Inequality as difference: A teaching note on the Gini coefficient

Samuel Bowles
Wendy Carlin

SFI WORKING PAPER: 2017-02-003

INEQUALITY AS DIFFERENCE: A TEACHING NOTE ON THE GINI COEFFICIENT

Samuel Bowles and Wendy Carlin[1]
18 February 2017

Abstract. Some of the dimensions along which we measure inequality are best conceived of as individual attributes that is, something that people simply have more or less of, like height. But on both descriptive and normative grounds other dimensions are best conceived of as differences between people. Economic inequalities – in wealth or income, for example – are in this latter class. Treating an economy as a complete undirected network the edges of which (not the nodes) are the fundamental data of the measurement allows a simple derivation of a standard inequality measure – the Gini coefficient – that is both intuitive and ranges over the unit interval irrespective of population size, thus avoiding two shortcomings of the conventional difference-based algorithm for calculating the Gini. We also illustrate the application of the Gini coefficient to settings in which individuals are members of homogeneous classes. Intuitions readily understood from these applications are given.

Keywords: inequality, Gini coefficient, relative mean difference, class, Lorenz curve

JEL codes: A2 (economic education and teaching economics); D31 (income, wealth, and their distributions)

Key points

- An intuitive explanation of an inequality measure based on differences between people
- Gini coefficient algorithm that returns values from 0 to 1 without requiring population size adjustment
- Applications to a class-divided economy

---

[1] Affiliations: Bowles: Santa Fe Institute; Carlin: University College London. We are grateful to the Behavioral Sciences Program at SFI for their support of this work and to the Curriculum Open Access Resources for Economics project of UCL (CORE) and especially Antonio Cabrales and Arjun Jayadev for stimulating us to find a simpler way to explain the Gini coefficient.

To explain the Gini coefficient as a measure of economic inequality to novices it is often effective to use small numbers examples such as cake cutting and or simplified applications to a population composed of homogeneous income classes. Here we point out and resolve a technical difficulty with the small numbers illustration of the Gini coefficient and provide examples of applications to class-divided populations.

1. An unbiased relative mean difference expression for the Gini coefficient

The common definition of the Gini coefficient is the area between the Lorenz curve and the line of perfect equality divided by the total area under the perfect equality line. For finite populations this formulation is known to be inconsistent with the equally common explanation that the coefficient ranges from 0 (no differences in the quantity in question) to 1 (a single recipient accounts for all of the quantity.) As a result, teachers and others attempting to explain the Gini coefficient to novices often find that intuitive explanations illustrated by small populations do not work.

The source of the problem is the population size bias in the usual formulation of the Gini coefficient. For example in a small population the Gini will fall far short of 1 even if a single individual has all of the wealth.

The standard expression for this Lorenz curve based interpretation (which we will call G^L) is due to Kendall and Stuart 1969, Dasgupta, Sen and Starrett 1973 and others and when applied to some quantity, y, is

$$1) \quad G^L = \frac{\sum_{k=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|}{n^2}\frac{1}{\underline{y}}\frac{1}{2}$$

which can be seen to be the mean difference among all pairs in the population relative to the average value of the quantity in question ($\underline{y}$) times one half. A population size adjustment is required so that this value will vary over the unit interval and take a value of 1 when a single person receives all of the quantity, namely, using G as the unbiased Gini coefficient:

$$2) \quad G = G^L \frac{n}{n-1}$$

The discrepancy between the two explanations of the Gini coefficient is substantial when the population size under consideration is small, as is often convenient in explaining the Gini coefficient to students or others unfamiliar with the measures. For example for a two person economy in which one person has all of the income, G^L is equal to one half, not 1. Asking students to remember that they must multiply G^L by $n/(n\text{-}1)$ to get the correct measure seems an unnecessary step, if it can be avoided.

The population size bias arises because of a non-intuitive way that equation (1) counts differences among pairs in the population. The relative mean difference among members of a population can be computed in two equivalent ways depending on whether we count an individual paired with itself (and hence a difference of zero) among the possible pairs and whether a pair is counted twice (k paired with j and j paired with k) or just once.

The idea of counting a person "paired" with itself as a pair is mathematically convenient, but not very intuitive if we are studying differences among people. Similarly counting a pair twice makes sense if we are interested in something transmitted one to the other on a directed graph, but not when studying a difference between the two nodes of which there can be only one. So we model the population as an undirected network, which means we count each pair (k and j) only once as a pair, and only once as a difference.

The difference between the two approaches is clarified if we consider the population as a complete undirected network the edges of which are the object of study, as in the left panel of Figure 1. The right panel represents the conventional measure (equation 1) as a directed network. If there are n members of the population then the total number of unique non identical pairs is $(n^2 - n)/2$, shown as the three edges in the left panel instead of n^2 pairs (the nine edges in the right panel).

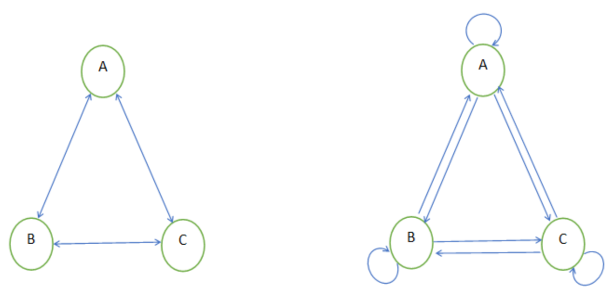


Figure 1. Unique non identical differences (left panel) and the conventional directed graph measure.

Using the unique non identical pairing setup we now derive an unbiased measure of G. Let Δ be the sum of the absolute differences among the (unique non identical) pairs, we have the following measure

$$3)\quad G = \frac{\Delta}{(n^2 - n)/2}\frac{1}{\underline{y}}\frac{1}{2} = \frac{\Delta}{n(n\ -1)}\frac{1}{\underline{y}}$$

Equation 3 gives the Gini as the mean difference among all pairs (the first term) relative to (divided by) the mean value of y (the "relative mean difference") times one half.

Like equation (1) it is half the relative mean difference, but we will see that it is unbiased in that when a single individual receives all of the measure $G = 1$ independently of population size. Suppose that a single individual in a group of finite size n owns the entire quantity which is y. Then there is a total of $n(n-1)/2$ undirected edges among all pairs in the population, of which all are zeros except the $(n-1)$ edges connecting the "have nots" with the single "have". The sum of the differences is thus $y(n-1) = \Delta$. Inserting this value in 3) confirms that G = 1 in this case for any $n > 1$.

Equation (3) is related to the standard expression as follows. First in our proposed reformulation, there are as we have seen $(n^2 - n)/2$ instead of n^2 pairs. Second, in the conventional approach, the sum of the differences between the pairs taking account of the double counting of identical pairs is 2Δ (the addition of the self-on-self pairs does not add to Δ because these "differences" are all zero). So we have

$$4)\quad G^L = \frac{2\Delta}{n^2}\frac{1}{\underline{y}}\frac{1}{2} = \frac{\Delta}{n^2}\frac{1}{\underline{y}} = \frac{A}{A+B}$$

Now correcting this approximation for the small numbers bias we have:

$$5)\quad G^L\left(\frac{n}{n-1}\right) = \frac{\Delta}{n^2}\frac{1}{\underline{y}}\left(\frac{n}{n-1}\right).$$

This expression based on double counting pairs and including "self on self" pairs when corrected for small numbers bias is equivalent to the more intuitive expression above as can be seen in the following

$$6)\quad G^L\left(\frac{n}{n-1}\right) = \frac{\Delta}{n^2}\frac{1}{\underline{y}}\left(\frac{n}{n-1}\right) = \frac{\Delta}{n}\frac{1}{\underline{y}}\left(\frac{1}{n-1}\right) = \frac{\Delta}{n(n-1)}\frac{1}{\underline{y}} = G$$

The unique non identical pairs approach here also provides an intuitive explanation of why the relative mean difference is multiplied by half to get the Gini coefficient. We find the maximum relative mean difference (one recipient receives all of the measure as above) as

7) $$\frac{y(n-1)}{n(n-1)/2}=\frac{2y}{n}$$

which means that:

8) $$\frac{\text{Mean difference}}{\text{Mean income}}=\frac{2y/n}{y/n}=2$$

To bound the measure of inequality so that its maximum value is one, we therefore multiply the relative mean difference by half to get the Gini coefficient for a population in which a single individual owns all of the wealth, which note is equal to 1 independent of population size.

9) $$G=\frac{1}{2}\frac{y(n-1)\Big/ n(n-1)\frac{1}{2}}{y/n}=1$$

2. ILLUSTRATING CLASS DIFFERENCES WITH THE GINI COEFFICIENT.

The Lorenz curve is typically presented as a continuously differentiable convex function representing a large population with differing amounts of human or material assets. But the same apparatus can be applied to income differences between classes in an economy of "producers" "owners" and the "inactive." Members of these three classes are identical in the sense that the income of each is the same; so we abstract from within class differences.

Output is average product of the producers (q) times the number of producers (*n*) normalized by setting total population equal to 1. Output is composed of the following fractions *w/q* received by the workers (producers) and 1-*w/q* received by their employers (or other non-producers) Total output (nq) is normalized in the figure to be unity. The share of output received by the producers is $s = w/q$. We assume the population is sufficiently large to use the Lorenz curve based measure without the population bias adjustment.

We know from the derivation in the appendix that the conventional Gini coefficient is

10) $$G=u+n-(1-u)w/q$$

This does not seem very transparent. But we can make it more so by considering the following.

First notice that as the non-producing income claiming class gets smaller (relatively), that is, as u or n increase (holding constant the share of income or the non-producing claimants) inequality goes up, as one would expect. This could depict the evolution of capitalism from an economy of smallish family owned firms and manufacturers employing a few workers to a modern economy with concentrated wealth and hence fewer "owners." Then notice that an increase in the wage share (or share of total output retained by the producers if we have a sharecropping economy) will reduce the Gini. And observe that if the identical producers keep all that they produce ($w/q = 1$) then there are no non-producing income claimants, and if there is no unemployment, then $G = 0$.

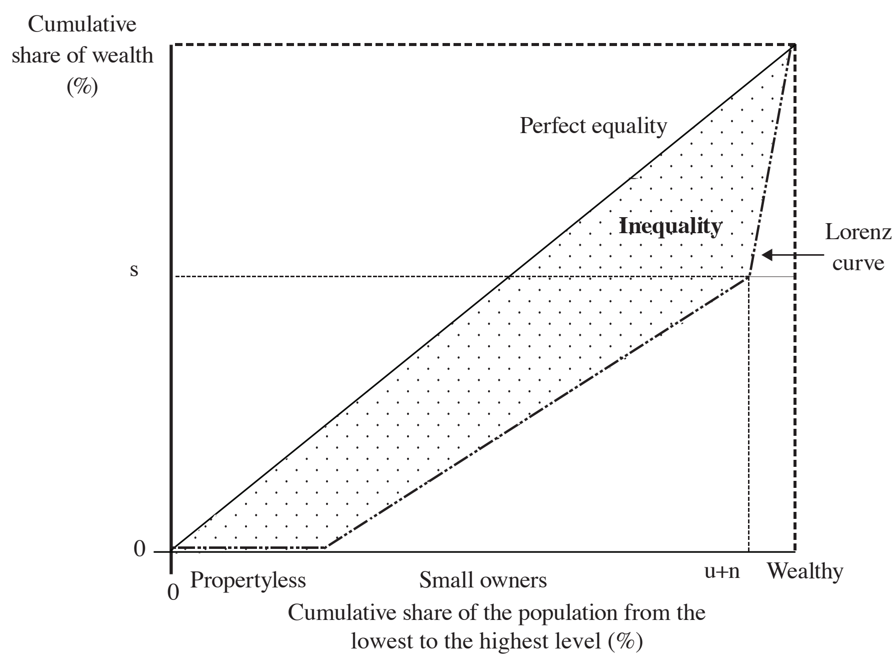


Figure 2. The Lorenz curve for a class-divided ecoomy.

But more interesting are the following.

An extreme class society: Now suppose that in a large population there is just a single employer so $u + n \approx 1$ and using this approximation, we have

$$G = 1 - (1 - u)(w / q)$$

This allows us to see that if there are no unemployed, then $G = 1 - (w / q)$ = the profit share (or the single landlord's crop share, etc). Here Gini-inequality is entirely determined by the class or producer to non-producer relationship.

The same analysis could be used for a large population of producers with a single state elite ("king") that taxes the producers. Then the Gini is the tax rate, that is, the share of output that the producers hand over to the king.

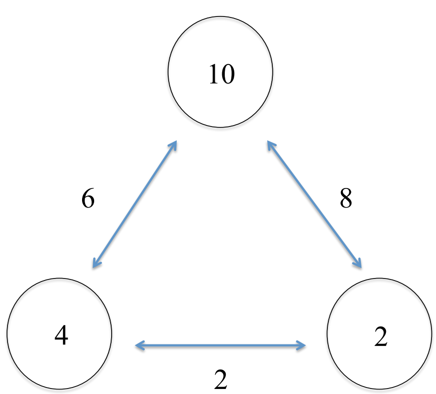
Demographic segmentation: if producers receive their entire output ($w/q = 1$) then there are no non producer income claimants, and $u + n = 1$ so $G = u$: the Gini is the rate of unemployment (fraction of population not engaged as producers). Here the Gini is determined entirely by "demographic structure" i.e. the fraction of the population who are producers.

3. CONCLUSION

Both the undirected network algorithm for calculating the Gini coefficient, and the applications to class divided economies suggested have been effectively used in teaching first year university students about the measurement and determinants of economic inequality. Both approaches convey intuitions about inequality as the result of relationships between people rather than being entirely a matter of individuals' diverse endowments. Extensive applications (conceptual and empirical) of these methods can be found in *The Economy* by The CORE Team at www.core-econ.org.

Teaching Exercises

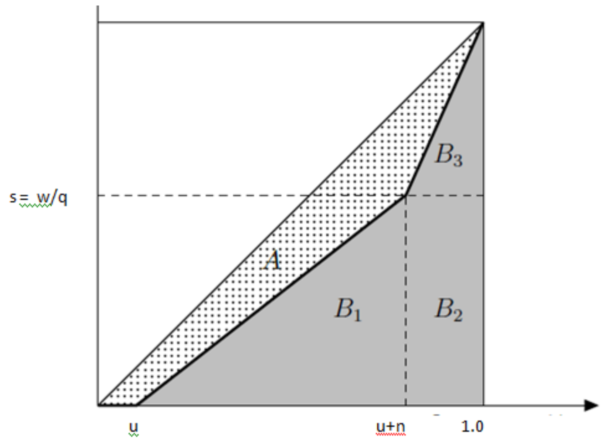
1. Calculate the Gini coefficient for the 3-person economy below (numbers in the circles are the wealth of person represented by the node, numbers on the arrows are the differences between the pairs)



Using equation 3, the Gini coefficient for the three-person economy shown is 16/3 x 1/(16/3) x ½ = 0.5

2. If 80 employees of 10 employers receive in wages three quarters of the economy's output, their employers receiving the rest, and 10 workers are unemployed (receiving nothing), use equation 10 to calculate the Gini coefficient.

Appendix: Derivation of the Gini for the class-divided economy



$$B_1 = \frac{1}{2} nw/q$$

$$B_2 = (1-u-n)w/q$$

$$B_3 = \frac{1}{2}(1-u-n)(1-w/q)$$

$$B = B_1 + \frac{1}{2}B_2 + \frac{1}{2}B_2 + B_3$$

$$B = \frac{1}{2}nw/q + \frac{1}{2}(1-u-n)w/q + \frac{1}{2}(1-u-n)w/q + \frac{1}{2}(1-u-n)(1-w/q)$$

$$B = \frac{1}{2}(1-u)w/q + \frac{1}{2}(1-u-n) = \frac{1}{2}\{(1-u-n)+(1-u)w/q\}$$

$$g = \frac{A}{A+B} = \frac{1/2 - B}{1/2} = 1 - \frac{B}{1/2} = 1 - 2B = u + n - (1-u)w/q$$

Works cited.

Dasgupta, Partha, Amartya Sen, and David Starrett. "Notes on the Measurement of Inequality." *Journal of Economic Theory* 6 (1973): 180-87.

Kendall, Maurice G., and Alan Stuart. *The Advanced Theory of Statistics* London: Charles Griffin, 1969.