Gini Coefficient Calculations

March 30, 2020

1 Gini Coefficient Definitions

The Gini coefficient is a measure of inequality popular in biology and economics, among other disciplines. It has two definitions. The first, as shown in Figure 1, is the area between the Lorenz Curve and the Line of Equality divided by the total area under the line of equality. We represent that by the following equation:

$$G = \frac{A}{A+B} \tag{1}$$

The second definition of the Gini coefficient is half the relative mean difference across an income distribution. In other words, it is the expected value of the operation of randomly selecting two people from a distribution and taking the ratio of the absolute value of the difference between their incomes to the mean income. This can be represented mathematically as follows:

$$G = \frac{E[|m_1 - m_2|]}{2E[m]} \tag{2}$$

where m_1, m_2, and m are identical independentally distributed probability density functions representing the amount of money an individual in a society holds. I don't have a graphic for this but hopefully **@Danny** makes one out of gratitude for this proof.

It should be noted that there is no immediately obvious reason why the two definitions above should give the same answer. This paper seeks to provide a somewhat unintiutive but hopefully enlightening proof of their equivalence.

2 Proof of Equivalence

First, we write the above definitions more precisely. For a given distribution m, let M represent the cumulative distribution function:

$$M(n) = \int_{-\infty}^{n} m(x)dx, 0 \leq n \leq 1 \tag{3}$$

We further note that the problem has been scaled as in Figure 1: namely, n represents where in the income distribution a person is, with 0 representing the person who makes

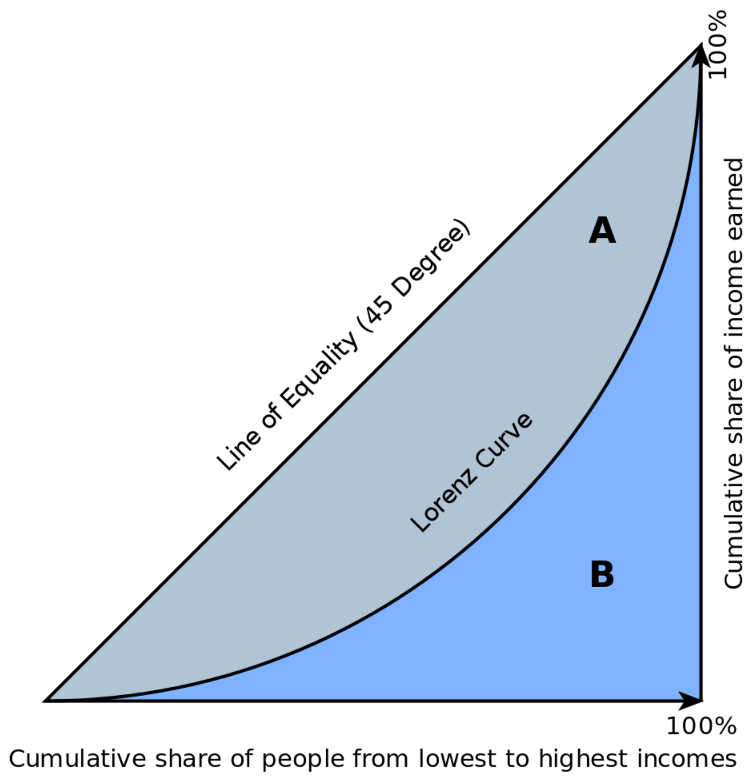

Figure 1: Graphical representation of the Gini coefficient

the least money and 1 representing the person who makes the most. Lastly, we assume the problem has been scaled such that the total amount of income in the society is equal to one. These assumptions are not necessary, but make calculation easier.

Next, we more precisely define the two definitions of the Gini coefficient. The first definition is a graphical one involving areas, which can be rewritten as an integral. Note that the Lorenz curve is simply M, the cumulative distribution function of income. From this point on, we will use G_1 and G_2 to represent the first and second definitions of the Gini coefficient, respectively.

$$G_1 = \frac{\int_0^1 (x - M(x))dx}{\int_0^1 x dx} \tag{4}$$

Because the area of the triangle is $\frac{1}{2}$, we can rewrite the above equation:

$$G_1 = 1 - 2\int_0^1 M(x)dx \tag{5}$$

The second definition is a bit trickier. However, it can be written as follows:

$$G_2 = \frac{\int_0^1 \int_0^1 |m(x_1) - m(x_2)| dx_2 dx_1}{2\int_0^1 m(x)dx} \tag{6}$$

We note that as $m(x_1)$ and $m(x_2)$ are identical independently distributed random variables, we can take advantage of symmetry to rewrite the above integral:

$$G_2 = \frac{\int_0^1 \int_0^{x_1} (m(x_1) - m(x_2)) dx_2 dx_1}{\int_0^1 m(x) dx} \tag{7}$$

Now, all that remains is to solve the integral. We take advantage of the definition of M:

$$\begin{aligned} G_2 &= \frac{\int_0^1 \big(xm(x_1) - M(x_1)\big) dx_1}{M(1) - M(0)} \\ &= \int_0^1 xm(x_1) dx_1 - \int_0^1 M(x_1) dx_1 \end{aligned} \tag{8}$$

Now, we use integration by parts to solve the first part of the above integral:

$$\begin{aligned} \int_0^1 xm(x) dx &= xM(x)|_0^1 - \int_0^1 M(x) dx \\ &= 1 - \int_0^1 M(x) dx \end{aligned} \tag{9}$$

Finally, our expression for the Gini coefficient can be written:

$$G_2 = 1 - 2 \int_0^1 M(x) dx \tag{10}$$

which is seen to equal G_1 $\square$