \mu - \sigma^2}{\sigma}\right). \quad (C32)$$

Using the staffs $Q_1(m_x) = 1 - m_x^{1-\alpha}$, $Q_1(m_0) = Q_2(\infty) = 0$, $Q_2(m_x) = H\left(\frac{\log m_x - \mu}{\sigma}\right)$ and $R_1(m_x) = (\alpha - 1)(1 - m_x^{2-\alpha})/(\alpha - 2)$, $R_2(m_x) = e^{\mu + \frac{\sigma^2}{2}} H\left(\frac{\log m_x - \mu - \sigma^2}{\sigma}\right)$, accompanying

$$S_1(m_0, m_x) = \frac{\alpha - 1}{2\alpha - 3} - \frac{(\alpha - 1)(2\alpha - 3)m_x^{1-\alpha} - (\alpha - 1)^2m_x^{3-2\alpha}}{(\alpha - 2)(2\alpha - 3)}, \quad (C33)$$

$$T_2(m_x) = -e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\log m_x - \mu}{\sigma}}^{\infty} Dx H(x - \sigma), \quad (C34)$$

the Lorenz curve is given by

$$Y = \begin{cases} \frac{1 - [1 - \{1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu}{\sigma}\right)\}X]^{\frac{2-\alpha}{1-\alpha}}}{1 - m_x^{2-\alpha} + e^{\mu + \frac{\sigma^2}{2}} \left(\frac{\alpha-2}{\alpha-1}\right) H\left(\frac{\log m_x - \mu - \sigma^2}{\sigma}\right)}, & 0 \leq X \leq \frac{1 - m_x^{1-\alpha}}{1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu - \sigma^2}{\sigma}\right)}, \\ 1 - \frac{\left(\frac{\alpha-2}{\alpha-1}\right) e^{\mu + \frac{\sigma^2}{2}} H\left(H^{-1}\left[\{1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu}{\sigma}\right)\}(1-X)\right] - \sigma\right)}{1 - m_x^{2-\alpha} + e^{\mu + \frac{\sigma^2}{2}} \left(\frac{\alpha-2}{\alpha-1}\right) H\left(\frac{\log m_x - \mu - \sigma^2}{\sigma}\right)}, & \frac{1 - m_x^{1-\alpha}}{1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu - \sigma^2}{\sigma}\right)} < X \leq 1. \end{cases} \quad (C35)$$

and the g index is given by

$$g = \frac{(1 - m_x^{1-\alpha})^2 - H\left(\frac{\log m_x - \mu}{\sigma}\right)^2}{(1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu}{\sigma}\right))^2} - \frac{2\left(\frac{\alpha-2}{\alpha-1}\right)\left\{\frac{\alpha-1}{2\alpha-3} - \frac{(\alpha-1)(2\alpha-3)m_x^{1-\alpha} - (\alpha-1)^2m_x^{3-2\alpha}}{(\alpha-2)(2\alpha-3)} - e^{\mu + \frac{\sigma^2}{2}} \int_{\frac{\log m_x - \mu}{\sigma}}^{\infty} Dx H(x - \sigma)\right\}}{(1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu}{\sigma}\right))(1 - m_x^{2-\alpha} + e^{\mu + \frac{\sigma^2}{2}} \left(\frac{\alpha-2}{\alpha-1}\right) H\left(\frac{\log m_x - \mu - \sigma^2}{\sigma}\right))}, \quad (C36)$$

and k index is

$$k = \frac{1 - [1 - \{1 - m_x^{2-\alpha} + e^{\mu + \frac{\sigma^2}{2}} \left(\frac{\alpha-2}{\alpha-1}\right) H\left(\frac{\log m_x - \mu - \sigma^2}{\sigma}\right)\}(1-k)]^{\frac{1-\alpha}{2-\alpha}}}{1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu}{\sigma}\right)}, \quad 0 \leq k \leq \frac{1 - m_x^{1-\alpha}}{1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu}{\sigma}\right)}. \quad (C37)$$

and

$$k = H^{-1} \left[\left\{ 1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu}{\sigma}\right) \right\} (1-k) \right] - H^{-1} \left[\frac{1}{e^{\mu + \frac{\sigma^2}{2}} \left(\frac{\alpha-2}{\alpha-1}\right)} \left\{ 1 - m_x^{2-\alpha} + e^{\mu + \frac{\sigma^2}{2}} \left(\frac{\alpha-2}{\alpha-1}\right) H\left(\frac{\log m_x - \mu - \sigma^2}{\sigma}\right) \right\} k \right], \quad \frac{1 - m_x^{1-\alpha}}{1 - m_x^{1-\alpha} + H\left(\frac{\log m_x - \mu}{\sigma}\right)} < k \leq 1. \quad (C38)$$

-
- [1] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2009.
- [2] K. J. Arrow, S. Bowles, and S. N. Durlauf. *Meritocracy and economic inequality*. Princeton Univ. Press, 2000.
- [3] J. E. Stiglitz. *The price of inequality: How today's divided society endangers our future*. WW Norton & Company, 2012.
- [4] K. Neckerman. *Social Inequality*. Russell Sage Foundation, 2004.
- [5] J. H. Goldthorpe. Analysing social inequality: a critique of two recent contributions from economics and epidemiology. *Eur. Sociological Rev.*, 26(6):731–744, 2010.
- [6] C. E. Hurst. *Social Inequality: Forms, Causes, and Consequences*. Allyn and Bacon, Boston, 1995.
- [7] V. M. Yakovenko and J. Barkley Rosser Jr. Statistical mechanics of money, wealth, and income. *Rev. Mod. Phys.*, 81(4):1703, 2009.
- [8] B. K. Chakrabarti, A. Chakraborti, S. R. Chakravarty, and A. Chatterjee. *Econophysics of income and wealth distributions*. Cambridge Univ. Press, Cambridge, 2013.
- [9] H. Aoyama, Y. Fujiwara, and Y. Ikeda. *Econophysics and companies: statistical life and death in complex business networks*. Cambridge Univ. Press, Cambridge, 2010.
- [10] S. Lawrence, Q. Liu, and V. M. Yakovenko. Global inequality in energy consumption from 1980 to 2010. *Entropy*, 15(12):5565–5579, 2013.
- [11] T. Piketty and E. Saez. Inequality in the long run. *Science*, 344(6186):838–843, 2014.
- [12] A. Cho. Physicists say it's simple. *Science*, 344(6186):828, 2014.
- [13] G. Chin and E. Culotta. What the numbers tell us. *Science*, 344(6186):818–821, 2014.
- [14] Y. Xie. Udemocracy: Inequalities in science. *Science*, 344(6186):809–810, 2014.
- [15] C. Gini. Measurement of inequality of incomes. *Econ. J.*, 31(121):124–126, 1921.
- [16] H. Theil. *Economics and information theory*, volume 7. North-Holland Amsterdam, 1967.
- [17] I. I. Eliazar and I. M. Sokolov. Measuring statistical heterogeneity: The pietra index. *Physica A*, 389(1):117–125, 2010.
- [18] A. A. Drăgulescu and V. M. Yakovenko. Exponential and power-law probability distributions of wealth and income in the united kingdom and the united states. *Physica A*, 299(1):213–221, 2001.
- [19] M. O. Lorenz. Methods for measuring the concentration of wealth. *Am. Stat. Assoc.*, 9:209–219, 1905.
- [20] A. Ghosh, N. Chattopadhyay, and B. K. Chakrabarti. Inequality in societies, academic institutions and science journals: Gini and k-indices. *Physica A*, 410(14):30–34, 2014.
- [21] Internal Revenue Service (IRS) – Individual Tax Tables, retrieved June, 2014. <http://www.irs.gov/uac/SOI-Tax-Stats—Individual-Statistical-Tables-by-Size-of-Adjusted-Gross-Income>.
- [22] Statistics Denmark, retrieved June, 2014. <http://www.dst.dk>.
- [23] N. Sazuka and J.-I. Inoue. Fluctuations in time intervals of financial data from the view point of the gini index. *Physica A*, 383(1):49–53, 2007.
- [24] N. Sazuka, J.-I. Inoue, and E. Scalas. The distribution of first-passage times and durations in forex and future markets. *Physica A*, 388(14):2839–2853, 2009.
- [25] A. A. Drăgulescu and V. M. Yakovenko. Evidence for the exponential distribution of income in the usa. *Eur. Phys. J. B*, 20(4):585–589, 2001.
- [26] I. I. Eliazar and M. H. Cohen. On social inequality: analyzing the rich-poor disparity. *Physica A*, 401(1):148 – 158, 2014.
- [27] A. Chatterjee, M. Mitrović, and S. Fortunato. Universality in voting behavior: an empirical analysis. *Sci. Reports*, 3:1049, 2013.
- [28] ISI Web of Science, retrieved April, 2014. <http://portal.isiknowledge.com>.
- [29] DATASUS, retrieved August, 2013. <http://www2.datasus.gov.br/DATASUS/>.
- [30] Población de España - datos y mapas, retrieved August, 2013. <http://alarcos.inf-cr.uclm.es/per/fruiz/pobesp/>.
- [31] e-Stat: Official Statistics of Japan, retrieved March, 2014. <https://www.e-stat.go.jp/SG1/chiiki/Welcom.do?lang=02>.
- [32] Science, Special Issue: The Science of Inequality, 23 May 2014. <http://www.sciencemag.org/site/special/inequality/>.
- [33] V. Pareto. *Manuale di economia politica*, volume 13. Societa Editrice, 1906.