

Notes on the Gini Index Decomposition

Note sulla Scomposizione dell'Indice di Gini

Michele Costa

Università di Bologna, Dipartimento di Scienze Statistiche, Via Belle Arti 41, Bologna,
email: costa@stat.unibo.it

Riassunto: Il rapporto di concentrazione di Gini è da tempo al centro di un intenso dibattito relativo alla scomposizione della disuguaglianza complessiva nelle due componenti di disuguaglianza “entro” e “tra”. Tradizionalmente, la disuguaglianza “tra” è misurata seguendo un percorso analogo all’analisi della varianza, e quindi sulla base delle sole medie dei gruppi. Alternativamente, viene suggerito di considerare, nella misura della disuguaglianza “tra”, tutti gli elementi di diversità tra le distribuzioni dei gruppi. In questo secondo contesto, il lavoro propone una semplificazione della misura della disuguaglianza “tra”. Si dimostra, in particolare, che la disuguaglianza attribuibile alla diversità tra due gruppi è uguale alla differenza tra la quota del gruppo più povero nella popolazione e la quota del gruppo più povero nella variabile oggetto di studio.

Keywords: Gini index; inequality decomposition; inequality between

1. Introduction

Gini inequality index (Gini, 1939, 1959) represents one of the most used and widespread inequality measure, both in methodological studies and in applied researches. Among its advantages with respect to entropy inequality measures, such as Theil, Hirschman-Herfindahl and Bourguignon indexes, there is, for example, the possibility to include also negative and null values of the variable object of the analysis. In the case of a population divided in k subgroups, a long standing tradition states that inequality indexes decomposition is obtained in the framework of analysis of variance, and therefore inequality between subgroups is calculated as a function of only the means of the k subgroups. Since contribution by Bhatthacharya and Mahalanobis (1967), also Gini index is usually decomposed following this approach. Furthermore, the property of additive decomposition, introduced by Shorrocks (1980) and valid for entropy indexes, but not for Gini index, greatly contributed to increase the use of entropy indexes in inequality and poverty studies. For a throughout discussion about Gini index decomposition into inequality within and inequality between, its limits and its interpretation, see Giorgi (1999). However, measuring the contribution to total inequality due to the differences among population subgroups only on the basis of subgroups means completely disagree with the most common and evident features of income and wealth distributions. Typically subgroups income or wealth distributions strongly differ not only for the mean, but also for variance and asymmetry. And it is straightforward to note how also analysis of variance does hold only under the hypothesis of homoscedasticity. Furthermore, it is possible to demonstrate (Dagum, 1997b) how the measure of inequality between obtained in the framework of entropy indexes, for example for the Theil index, notwithstanding its additive decomposition,

leads to underestimate the real contribute to total inequality given by the differences among population subgroups. A final element of analysis which it is worth to consider refers to the relevance attributed to the additive decomposition hypothesis. However mathematically elegant and appealing, this property seems to be uncritically adopted if compared with either logical reference of distance between population subgroups and empirical analysis of subgroups distributions. In neither event a measure obtained only on the basis of subgroups means can be considered as acceptable and satisfactory.

A decomposition of the Gini index able to provide a more complete and correct measurement of inequality between component is suggested by Dagum (1997a), by taking into account not only the differences between the means of population subgroups, but all differences between subgroups distributions. Starting from this decomposition, the paper introduces a relevant simplification of the measurement of inequality between for the case of a population divided in two subgroups.

2. Methodology

In a population of n units, with income (or wealth, ...) vector $\{y_1, y_2, \dots, y_n\}$ and income mean $\bar{y}$, disaggregated in k subgroups of n_j units, with $n = \sum_{j=1}^k n_j$, and income mean $\bar{y}_j$, the Gini index is:

$$G = \frac{1}{2n^2 \bar{y}} \sum_{j=1}^k \sum_{h=1}^k \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|$$

By setting the subgroup j population share $p_j = n_j / n$, and the subgroup j income share $s_j = p_j \bar{y}_j / \bar{y}$, the Gini index between subgroup j and subgroup h can be expressed as

$$G_{jh} = \frac{1}{n_j n_h (\bar{y}_j + \bar{y}_h)} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|$$

and the Gini index can be obtained as a weighted sum of G_{jh} with weight $p_j s_h$

$$G = \sum_{j=1}^k \sum_{h=1}^k G_{jh} p_j s_h$$

In this framework the inequality within the k subgroups can be easily expressed as

$$G_w = \sum_{j=1}^k G_{jj} p_j s_j$$

i.e. the sum of the k Gini indexes in the k subgroups weighted by $p_j s_j$. This measure of the inequality within is quite generally accepted in the literature, where the inequality between subgroups is usually obtained as variant of Bhatthacharya and Mahalanobis (1967)

$$G_b = \sum_{j=1}^k \sum_{\substack{h=1 \\ j \neq h}}^k p_j p_h | \bar{y}_j - \bar{y}_h |$$

An alternative proposal (Dagum, 1997a; Mehran, 1975) in order to measure the contribution to total inequality attributable to the differences between the k population subgroups is

$$G_b = \sum_{j=1}^k \sum_{\substack{h=1 \\ j \neq h}}^k G_{jh} p_j s_h$$

that is a quantity which takes into account not only the differences between the income means of the k population subgroups, but also all other possible differences. Dagum suggests a further decomposition of the inequality between subgroups distinguishing the net Gini inequality between and the contribution of the intensity of transvariation; more detailed discussion is provided in the original references.

3. Results and conclusions

The analysis and the measurement of the inequality in a population divided in only two subgroups collects a huge anthology of preeminent situations in many fields.

By referring to the context of previous paragraph, the Gini index decomposition for the case of two subgroups results

$$G = (G_1 p_1 s_1 + G_2 p_2 s_2) + (G_{12} p_1 s_{12} + G_{21} p_2 s_1)$$

The purpose of the paper is to demonstrate that it is possible to obtain a relevant simplification of the component $(G_{12} p_1 s_{12} + G_{21} p_2 s_1)$, related to the measurement of the contribution to total inequality attributable to the differences between the two subgroups. When the population is divided in only two subgroups it results

$$p_2 = 1 - p_1, \quad s_2 = 1 - s_1, \quad G_{12} = G_{21}$$

By using these properties, and assuming that the poor are in the first group and the rich in the second, i.e. that $|y_{1i} - y_{2r}| = (y_{2r} - y_{1i}) \quad \forall i, r$, the measurement of the inequality between component can be simplified into:

$$\begin{aligned} G_b &= G_{12} p_1 s_2 + G_{21} p_2 s_1 = G_{12} (p_1 s_2 + p_2 s_1) = \\ &= \frac{p_1 s_2 + p_2 s_1}{n_1 n_2 (\bar{y}_1 + \bar{y}_2)} \sum_{i=1}^{n_1} \sum_{r=1}^{n_2} |y_{1i} - y_{2r}| = \frac{p_1 s_2 + p_2 s_1}{n_1 n_2 (\bar{y}_1 + \bar{y}_2)} \sum_{i=1}^{n_1} \sum_{r=1}^{n_2} (y_{2r} - y_{1i}) = \\ &= \frac{p_1 s_2 + p_2 s_1}{n_1 n_2 (\bar{y}_1 + \bar{y}_2)} n_1 n_2 (\bar{y}_2 - \bar{y}_1) = \frac{p_1 s_2 + p_2 s_1}{(\bar{y}_1 + \bar{y}_2)} (\bar{y}_2 - \bar{y}_1) = \end{aligned}$$

$$\begin{aligned}
&= p_1(s_2 + p_2\bar{y}_1/\bar{y}) \frac{(\bar{y}_2 - \bar{y}_1)}{(\bar{y}_1 + \bar{y}_2)} = p_1 p_2 \frac{(\bar{y}_2 + \bar{y}_1)}{\bar{y}} \frac{(\bar{y}_2 - \bar{y}_1)}{(\bar{y}_1 + \bar{y}_2)} = \\
&= p_1 p_2 \bar{y}_2 / \bar{y} - p_1 p_2 \bar{y}_1 / \bar{y} = p_1(1 - s_1) - (1 - p_1)s_1 = \\
&= p_1 - p_1 s_1 - s_1 + p_1 s_1 = p_1 - s_1
\end{aligned}$$

That is the inequality between measured as the difference of the poor share in the population, p_1 , and the poor share in the variable of interest, s_1 . A similar result is also achieved by Yitzhaki (2002), decomposing Gini index into Sen's poverty index and other components.

Furthermore, this simplification is robust with respect to the presence of transvariation between the two groups, that is, also overlapping subgroups distributions lead to the same measurement of the inequality between. Therefore, for the case of only two population subgroups, the Gini inequality index can be easily decomposed as

$$G = (G_1 p_1 s_1 + G_2 p_2 s_2) + (p_1 - s_1)$$

where the contribute to total inequality given by the differences between the two subgroups seems far more correctly measured than by referring to the traditional simple difference between the means of the two subgroups.

The measurement of inequality between performed by resorting to the Gini index decomposition allows to avoid the major difficulties related to the traditional income inequality measures decompositions and, at the same time, allows to correctly evaluate and compare the effects of different factors on total inequality. When a population is divided in two groups, Gini inequality index enables to obtain an immediate and straightforward measure of the contribution to total inequality attributable to the differences between the groups.

References

- Bhattacharya N., Mahalanobis B. (1967) Regional disparities in household consumption in India, *Journal of the American Statistical Association*, 62, 143-161.
- Dagum C. (1997a) A new decomposition of the Gini income inequality ratio, *Empirical Economics*, 22, 515-531.
- Dagum C. (1997b) Decomposition and interpretation of Gini and Theil entropy inequality measures, *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 200-205.
- Gini C. (1939) *Variabilità e concentrazione*, Giuffrè, Milano.
- Gini C. (1959) *Transvariazione*, Libreria goliardica, Roma.
- Giorgi G. M. (1999) Income inequality measurement: the statistical approach, in: *Handbook on Income Inequality Measurement*, Silber, J. (Ed.), Kluwer, 245-267.
- Mehran F. (1975) A statistical analysis of income inequality based on a decomposition of the Gini index, *Proceedings of the 40th ISI Session, Warsaw*, 580-585.
- Shorrocks A. F. (1980) The class of additively decomposable inequality measures, *Econometrica*, 48, 613-625.
- Yitzhaki S. (2002) Do we need a separate poverty measurement?, *European Journal of Political Economy*, 61-85.