D I D A C T I C S O F M A T H E M A T I C S

No. 12 (16) 2015

THE GINI COEFFICIENT AS A MEASURE OF DISPROPORTIONALITY*

Piotr Dniestrzański

Abstract. Measures of inequality, properly adapted, often tend to be used as a tool to address the issue of disproportionality. The most popular of them, such as the Gini or Atkinson coefficient, or entropy coefficient can, under certain circumstances, act as measures of disproportionality. However, one must specify precisely what is to be measured and interpret the results consistently. In this paper we analyze what confusion or outright errors can be committed when using inequality coefficients. The presented analysis is aimed at the Gini coefficient, however, the problem also applies to the rest of the coefficients.

Keywords: inequality measure, the Gini coefficient, mathematics teaching.

JEL Classification: D31, D63.

DOI: 10.15611/dm.2015.12.03.

1. Introduction

Measures of disproportionality may be helpful in the assessment of the degree of disproportionality of a given allocation of goods or burdens. Evaluating the disproportionality with the use of adopted measures of equality is commonly known. Let us suppose that there are two vectors $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$. Throughout the work, we will assume that the analyzed vectors have non-zero coordinates and are not zero vectors, which means that they belong to a set of $\mathbb{R}_+^n$. We put forth the problem of assessing the degree of disproportionality of vectors x and y. These vectors are strictly proportionate if there is a real number α such that $y = \alpha x$.

Piotr Dniestrzański
Department of Mathematics and Cybernetics, Wrocław University of Economics
piotr.dniestrzanski@ue.wroc.pl

* The results presented in this paper have been supported by the Polish National Science Centre under grant no. 2013/09/B/HS4/02702.

From the assumption that the $x, y \in \mathbb{R}_{+}^{n}$ it follows, of course, that the coefficient α is different from zero. Equality $y = \alpha x$ is equivalent to the equality $x = \frac{1}{\alpha} y$. If the vectors x and y are strictly proportionate then there occur equalities $\frac{y_i}{x_i} = \alpha$ for all $i = 1, 2, ..., n$. The disorder of the proportionality causes the quotients $\frac{y_i}{x_i}$ are not the same for all i. In such a situation, it seems to be reasonable to assess the degree of disproportionality of vectors x and y with the use of the degree of inequality of vector $\frac{y}{x} = \left(\frac{y_1}{x_1}, \frac{y_2}{x_2}, ..., \frac{y_n}{x_n} \right)$. This approach to analyze disproportionality is widely accepted. It is even considered that, for example, the Gini coefficient [Karpov 2008] is in this case an appropriate tool. Discussions and examples of such an adaptation of the Gini coefficient can be found, inter alia, in the work of [White 1986; Taagepera, Shugart 1989; Monroe 1994; Taagepera, Grofman 2003]. The application in this case of the Gini coefficient has a major flaw, it is burdened with a certain ambiguity. Measure of disproportionality constructed in such a way does not meet the condition of symmetry which is necessary in the analysis of disproportionality. This means that the measures of inequality of vectors $\frac{x}{y}$ and $\frac{y}{x}$ are mostly not equal. What is more, if one of the coordinates of the vector y is equal to zero then the quotient $\frac{x}{y}$ is incorrectly defined.

2. The Gini coefficient

The Gini coefficient is one of the most famous and widely used measures of inequality. It has been present in the world of science for over a hundred years [Gini 1912], and has been included in many thousands of scientific papers in the form of monographs and papers. It is used mainly as a tool to study the degree of social and economic inequalities. The main area of use of the Gini coefficient is the analysis of income inequality. Analysis of the Gini coefficient's features, possibility of applications and comparisons with other measures of inequality can be found inter alia in [Cowell 2011]. The Gini coefficient and the associated Lorenz curve are a canon in most academic courses and textbooks on statistics [Ostasiewicz 2011; Starzyńska 2006]. However, none of the studies known to me analyses the issue that is under consideration in this study.

There are many formulas that you can use to calculate the value of the Gini coefficient. Some of these formulas and their authors can be found in [Ceriani, Verme 2015]. In this study we will use the figure proposed by Kendal and Stuart [1958]:

$$G(a) = \frac{\sum_{i,j=1}^{n} \left| a_i - a_j \right|}{2n^2 \overline{a}} , \tag{1}$$

where $a = \left(a_1, a_2, ..., a_n\right) \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}$. The value of $G(a)$ belongs to the range $\left[0, \frac{n-1}{n}\right]$. The closer to zero the value of the $G(a)$ is, the smaller the inequality of the vector[1] – a. The closer the value of $G(a)$ is to the right end of the interval $\left[0, \frac{n-1}{n}\right]$, the bigger the inequality. The Gini coefficient is usually calculated for vectors with a high number of coordinates, so one can recognize that it takes values in range $\left[0,1\right)$.

The Gini coefficient satisfies many properties which are expected of the inequality coefficients. These include, for example:

A1. **Scale Independence**: $G(\lambda a) = G(a)$ for all $x \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}$ and $\lambda > 0$.

A2. **Symmetry**: $G(a^\theta) = G(a)$ for every permutation θ.

The paper by [Plata et al. 2015] is the first one to provide an elementary characterization of the Gini coefficient. The authors demonstrated that the Gini coefficient is the only measure of inequality which meets four natural properties. In addition to those listed above (A1 and A2), the features are (the authors call them axioms of): standarization and comonotone separability. In this paper we examine only cases in which the coordinates of the corresponding vectors are non-negative. In the literature [Raffinetti et al.

[1] By the inequality of the vector we understand the inequality of its next coordinates. A vector with zero inequality has all the same coordinates.

2014], there are also considerations of cases where some of the coordinates are negative. An index is then defined, which is a generalization of the Gini coefficient.

3. Disproportionality and the Gini coefficient

We will now present specific examples of what sort of ambiguities can occur when applying the Gini coefficient to analyze the matter of disproportionality.

Example 1. Consider the vectors $x=(1,2,4)$ and $y=(2\,3,5)$. They are not strictly proportionate. Let us assess the degree of their disproportionality using the Gini coefficient as described in the introduction. We then have $\frac{y}{x}=\left(\frac{y_1}{x_1},\frac{y_2}{x_2},\frac{y_3}{x_3}\right)=\left(2,\frac{3}{2},\frac{5}{4}\right)$. The Gini coefficient of the vector $\frac{y}{x}$ is equal to $G\left(2,\frac{3}{2},\frac{5}{4}\right)=G(8,6,5)=\frac{8}{57}\approx 0.1053$. The degree of disproportionality of vectors x and y estimated using the degree of inequality of vector $\frac{y}{x}$ differs from the one estimated using inequality of vector $\frac{x}{y}$. We have therefore:

$$\frac{x}{y}=\left(\frac{x_1}{y_1},\frac{x_2}{y_2},\frac{x_3}{y_3}\right)=\left(\frac{1}{2},\frac{2}{3},\frac{4}{5}\right)$$

and

$$G\left(\frac{1}{2},\frac{2}{3},\frac{4}{5}\right)=G(15,20,24)=\frac{6}{59}\approx 0.1017\,.$$

The discrepancies that were shown in Example 1 may be much greater. In Example 2 we see that this difference may be extremely large.

Example 2. Take vectors $\mathbf{1}_n=(1,1,\ldots,1)$ and $x=(k,1,1,\ldots,1)$ from a set of $\mathbb{R}_+^n$, where $k>0$. Then we have $\frac{x}{\mathbf{1}_n}=x$ and $\frac{\mathbf{1}_n}{x}=\left(\frac{1}{k},1,1,\ldots,1\right)$. We will assess the border values for the Gini coefficient for vectors $\frac{x}{\mathbf{1}_n}=x$ and $\frac{\mathbf{1}_n}{x}$

with k tends to infinity. In the calculation we will use the property of A1 and A2, i.e. the insensitivity of the Gini coefficient to permutation and scaling:

$$\lim_{k\to\infty} G\left(\frac{x}{\mathbf{1}_n}\right) = \lim_{k\to\infty} G(k,1,1,...,1) = \lim_{k\to\infty} G\left(1,\frac{1}{k},\frac{1}{k},...,\frac{1}{k}\right) =$$

$$G(1,0,0,...,0) = \frac{n-1}{n} = 1-\frac{1}{n},$$

$$\lim_{k\to\infty} G\left(\frac{\mathbf{1}_n}{x}\right) = \lim_{k\to\infty} G\left(\frac{1}{k},1,1,...,1\right) = G(0,1,1,...,1) = \frac{1}{n}.$$

Therefore, $\lim_{k\to\infty}\left[G\left(\frac{x}{\mathbf{1}_n}\right) - G\left(\frac{\mathbf{1}_n}{x}\right)\right] = 1-\frac{2}{n}$ and $\lim_{\substack{k\to\infty\\ n\to\infty}}\left[G\left(\frac{\mathbf{1}_n}{x}\right) - G\left(\frac{x}{\mathbf{1}_n}\right)\right] = 1$.

In Example 2, we can see that the assessment of disproportionality of vectors made using the Gini coefficient can be extremely variable depending on how it was used for this purpose. Therefore, while using the Gini coefficient to estimate disproportionality we should mention the degree of disproportionality of the vector with relation to another vector, and not the disproportionality of a pair of vectors. This does not change the fact that estimating disproportionality with the use of a measure with such a flaw is rather unfortunate.

The shortcoming of the Gini coefficient as a measure of disproportionality presented in the example is not present in the case of 2D vectors with non-zero coordinates. It is easy to demonstrate the veracity of the following proposition.

Proposition 1. If the vectors x, y from space R^2 have non-zero coordinates, then the values of the Gini coefficient for vectors $\frac{x}{y}$ and $\frac{y}{x}$ are equal.

Application of the Gini coefficient as a way of evaluating vector disproportionality encounters yet another deficiency from the mathematical point of view. We will see this in Example 3.

Example 3. Let us consider the two pairs of vectors $x=(0,1)$, $y=(1,1)$ and $a=(0,0,1)$, $b=(1,1,1)$. Then we will calculate the vector disproportionality for x and y as well as a and b. We then obtain

$G\left(\frac{x}{y}\right)=G(0,1)=0,5$ and $G\left(\frac{a}{b}\right)=G(0,0,1)=0,67$. Assessing on the basis of the Gini coefficient the degrees of proportionality for vectors x and y as well as a and b ,we come to the conclusion that more proportionate are the vectors x and y than a and b. This is in conflict with the idea of proportionality understood as a linear dependency between the vectors. Going further we will arrive at the conclusion that a degree of disproportionality for vectors $c=(0,0,...,0,1)\in R^n$ and $d=(1,1,...,1)\in R^n$ by $n\to\infty$ tends to unity.

4. The Gini coefficient and the European Parliament

Proportional division is one of the main elements of the subject of the distribution of goods and burdens in contemporary societies. It appears, inter alia, in the matter of distribution of seats in collegial bodies. For example, the Polish Constitution says that "the elections to Sejm shall be universal, equal, direct and **proportional** and shall be conducted by secret ballot". The Constitution does not define, however, how the effect of proportionality is to be achieved. There are relevant legal acts of lower rank dealing with this. Proportional distribution becomes troublesome in the case of goods which are indivisible, for example the already mentioned, seats in collegial bodies. Strict proportionality warrants almost always assigning non-integer values. It is obvious that in such a situation the fractional values are rounded to the integer values. This often results in problems as there are a lot of possibilities for such roundings. Some proportional distribution methods are susceptible to so-called paradoxes. For example, the method of the largest reminder (Hamilton's method) is sensitive to the so-called Alabama paradox[2]. If the ideal required distribution in a given problem is a proportional distribution and, at the same time, it is not possible to achieve, one can instead use a distribution method similar to the desired. What remains to be agreed in this situation is the question of how to measure which of the distribution methods is the closest to the ideal proportion.

[2] The Alabama Paradox was discovered in 1880 in the USA. It was noted then that an increase in the size of the United States House of Representatives from 299 to 300 would result in the State of Alabama losing one mandate. The discovery was one of the reasons for the House of Representatives to abandon (in 1911) this method of distribution of seats in favor of another proportional method proposed by Webster.

Different values of the Gini coefficient for the same pair of vectors can be useful if these vectors represent the data that we can interpret. Let us look at the distribution of seats in the European Parliament (EP) among the Member States of the European Union (EU). Table 1 shows populations and the number of seats in the EP for all of the Member States within the term 2014-2019. Let x_i be a population of Member State number i and y_i the number of seats held by that State. How to estimate the level of disproportionality[3] of this allocation using the Gini coefficient? Following the trail of earlier considerations we can calculate the value of the Gini coefficient for vectors $\frac{x}{y}$ and $\frac{y}{x}$. They are $G\left(\frac{x}{y}\right) = 0{,}1889$ and $G\left(\frac{y}{x}\right) = 0{,}3076$ respectively. In this case, the calculated values of the coefficients can be easily interpreted. Vector $\frac{x}{y}$ inequality is a differentiation of the number of citizens per one seat in EP distinct to the individual Member States. Vector inequalities $\frac{y}{x}$ is a differentiation of the amount of seats in EP per capita, again distinct to the individual Member States. Hence, there are two different kinds of inequality. In each of the cases the value of the Gini coefficient will be zero if and only if the distribution is strictly proportionate.

In addition to the above two, in the case of the distribution of seats in the EP, one can calculate the Gini coefficient in yet another way. Let us look at the citizens of the EU as one group of people. The number of seats which is assigned per capita can be treated as a kind of "income" and the question can be asked: what is the degree of inequality of that "income"? It is distributed unevenly, as for example any citizen of Malta has an "income" in the amount of $6/416110$ (number of seats for Malta divided by the population of Malta). Similarly, we define the income for the citizens of the rest of the Member States. We get a vector with the number of coordinates equal to the quantity of the EU population. The Gini coefficient designated for such vector is $G(EP) = 0{,}1692$.

[3] The distribution of seats in the EP is not strictly proportional. It is, in accordance with the provisions of the Treaty of Lisbon, of a degressively proportional nature. This is the result of the too big variations in the populations of the Member States, which makes it impossible to use any of the methods of proportional allocation. Under this restriction, UE countries seek distribution closest to the proportional. An analysis of how the distribution of seats in the EP can be proportioned is to be found inter alia in [Dniestrzański, Łyko 2014; Łyko 2012].

Table 1. The distribution of seats in the EP between the Member States of EU in the term 2014-2019 and the Gini coefficient for ratios population/seats and seats/population

Member State	Population *x*	Seats 2014-2019 *y*	*x*/*y*	*y*/*x*
Germany	81 843 743	96	852539	0,0000012
France	65 397 912	74	883756	0,0000011
United Kingdom	62 989 550	73	862871	0,0000012
Italy	60 820 764	73	833161	0,0000012
Spain	46 196 276	54	855487	0,0000012
Poland	38 538 447	51	755656	0,0000013
Romania	21 355 849	32	667370	0,0000015
Netherlands	16 730 348	26	643475	0,0000016
Greece	11 290 935	21	537664	0,0000019
Belgium	11 041 266	21	525775	0,0000019
Portugal	10 541 840	21	501992	0,0000020
Czech Republic	10 505 445	21	500259	0,0000020
Hungary	9 957 731	21	474178	0,0000021
Sweden	9 482 855	19	499098	0,0000020
Austria	8 443 018	19	444369	0,0000023
Bulgaria	7 327 224	17	431013	0,0000023
Denmark	5 580 516	13	429270	0,0000023
Slovakia	5 404 322	13	415717	0,0000024
Finland	5 401 267	13	415482	0,0000024
Ireland	4 582 769	11	416615	0,0000024
Croatia	4 398 150	11	399832	0,0000025
Lithuania	3 007 758	11	273433	0,0000037
Latvia	2 055 496	8	256937	0,0000039
Slovenia	2 041 763	8	255220	0,0000039
Estonia	1 339 662	6	223277	0,0000045
Cyprus	862 011	6	143669	0,0000070
Luxembourg	524 853	6	87476	0,0000114
Malta	416 110	6	69352	0,0000144
Total	**508 077 880**	**751**		
Gini			0,1889	0,3076

Source: own elaboration.

From the above analysis, it can be concluded that the Gini coefficient can be utilized in some notions of research of the degree of disproportionality in several ways. In any case, its value is usually different and can be properly interpreted.

5. Summary

The Gini coefficient can be used as a measure of disproportionality. In using it as a tool for this purpose, however, one should be aware of the limitations and ambiguity. Certainly it cannot be used as a measure of disproportionality in a purely mathematical sense since the property of symmetry is not met. When providing the value of the Gini coefficient in disproportionality analyses, one has to specify precisely how it was used. For example, the sentence "the Gini coefficient for the allocation of seats in the EP is..." is not precise enough. This specific property of the Gini coefficient (and any other measure of inequality) used in the notion of disproportionality should be clearly emphasized in the course of statistics.

References

Ceriani L., Verme P. (2015). *Individual diversity and the Gini decomposition*. Social Indicators Research 121. Pp. 637-646.

Cowell F. (2011). *Measuring Inequality*. Oxford University Press.

Dniestrzański P., Łyko J. (2014). *Influence of boundary conditions of digressively proportional division on the potential application of proportional rules*. Procedia – Social and Behavioral Sciences 109. Pp. 722-729.

Gini C. (1912). *Variabilita e mutabilita: contributo allo studio delle relazioni statistiche*. Studi Economico-giurdici. Facolta di Giurisprudenza della R. Universita di Cagliari. Anno III. Cuppini. Bologna.

Karpov A. (2008). *Measurement of disproportionality in proportional representation systems*. Mathematical and Computer Modelling 48. Pp. 1421-1438.

Kendall M.G., Stuart A. (1958). *The Advanced Theory of Statistics* (1st ed., vol. 1). Hafner Publishing Company. New York.

Łyko J. (2012). *The boundary conditions of degressive proportionality*. Procedia – Social and Behavioral Sciences 65. Pp. 76-82

Monroe B.L. (1994). *Disproportionality and malapportionment: Measuring electoral inequity*. Electoral Studies 13. Pp. 132-49.

Ostasiewicz W. (2011). *Badania statystyczne*. Wolters Kluwer.

Plata-Pérez L., Sánchez-Pérez J., Sánchez-Sánchez F. (2015). *An elementary characterization of the Gini index.* Mathematical Social Sciences 74. Pp. 79-83.

Raffinetti E., Silletti E., Vernizzi A. (2014). *On the Gini coefficient normalization when attributes with negative values are considered.* Statistical Methods & Applications.

Starzyńska W. (2006). *Statystyka praktyczna*. Wydawnictwo Naukowe PWN.

Taagepera R., Grofman B. (2003). *Mapping the indices of seats-votes disproportionality and inter-election volatility*. Party Politics 9(6). Pp. 659-677.

Taagepera R., Shugart M. (1989). *Seats and Votes: The Effects and Determinants of Electoral Systems.* Yale University Press. New Haven.

White M.J. (1986). *Segregation and diversity measures in population distribution*. Population Index 52. Pp. 193-221.