Tutorial Note: Curve of Concentration: Gini Coefficient

1 Definition

The Gini coefficient is usually defined mathematically based on the Lorenz curve, which plots the proportion of the total income of the population (Y axis) that is cumulatively earned by the bottom $x\%$ of the population (see diagram). The line at 45 degrees thus represents perfect equality of incomes. The Gini coefficient can then be thought of as the ratio of the area that lies between the line of equality and the Lorenz curve (marked A in the diagram) over the total area under the line of equality (marked A and B in the diagram); i.e., G = A / (A + B). If all people have non-negative income (or wealth, as the case may be), the Gini coefficient can theoretically range from 0 to 1; it is sometimes expressed as a percentage ranging between 0 and 100. In practice, both extreme values are not quite reached. If negative values are possible (such as the negative wealth of people with debts), then the Gini coefficient could theoretically be more than 1. Normally the mean (or total) is assumed positive, which rules out a Gini coefficient less than zero. A low Gini coefficient

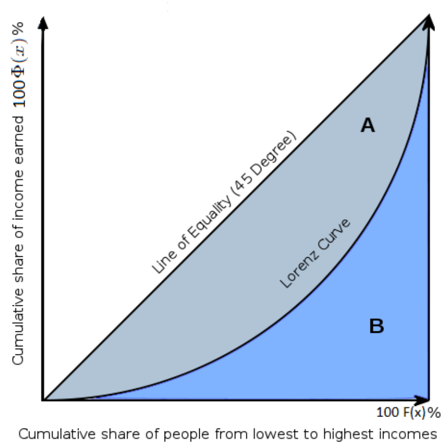

Figure 1: Graphical representation of the Lorenz Curve or the Curve of Concentration and the area of concentration

indicates a more equal distribution, with 0 corresponding to complete equality, while higher Gini coefficients indicate more unequal distribution, with 1 corresponding to complete inequality. When used as a measure of income inequality, the most unequal society will be one in which a single person receives 100% of the total income and the remaining people receive none (G = $11/N$); and the most equal society will be one in which every person receives the same income (G = 0).

An alternative approach would be to consider the Gini coefficient as half of the relative mean difference, which is a mathematical equivalence. The mean difference is the average absolute difference between two items selected randomly from a population, and the relative mean difference is the mean difference divided by the average, to normalize for scale.

The area between the line of equal distribution and the curve of concentration is called the *area of concentration*, this is an indicator of the degree of concentration. The Gini index is defined as a ratio of the areas on the Lorenz curve diagram. If the area between the line of perfect equality and the curve of concentration is A, and the area under the curve of concentration is B, then the Gini index is A / (A + B). Since A + B = 1/2, the Gini index is G = 2 $\times A$ or G = 1 - 2B.

Result: *Gini coefficient* is equal to the half of the *relative mean difference.*

Proof: Let $F(x)$ denote the cumulative frequency for the variable value upto x and let $\Phi(x)$ denote the cumulative total (i.e. value times frequency/density) upto same x. The curve obtained by plotting Φ against F for different fixed values of x is known as the Lorenz curve or the curve of concentration. Let μ be the population mean of the variable (say, income).

As we know $f(x) = \frac{d}{dx}F(x)$ is a standardize weight function in terms of frequency attached to the value x. Similarly on can think $\frac{d}{dx}\Phi(x) = \phi(x)$ is other type of weight function in terms of total income for x. Hence, $\phi(x) = \frac{xf(x)}{\int xf(x)dx} = \frac{xf(x)}{\mu}$. So, $\mu\frac{d\Phi}{dx} = xf(x)$.

$$\text{Now,}\mu\frac{d\Phi}{dF} = \frac{\mu d\Phi/dx}{dF/dx} = \frac{xf(x)}{f(x)} = x$$

and

$$\mu\frac{d^2\Phi}{dF^2} = \frac{dx}{dF} = f^{-1}(x) > 0$$

The line $\Phi = F$ is called the line of equal distribution. Then

$$\begin{aligned}
G &= \int_0^1 Fd\Phi - \int_0^1 \Phi dF \\
&= \frac{1}{\mu}\left[\int_{-\infty}^{\infty} F(x)xdF(x) - \int_{-\infty}^{\infty} \Phi(x)dF(x)\right] \\
&= \frac{1}{\mu}\left[\int_{-\infty}^{\infty} x\left\{\int_{-\infty}^{x} dF(y)\right\}dF(x) - \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{x} ydF(y)\right\}dF(x)\right] \\
&= \frac{1}{\mu}\int_{-\infty}^{\infty}\int_{-\infty}^{x}(x-y)dF(y)dF(x) \\
&= \frac{1}{2\mu}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{x}(x-y)dF(y)dF(x) + \int_{-\infty}^{\infty}\int_{x}^{\infty}(y-x)dF(y)dF(x)\right] \\
&= \frac{1}{2\mu}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mid x-y \mid dF(y)dF(x) \qquad (1) \\
&= \frac{1}{2} \times \text{relative mean difference}
\end{aligned}$$

Note:1. In case of a given set of n values of variable, say, x_1, x_2, ..., x_n,

$$G = \frac{1}{2n^2\mu}\Sigma_{i=1}^{n}\Sigma_{j=1}^{n} \mid x_i - y_j \mid$$

Note:2. For a cumulative distribution function F of a continuous non-negative variable that has a mean μ,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mid x - y \mid dF(y)dF(x) = 2\int_{-\infty}^{\infty} F(y)(1 - F(y))dy$$

. Then,

$$G = \frac{1}{\mu}\int_{-\infty}^{\infty} F(y)(1 - F(y))dy \qquad (Exercise)$$

Hint: Relative mean difference can be expressed as $2\mu - 2\int_0^{\infty}(1 - F(y))^2 dy$.

2 Features of Gini coefficient (G)

Gini coefficient has features that make it useful as a measure of dispersion in a population, and inequalities in particular. It also avoids references to a statistical average or position unrepresentative of most of the population, such as per capita income or gross domestic product. For a given time interval, Gini coefficient can therefore be used to compare diverse countries and different regions or groups within a country; for example states, counties, urban versus rural areas, gender and ethnic groups. Gini coefficients can be used to compare income distribution over time, thus it is possible to see if inequality is increasing or decreasing independent of absolute incomes. Other useful features of Gini coefficient are:

Anonymity: it does not matter who the high and low earners are.

Scale independence: the Gini coefficient does not consider the size of the economy, the way it is measured, or whether it is a rich or poor country on average.

Population independence: it does not matter how large the population of the country is.

Transfer principle: if income (less than the difference), is transferred from a rich person to a poor person the resulting distribution is more equal.

3 Other Uses

Although the Gini coefficient is most popular in economics, it can in theory be applied in any field of science that studies a distribution. For example, in ecology the Gini coefficient has been used as a measure of biodiversity, where the cumulative proportion of species is plotted against cumulative proportion of individuals. In health, it has been used as a measure of the inequality of health related quality of life in a population. In education, it has been used as a measure of the inequality of universities. In chemistry it has been used to express the selectivity of protein kinase inhibitors against a panel of kinases. In engineering, it has been used to evaluate the fairness achieved by Internet routers in scheduling packet transmissions from different flows of traffic.