Communications in Mathematics

Marian Genčev

A note on a property of the Gini coefficient

Communications in Mathematics, Vol. 27 (2019), No. 2, 81–88

Persistent URL: http://dml.cz/dmlcz/147983

Communications in Mathematics 27 (2019) 81–88
DOI: 10.2478/cm-2019-0008

A note on a property of the Gini coefficient

Marian Genčev

Abstract. The scope of this note is a self-contained presentation of a mathematical method that enables us to give an absolute upper bound for the difference of the Gini coefficients

$$|G(\sigma_1, \dots, \sigma_n) - G(\gamma_1, \dots, \gamma_n)|,$$

where $(\gamma_1, \dots, \gamma_n)$ represents the vector of the gross wages and $(\sigma_1, \dots, \sigma_n)$ represents the vector of the corresponding super-gross wages that is used in the Czech Republic for calculating the net wage. Since (as of June 2019) $\sigma_i = 100 \cdot \lceil 1.34\gamma_i/100 \rceil$, the study of the above difference seems to be somewhat inaccessible for many economists. However, our estimate based on the presented technique implies that the introduction of the super-gross wage concept does not essentially affect the value of the Gini coefficient as sometimes expected.

1 Introduction and motivation

For governments of most countries in the world, the analysis of the income distribution is an important tool as it can directly reflect the economic policy impact in the given country. In virtue of this, it is necessary to know the basic tools for measuring the so-called income inequality. The well-known measures are, e.g., the Gini coefficient, Theil index, Hoover index, and many other indicators can be found in the appropriate literature (see Atkinson and Bourguignon [2]). These measures can especially be used for assessments of the redistribution impact of the tax and public cash transfer system of every country and, therefore, their role in modern economics is indispensable.

In this paper, we focus on a special property of the measure for the income distribution inequality represented with the famous Gini coefficient that requires a convenient mathematical argumentation to deduce our main result. Since the

2010 MSC: 91B82, 26B35
Key words: Gini coefficient, finite sums, estimates
Affiliation:
VŠB – Technical University Ostrava, Faculty of Economics, Sokolská 33, 702 00 Ostrava, Czech Republic
E-mail: marian.gencev@vsb.cz

objectives of this paper require some technical details, we rather start this section with a methodological background that is followed by a concrete real motivation.

1.1 The Gini coefficient

Perhaps the most used index for measuring the income distribution inequality is the well-known Gini coefficient introduced by the statistician C. Gini in his famous book [5]. Selected extracts from Gini's book (published in Italian) were provided in 2012 by Ceriani and Verme [3] who described Gini's original suggestions of his index.

At present, we usually define the Gini coefficient of n non-negative values (incomes) $x_1, \ldots, x_n$ as

$$G(x_1, \ldots, x_n) := \frac{1}{2n\sum_{i=1}^{n} x_i} \cdot \sum_{1 \le i,j \le n} |x_i - x_j| \tag{1}$$

with $\sum_{i=1}^{n} x_i > 0$; for further details, see, e.g., Allison [1, p. 867], Genčev *et al.* [4], Lambert [6, p. 34] or Sen [9, p. 31]. Its calculation by using the defining expression (1) is, however, very inefficient due to the high computational complexity (the number of algebraic operations grows quadratically with n).

For a fast calculation of the Gini coefficient, we most frequently use the equivalent formula

$$G(x_1, \ldots, x_n) = \frac{2\sum_{i=1}^{n} i x_i}{n\sum_{i=1}^{n} x_i} - 1 - \frac{1}{n}, \tag{2}$$

where it is assumed that the non-negative income sequence $\{x_i\}_{i=1}^{n}$ is non-decreasing, i.e.,

$$0 \le x_1 \le x_2 \le \cdots \le x_n, \qquad x_n > 0.$$

Although identity (2) is widely known and generally accepted by economists, a complete mathematical proof of it does not seem to be presented elsewhere. E.g., Zenga *et al.* [10, p. 464] state without any detail that the double sum in (1) 'can be rewritten' to a specific one-dimensional sum. Sen [9, p. 30] even comments that the transformation of the double sum in question is rather 'tedious'. On the other hand, an interesting axiomatic characterization of the Gini coefficient containing a formula equivalent to (2) is given in Plata-Peréz *et al.* [8, p. 83] but the connection to (1) is not mentioned.

Since (2) will constitute our main tool in this paper, we give (among our main result) a short mathematical proof of formula (2) to make our paper as self-contained as possible.

1.2 The concept of the super-gross wage

The Czech Republic is the only country using the concept of the super-gross wage σ that is calculated from the gross wage γ. The gross wage γ is an individual's total personal income before taking taxes or deductions into account. In contradistinction to this, the super-gross wage σ is the gross wage γ increased by the social and health contributions[1] (25 % and 9 % respectively) paid by the employer.[2] Since

[1] The status from June 2019.

[2] For the precise calculation, see (3).

its introduction in the Czech Republic in 2008, its non-standard nature has often been discussed by economists.

The mathematical model for calculating σ (depending on γ) is presented by the relation

$$\sigma := 100 \cdot \left\lceil \frac{1.34\gamma}{100} \right\rceil, \tag{3}$$

where $\lceil z \rceil$ denotes the ceiling function that maps z to the least integer that is greater or equal to z. This means that the super-gross wage σ is equal to the amount 1.34γ that is rounded up to the whole hundreds. For our later needs, we introduce the quantity

$$d := \sigma - 1.34\gamma \tag{4}$$

that represents the rounding increase. In virtue of this and (3), we easily deduce that

$$0 \le d < 100. \tag{5}$$

1.3 Motivation and structure of the paper

The motivation for starting our research was a series of controversies of several economists that were not sure about the absolute change of the Gini coefficient when its calculation is performed with the vector of gross wages $\boldsymbol{\gamma}$ instead of the vector of super-gross wages $\boldsymbol{\sigma}$. Therefore, our aim was to develop a rigorous technique that definitely enable us to find a concrete absolute upper bound for the quantity $|\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma})|$, where

$$\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma}) := G(\sigma_1, \dots, \sigma_n) - G(\gamma_1, \dots, \gamma_n), \tag{6}$$

that only depends on the current amount of the minimum gross wage in the Czech Republic (see Corollary 1).

Unfortunately, the described procedure of calculating the super-gross wage σ with the specific kind of rounding makes the study of the difference $\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma})$ somewhat inaccessible. Note that if the introduced rounding would simply be neglected, i.e.,

$$\boldsymbol{\sigma}' = 1.34\, \boldsymbol{\gamma} = (1.34\gamma_1, \dots, 1.34\gamma_n),$$

then, by the homogeneity of the Gini coefficient, the corresponding Gini coefficients are equal each other. On the other hand, the rounding procedure certainly affects the value of the Gini coefficient and, consequently, the difference $\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma})$ will not be generally equal to zero (in fact, $\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma})$ can be both, positive and negative).

From the practical point of view, when assessing the global progression of tax and public cash transfer measured by the Musgrave-Thin index (see [7])

$$M := \frac{1 - G_{\text{after}}}{1 - G_{\text{before}}},$$

it is insignificant if we put $G_{\text{before}} = G(\gamma_1, \dots, \gamma_n)$ or, alternatively, $G_{\text{before}} = G(\sigma_1, \dots, \sigma_n)$ since their difference $\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma})$ is irrelevant by Corollary 1. Of course, when calculating the index M we usually set $G_{\text{after}} = G(\nu_1, \dots, \nu_n)$, where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)$ is the vector of the corresponding net wages.

The rest of the paper is organized in the following way. In Section 2, we present our main result (Theorem 1) containing a general upper bound of the quantity $|\Delta(\boldsymbol{\sigma},\boldsymbol{\gamma})|$, and the elegant Corollary 1 with an absolute bound. Section 3 is devoted to the proof of both statements from Section 2 and, for the sake of completeness, to the proof of the important identity (2).

Since our approach is rather elementary, the presented proofs are accessible for every researcher acquainted with basic sum manipulations.

2 Main result

The following assertion is our main result.

Theorem 1. *Let $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$ and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)$, respectively, be arbitrary two income vectors of gross wages and of the corresponding super-gross wages whose components are calculated by* (3). *Then*

$$|\Delta(\boldsymbol{\sigma},\boldsymbol{\gamma})| \le \frac{\sum_{i=1}^{n} \sigma_i}{0.67 \sum_{i=1}^{n} \gamma_i} - 2. \tag{7}$$

Note that the constant "2" in the estimation on the right-hand side of (7) is best possible since for certain special vectors $\boldsymbol{\gamma}$ the term on the right-hand side of (7) can vanish.

Remark 1 (the case of extremely small wages). Observe that for $\sum_{i=1}^{n} \gamma_i \to 0^+$, the expression on the right-hand side of (7) is not bounded from above. Since the difference of two arbitrary Gini coefficients is always less than 1, one can write (7) as

$$|\Delta(\boldsymbol{\sigma},\boldsymbol{\gamma})| \le \min\left(1, \frac{\sum_{i=1}^{n} \sigma_i}{0.67 \sum_{i=1}^{n} \gamma_i} - 2\right). \tag{7'}$$

However, the practical importance of the case $\sum_{i=1}^{n} \gamma_i \to 0^+$ is negligible and we utilize (7) instead of (7') in the rest of the paper.

Applying the current[3] minimum gross wage amount 13 350 CZK valid in the Czech Republic, Theorem 1 enables us to obtain the following elegant absolute bound documenting that the Gini coefficients under consideration are almost identical.

Corollary 1. *Let $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)$ and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)$ respectively, be arbitrary two income vectors of gross wages and of the corresponding super-gross wages whose components are calculated by* (3). *Then, by taking the current*[3] *amount of the minimum gross wage in the Czech Republic, we obtain*

$$|\Delta(\boldsymbol{\sigma},\boldsymbol{\gamma})| < \frac{1}{89}.$$

[3]The status from June 2019.

3 Proofs

3.1 Proof of identity (2)

Proof. Assume that the income sequence $\{x_i\}_{i=1}^n$ is non-decreasing. Then

$$\begin{aligned}\sum_{1\le i,j\le n} |x_i - x_j| &= \sum_{\substack{1\le i,j\le n\\ j<i}} |x_i - x_j| + \sum_{\substack{1\le i,j\le n\\ j>i}} |x_i - x_j| \\ &= 2\sum_{1\le j<i\le n} (x_i - x_j) \\ &= 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}(x_i - x_j) \\ &= 2\sum_{i=2}^{n}\left(\sum_{j=1}^{i-1} x_i - \sum_{j=1}^{i-1} x_j\right).\end{aligned}$$

Since, obviously,

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1} x_i = \sum_{i=1}^{n}(i-1)x_i,$$

and,

$$\sum_{i=2}^{n}\sum_{j=1}^{i-1} x_j = \sum_{i=1}^{n}(n-i)x_i$$

(the last one can be obtained by induction or by applying Abel partial summation), we deduce immediately that

$$\begin{aligned}\sum_{1\le i,j\le n} |x_i - x_j| &= 2\sum_{i=1}^{n}(i-1)x_i - 2\sum_{i=1}^{n}(n-i)x_i \\ &= 2\sum_{i=1}^{n}(i-1-n+i)x_i \\ &= 4\sum_{i=1}^{n} ix_i - 2(n+1)\sum_{i=1}^{n} x_i.\end{aligned}$$

Consequently, for $\sum_{i=1}^n x_i > 0$ we deduce from (1) that

$$\begin{aligned}G(x_1,\dots,x_n) &= \frac{1}{2n\sum_{i=1}^n x_i}\cdot\left(4\sum_{i=1}^{n} ix_i - 2(n+1)\sum_{i=1}^{n} x_i\right) \\ &= \frac{2\sum_{i=1}^n ix_i}{n\sum_{i=1}^n x_i} - 1 - \frac{1}{n},\end{aligned}$$

as required. □

3.2 Proof of the main theorem

For the proof of Theorem 1, we utilize the following auxiliary result.

Lemma 1. *Assume that* $\{\gamma_i\}_{i=1}^n$ *is a non-decreasing sequence with* $\gamma_i \geq 0$. *If*

$$\sigma_i = 100 \cdot \left\lceil \frac{1.34\gamma_i}{100} \right\rceil, \qquad i = 1, \ldots, n,$$

then the sequence $\{\sigma_i\}_{i=1}^n$ *is also non-decreasing.*

The proof of Lemma 1 is self-evident and we leave its detailed proof as an exercise for the reader.

Remark 2. Since the income sequences $\{\gamma_i\}_{i=1}^n$ and $\{\sigma_i\}_{i=1}^n$ are both non-decreasing by Lemma 1, we can apply formula (2) to both of them when investigating the difference $\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma})$. This simplifies our further investigation.

Proof. [Proof of Theorem 1] Suppose that the sequence $\{\gamma_i\}_{i=1}^n$, $\gamma_i \geq 0$, is non-decreasing. Then, according to Lemma 1, (2) and (6), we obtain

$$\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma}) = \frac{2}{n} \cdot \left(\frac{\sum_{i=1}^n i\sigma_i}{\sum_{i=1}^n \sigma_i} - \frac{\sum_{i=1}^n i\gamma_i}{\sum_{i=1}^n \gamma_i} \right). \tag{8}$$

Next, by (4), we construct the following estimate:

$$\begin{aligned}
\frac{\sum_{i=1}^n i\sigma_i}{\sum_{i=1}^n \sigma_i} - \frac{\sum_{i=1}^n i\gamma_i}{\sum_{i=1}^n \gamma_i} &= \frac{\sum_{i=1}^n 1.34i\gamma_i + \sum_{i=1}^n id_i}{\sum_{i=1}^n 1.34\gamma_i + \sum_{i=1}^n d_i} - \frac{\sum_{i=1}^n i\gamma_i}{\sum_{i=1}^n \gamma_i} \\
&\leq \frac{\sum_{i=1}^n 1.34i\gamma_i + n\sum_{i=1}^n d_i}{\sum_{i=1}^n 1.34\gamma_i} - \frac{\sum_{i=1}^n i\gamma_i}{\sum_{i=1}^n \gamma_i} \\
&= \frac{n\sum_{i=1}^n d_i}{\sum_{i=1}^n 1.34\gamma_i}.
\end{aligned}$$

Analogously, we deduce

$$\begin{aligned}
\frac{\sum_{i=1}^n i\sigma_i}{\sum_{i=1}^n \sigma_i} - \frac{\sum_{i=1}^n i\gamma_i}{\sum_{i=1}^n \gamma_i} &= \frac{\sum_{i=1}^n 1.34i\gamma_i + \sum_{i=1}^n id_i}{\sum_{i=1}^n 1.34\gamma_i + \sum_{i=1}^n d_i} - \frac{\sum_{i=1}^n i\gamma_i}{\sum_{i=1}^n \gamma_i} \\
&\geq \frac{\sum_{i=1}^n 1.34i\gamma_i}{\sum_{i=1}^n 1.34\gamma_i + \sum_{i=1}^n d_i} - \frac{\sum_{i=1}^n i\gamma_i}{\sum_{i=1}^n \gamma_i} \\
&= \frac{-\left(\sum_{i=1}^n d_i\right) \cdot \left(\sum_{i=1}^n i\gamma_i\right)}{\left(\sum_{i=1}^n 1.34\gamma_i + \sum_{i=1}^n d_i\right) \cdot \left(\sum_{i=1}^n \gamma_i\right)}.
\end{aligned}$$

Consequently, from (8) and from the previous estimates we obtain

$$\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma}) \leq \frac{2}{n} \cdot \frac{n\sum_{i=1}^n d_i}{\sum_{i=1}^n 1.34\gamma_i} = \frac{1}{0.67} \cdot \frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n \gamma_i},$$

and,

$$\begin{aligned}\Delta(\boldsymbol{\sigma},\boldsymbol{\gamma}) &\geq \frac{2}{n}\cdot\frac{-\left(\sum_{i=1}^n d_i\right)\cdot\left(\sum_{i=1}^n i\gamma_i\right)}{\left(\sum_{i=1}^n 1.34\gamma_i+\sum_{i=1}^n d_i\right)\cdot\left(\sum_{i=1}^n \gamma_i\right)}\\ &\geq \frac{-2\sum_{i=1}^n d_i}{\sum_{i=1}^n 1.34\gamma_i+\sum_{i=1}^n d_i}\\ &\geq -\frac{1}{0.67}\cdot\frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n \gamma_i}.\end{aligned}$$

Finally, the relation in (4) and the last two estimates of $\Delta(\boldsymbol{\sigma},\boldsymbol{\gamma})$ imply that

$$\begin{aligned}|\Delta(\boldsymbol{\sigma},\boldsymbol{\gamma})| &\leq \frac{1}{0.67}\cdot\frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n \gamma_i} \qquad (9)\\ &= \frac{1}{0.67}\cdot\frac{\sum_{i=1}^n (\sigma_i-1.34\gamma_i)}{\sum_{i=1}^n \gamma_i}\\ &= \frac{\sum_{i=1}^n \sigma_i}{0.67\sum_{i=1}^n \gamma_i}-2.\end{aligned}$$

This concludes the proof of the main theorem. □

3.3 Proof of Corollary 1

Proof. Since the current minimum gross wage in the Czech Republic is assessed by the amount 13 350 CZK, we obviously obtain

$$\sum_{i=1}^n \gamma_i \geq 13\,350n.$$

Also, by (5) we know that

$$\sum_{i=1}^n d_i < 100n.$$

Therefore, from this and (9) one immediately obtains

$$\begin{aligned}|\Delta(\boldsymbol{\sigma},\boldsymbol{\gamma})| &\leq \frac{1}{0.67}\cdot\frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n \gamma_i}\\ &< \frac{1}{0.67}\cdot\frac{100n}{13\,350n}=\frac{200}{17\,889}\\ &< \frac{1}{89}.\end{aligned}$$

The proof is complete. □

4 Concluding remarks

In our paper, we have presented a simple mathematical method that definitely enables us to find a general upper bound for the quantity $|\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma})|$ that is often discussed when using the concept of the super-gross wage (Theorem 1). On the other hand, Corollary 1 provides even an absolute bound that is small enough for ensuring that the difference between the Gini coefficients $G(\gamma_1, \dots, \gamma_n)$ and $G(\sigma_1, \dots, \sigma_n)$ is insignificant. Accordingly, both these coefficients can be used for calculating the Musgrave-Thin index M of the global tax progression without any considerable impact on its value.

One easily inspects that Theorem 1 and Corollary 1 can both be generalized. The proofs of these statements remain analogous when considering an arbitrary tax rate r instead of the value 1.34 or other minimum gross wage amount m. In the case of such changes, the estimate in Corollary 1 has the form

$$|\Delta(\boldsymbol{\sigma}, \boldsymbol{\gamma})| < \frac{200}{r \cdot m}.$$

We leave the simple details to interested readers.

Acknowledgment

The author wish to thank the unknown referee and prof. J. Šustek for helpful comments that improved the presentation of this paper.

References

[1] P.D. Allison: Measures of inequality. *American Sociological Review* 43 (1978) 865–880.

[2] A.B. Atkinson, F. Bourguignon: *Handbook of Income Distribution*. New York: Elsevier (2015).

[3] L. Ceriani, P. Verme: The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality* 10 (3) (2012) 1–23.

[4] M. Genčev, D. Musilová, J. Široký: A Mathematical Model of the Gini Coefficient and Evaluation of the Redistribution Function of the Tax System in the Czech Republic. *Politická ekonomie* 66 (6) (2018) 732–750. (in Czech)

[5] C. Gini: *Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche*. Bologna: C. Cuppini (1912).

[6] P.J. Lambert: *The Distribution and Redistribution of Income*. Manchester: Manchester University Press (2002).

[7] R.A. Musgrave, T. Thin: Income tax progression, 1929–48. *Journal of Political Economy* 56 (1948) 498–514.

[8] L. Plata-Peréz, J. Sánchez-Peréz, F. Sánchez-Sánchez: An elementary characterization of the Gini index. *Mathematical Social Sciences* 74 (2015) 79–83.

[9] A.K. Sen: *On Economic Inequality*. Oxford: Oxford University Press (1997).

[10] M. Zenga, M. Polisicchio, F. Greselin: The variance of Gini's mean difference and its estimators. *Statistica* 64 (3) (2004) 455–475.

Received: 21 March 2018
Accepted for publication: 11 August 2019
Communicated by: Karl Dilcher