UNIVERSITÀ
DEGLI STUDI
DI TORINO

ALMA UNIVERSITAS
TAURINENSIS

Founded in 1404

DEPARTMENT OF
ECONOMICS AND STATISTICS
WORKING PAPER SERIES

Quaderni del Dipartimento di Scienze
Economico-Sociali e Matematico-Statistiche

ISSN 2279-7114

THE GINI COEFFICIENT: ITS ORIGINS

SIMONE PELLEGRINO

Working paper No. 70 - December 2020

The Gini Coefficient: Its Origins

Simone Pellegrino*

December 22, 2020

Abstract

This essay retraces the fundamental steps and analyses the theoretical motivation that influenced the definition of the Gini index and its application today. It starts with the concept of mean difference, proposed by Corrado Gini in 1912, for applications in statistics and economics. The difference between the concentration ratio Gini proposed in 1914 and the Gini index, as it is usually used today, is highlighted in light of its geometrical interpretation with the Lorenz piecewise linear function proposed by Gaetano Pietra in 1915.

JEL-Codes: D63.
Keywords : Gini Coefficient, Lorenz Curve.

*Department of Economics and Statistics – ESOMAS, Università degli Studi di Torino. ORCID: 0000-0001-8372-1054. Phone: 0039 011 670 6060. Email: simone.pellegrino@unito.it. Address: Corso Unione Sovietica 218bis, 10134, Turin (IT).

1. Introduction

For more than a century, the Gini index has been the most studied index and used to assess income and wealth inequality. Its range of applications is very wide and endless literature has followed over the decades (Giorgi, 1990; Xu, 2003).

Curiously, the bibliographies of many international scientific articles do not contain references to the original articles, even those immediately following Corrado Gini's publications. This essay reconstructs the historical origins of the Gini index by considering two fundamental works. The first is the pioneering work of Corrado Gini in 1914, while the second is a note published by Gaetano Pietra the following year. These two articles help us understand the steps and theoretical motivations that led to the definition of the Gini index, G, and its application today. In particular, Gini (1914) proposes the concentration ratio, R, and aims to demonstrate its connection to the Lorenz curve, a link subsequently investigated by Gaetano Pietra (1915).

Very briefly, this essay covers the following points:

In 1912, Corrado Gini proposed the concept of simple mean difference (with and without repetition) as an index of variability for quantitative values, which soon became a fundamental indicator for studies in statistics and economics. The objective of his work was to highlight[1] '*how the procedures followed thus far, to measure the variability of statistical phenomena . . . do not always respond well to the scope of the statistical investigation.* Gini discusses the application of the simple mean difference of observed quantities as an indicator that may be preferred over others in some areas of study.[2]

In 1914, in his publication '*Sulla misura della concentrazione e della variabilità dei caratteri*' ['*On the measurement of concentration and variability of characters*'], Gini presents three fundamental aspects that would later revolutionise the study of income and wealth inequality, as well as other areas.[3]

First, he proposes the concentration ratio (pp. 1203-1228), which would result in the Gini index, as it is applied today. This ratio presents some particular aspects. The denominator contains the sum of the cumulative portions of statistical units, while the numerator contains the sum of the differences between the cumulative portions of statistical units and the cumulative portions of the quantitative variable whose concentration is being calculated. This ratio can be represented graphically, as a ratio of the sums of segments. The concentration ratio, as originally conceived by Gini, is equal to zero for a perfectly egalitarian distribution of values and equal to one for maximum concentration.

Secondly (pp. 1229-1236), he devotes a few pages to observe that[4] '*The ratio, that we are proposing in this note as the appropriate measure of concentration, can also be obtained by improving a graphical method already introduced by some authors, as Lorenz (1905), Chatelain (1910), Séailles (1910) in order to evaluate inequality in the distribution of wealth.*' In reality, as highlighted in this essay, Gini does not propose a rigorous comparison between the two approaches but limits himself to giving some insight about it. He notes that the two methods (the concentration ratio and the ratio of the area between the equal distribution line and the Lorenz curve over the maximum area) yield the same result when the number N of quantities that measure the intensity of a certain value becomes very large.

Finally, in the last part (pp. 1236-1240), Gini verifies that[5] '*the concentration ratio coincides with*

[1] '*come i procedimenti finora seguiti per misurare la variabilità dei fenomeni statistici . . . non rispondano sempre bene allo scopo che l'indagine statistica si propone.*'

[2] Prior to this, Gini (1910) had already begun to examine concentration indices (Forcina and Giorgi, 2005).

[3] Non-Italian readers may read the English translation of this article, which was published in 2005 in Metron (Gini, 2005).

[4] '*Al rapporto, che noi proponiamo in questa nota, come misura appropriata della concentrazione, si giunge anche perfezionando un metodo grafico che alcuni autori, il Lorenz (1905), il Chatelain (1910), il Séailles (1910), hanno già proposto per giudicare della maggiore o minore disuguaglianza di distribuzione della ricchezza.*'

[5] '*il rapporto di concentrazione coincide col rapporto della differenza media [senza ripetizione] al valore massimo che questa può assumere, o in altre parole, col rapporto della differenza media al doppio della media aritmetica del carattere.*'

the ratio between the mean difference [without repetition] and its maximum value, or, in other words, with the ratio of the mean difference with twice the arithmetic mean of the character.'

The following year, in 1915, Gaetano Pietra[6] studied the link between the concentration ratio proposed by Gini and the ratio between the area of observed concentration and the area of maximum concentration, providing an elegant geometrical interpretation. This version of the index — given by the ratio between the mean difference with repetition in the observed series and the mean difference without repetition in the corresponding maximising series — has been the most-used index of inequality for more than a century. The same article contains the first definition of the Gini index for the continuous case[7] and the introduction of the concept of 'graduation', that is, the inverse of the distribution function.

In fact, the formula for the Gini index commonly applied today does not vary between zero and one, but between zero and $\frac{N-1}{N}$, as defined by Pietra (1915). The scientific community[8] prefers the latter version of the Gini index for three main reasons. First, in part, because Corrado Gini's work was not initially widely disseminated at the international level[9] because it was written in Italian.[10] Second, because the concentration ratio introduced by Gini in 1914 did not perfectly correspond to the graphical view of inequality in an ordered series of values, as proposed by Otto Max Lorenz in 1905 (Lorenz, 1905; Gini, 1914; Pietra, 1915). Finally, because the international literature focused on the study of continuous rather than discrete income and wealth distributions.

The essay is organised as follows: Subsection 2.1 introduces the concept of mean absolute difference, while Subsection 2.2 highlights some particular aspects of its historical origins. Section 3 presents Gini's concentration ratio, while Section 4 analyses the connection between the concentration ratio and the mean absolute difference. Following this, Section 5 summarises the original concepts of Otto Max Lorenz's curve. Section 6 describes the Lorenz curve in light of Corrado Gini's interpretation, while Section 7 examines the Lorenz curve in light of Gaetano Pietra's interpretation. Section 8 provides a conclusion.

2. The simple mean difference

2.1. The proposal in Corrado Gini's writings

Gini (1912) sets out to '*find a formula that expresses the arithmetic average of the differences among*' N '*quantities*'. In order to reach this goal, he considers a non decreasing series of non negative quantities

[6] Non-Italian readers can read the translation of this article in English, as published in 2013, in Statistica & Applicazioni (Pietra, 2013).

[7] More than 60 years later, Dorfman (1979) would propose an approach to unite the calculation of the continuous and discrete Gini indices.

[8] In his book '*Il rapporto di concentrazione di Gini*' ['*Gini's concentration ratio*'], Giorgi (1992) dedicates an entire chapter to the literature, underlining how hundreds of contributions on the topic have been written since 1914, most of them starting in the 1970s. For an exhaustive bibliography of all of Corrado Gini's writings (827 publications), see Castellano (1965). For an account of his personality, see Giorgi (2011).

[9] For the years immediately following its introduction, see Castellano (1965).

[10] Gini (1921) himself replies to the article '*Measurement of the Inequality of Income*' by Dalton (1920), thanking him for having introduced the writings of Italian statisticians to international economists and suggesting a more in-depth interpretation of some writings, in particular, those by Czuber (1914), Gini (1914), and Pietra (1915). In Gini's words: '*The methods of Italian writers, which are explained by Mr. Dalton, are not, as a matter of fact, comparable to his own, inasmuch as their purpose is to estimate, not the inequality of economic welfare, but the inequality of incomes and wealth, independently of all hypotheses as to the functional relations between these quantities and economic welfare or as to the additive character of the economic welfare of individuals.*' And then: '*Mr. Dalton explains these methods with precision and brevity, and Italian writers must be most grateful to him for having directed the attention of English economists to the subject. Perhaps, however – as a supplement to Mr. Dalton's article – I may be permitted to draw the attention of readers of the Economic Journal to certain papers, a perusal of which, in my opinion, is necessary to enable one to form an exact idea of the applicability and character of the methods in question. . . . Probably these papers have escaped Mr. Dalton's attention owing to the difficulty of access to the publications in which they appeared.*'

$x_1, x_2, \ldots, x_{N-1}, x_N$, with $x_{i-1} \leq x_i \ \forall i$. He observes that the sum of the $N-1$ possible differences between x_1 and all the other quantities is

$$\begin{aligned}
&(x_2 - x_1) + (x_3 - x_1) + \cdots + (x_{N-1} - x_1) + (x_N - x_1) = \\
&(x_2 + x_3 + \cdots + x_{N-1} + x_N) - (N-1)x_1 = \\
&x_1 + x_2 + x_3 + \cdots + x_{N-1} + x_N - N x_1.
\end{aligned} \tag{1}$$

Similarly to Eq. (1), for x_2 he gets

$$\begin{aligned}
&(x_2 - x_1) + (x_3 - x_2) + \cdots + (x_{N-1} - x_2) + (x_N - x_2) = \\
&x_3 + \cdots + x_{N-1} + x_N - (N-2)x_2 + (x_1 - x_2) + (2x_2 - 2x_1) = \\
&x_2 + x_3 + \cdots + x_{N-1} + x_N - (N-1)x_2 + 2x_2 - x_1 - x_2
\end{aligned} \tag{2}$$

and, for x_3,

$$x_3 + \cdots + x_{N-1} + x_N - (N-2)x_3 + 3x_3 - x_1 - x_2 - x_3. \tag{3}$$

And so on up to the last value of the series of N quantities:

$$x_N - x_N + N x_N - x_1 - x_2 - \cdots - x_{N-2} - x_{N-1} - x_N. \tag{4}$$

Adding up all the N equations and rearranging them, Gini (1912) gets[11] a first formulation able to express the arithmetic mean of the $N(N-1)$ possible differences between the N quantities, i.e. the simple mean difference without repetition Δ:

$$\Delta = \frac{2}{N(N-1)} \sum_{i=1}^{\frac{N+1}{2}} (N+1-2i)(x_{N-i+1} - x_i). \tag{5}$$

The arithmetic mean of the N^2 possible differences between the N quantities, i.e. the simple mean difference with repetition Δ_R can be instead written as

$$\Delta_R = \frac{2}{N^2} \sum_{i=1}^{\frac{N+1}{2}} (N+1-2i)(x_{N-i+1} - x_i) \tag{6}$$

from which

$$\Delta_R = \frac{N-1}{N} \Delta \tag{7}$$

is derived. It has to be noted that the original formulas, that is Eq. (5) and Eq. (6), are now in disuse; most frequently, labeling Y the sum of differences in absolute value (De Finetti and Paciello, 1930; De

[11] In the book by Gini (1912) dozens of alternative formulas for the computation of the mean difference are then discussed (for a collection, see Ceriani and Verme (2012) and Yitzhaki and Schechtman (2013)). In the immediately following years the debate of the Italian school of statistics on the simple mean difference and the concentration ratio is born (Bresciani-Turroni, 1916; Ricci, 1916; Pietra, 1917, 1932; Yntema, 1933; Pietra, 1935; Castellano, 1935, 1937; Pietra, 1937) and some authors try their hand at identifying more efficient and faster formulas for calculating both the simple mean difference (De Finetti and Paciello, 1930; De Finetti, 1931), and the concentration ratio (de Vergottini, 1940; Amato, 1947; de Vergottini, 1950; Pizzetti, 1955; Fortunati, 1955, 1957; Benedetti, 1980). The last cited essay contains a peculiarity: Giorgi (1990) writes '*In more recent years, the aforementioned theme is analysed by Benedetti (1980) which brings to the attention of scholars a general formula he deduced in the early 1950s but not immediately published due to hostility, as claimed by the author himself, of Gini towards everything that tends to diminish his concentration ratio making it seem an index in the same way as many others.*' *['In anni più prossimi a noi la suddetta tematica è ripresa da Benedetti (1980) che pone all'attenzione degli studiosi una formula generale da lui desunta agli inizi degli anni '50 ma non pubblicata subito per l'ostilità, come sostiene lo stesso Autore, di Gini verso tutto ciò che tende a sminuire il suo rapporto di concentrazione facendolo sembrare un indice alla stessa stregua di tanti altri.']* Hence the importance, even today, of studying the original writings of Italian authors (Giorgi, 1990, 2005, 2014).

Finetti, 1931)

$$Y = \sum_{i=1}^{N}\sum_{j=1}^{N} |x_i - x_j| \tag{8}$$

it is very common to make them explicit as

$$\Delta = \frac{Y}{N(N-1)} \tag{9}$$

and

$$\Delta_R = \frac{Y}{N^2}. \tag{10}$$

As Gini (1912) observes,[12] '*...the probable deviation is given by a quantity which is greater in absolute value by one half of the deviations and not exceeded by the other half*'; therefore for the calculation of Y it is sufficient to consider the differences above (or below) the main diagonal, because the difference matrix is symmetrical. As a consequence,

$$Y = \sum_{i=1}^{N}\sum_{j=1}^{N} |x_i - x_j| = 2\sum_{i=1}^{N}\sum_{j=1}^{i} (x_i - x_j). \tag{11}$$

2.2. Historical origins

The historical origins and effective paternity of the formula for the mean absolute difference merits some further consideration. In fact, the work of some astronomers in the second half of the nineteenth century already contained this concept (Jordan, 1869; von Andrae, 1869; Jordan, 1872; von Andrae, 1872; Helmert, 1876); however, Corrado Gini independently proposed the definition of mean absolute difference, emphasising that he only became aware of the articles by German astronomers when his book was practically finished.

In particular, in the section '*La differenza media tra più quantità*' ['The mean difference between multiple quantities'], Gini (1912, pp. 20-23) derives Δ (Formula 5 on p. 22),[13] that is, the mean absolute difference without repetition, and Δ_R, that is, the mean absolute difference with repetition (Formula 7).[14] Later, on p.49, in the section '*Degli indici di variabilità dei caratteri in alcuni tipi di seriazioni*' ['Indices of variability of values in some types of seriation'], note 2 on p. 56, discusses Formulas 76 and 77, which define Δ_R and Δ, respectively, in an alternative way. Gini observes,[15] '*This formula, and therefore, also Formula 77, which derives from it, are for now empirical formulas. In fact, as of now, I am not able to provide a general demonstration, not even through mathematical induction. The numerous comparisons I have made and the fact that ... these formulas reduce to expressions that were already obtained in another way by von Andrae and Helmert make their mathematical exactness very probable.*'

In note 1 on p. 58 he further specifies,[16] '*This study was already complete when I learnt about the research by W. Jordan, von Andrae, and Helmert, who, many years ago, were calculating the mean*

[12] '*... lo scostamento probabile è dato da una quantità che è superiore in valore assoluto da una metà degli scostamenti e non superata dall'altra metà*'

[13] To avoid confusion, the formulas discussed in this essay are given in parentheses while the formulas in the original text are reported without parentheses.

[14] In the same section, Gini also introduces the concept of gradual distance, what we today call the *rank* of an income distribution.

[15] '*Questa formula, e quindi anche la 77 che ne discende, sono dunque, per ora, formule empiriche. Non mi riuscì infatti finora di darne una dimostrazione generale, neppure mediante induzione matematica. I numerosi riscontri che ne ho eseguito e il fatto che ... tali formule si riducono ad espressioni a cui erano già giunti per altra via von Andrae ed Helmert, fanno ritenere molto probabile la loro esattezza matematica.*'

[16] '*Questo studio era completamente scritto quando ho potuto conoscere alcune ricerche di W. Jordan, von Andrae e Helmert, che, molti anni or sono, si sono occupati, da un punto di vista del tutto diverso, del calcolo della differenza media tra più quantità.*'

difference between multiple quantities, from an entirely different perspective.'[17] He then indicates the field of investigation of the German astronomers and underlines that these articles contain some of the formulas that he came by independently.[18]

The same chronicle is found in note 2 on p. 77 of De Finetti (1931) in Metron, which at the time, was directed by Corrado Gini. In particular, De Finetti observes that the German astronomers were occupied by the simple mean difference[19] '*for the question of calculating probabilities relating to the theory of observational errors.*' The author concludes the discussion by stating,[20] '*As for the scope of this work mentioned above, it is clear that it lies rather far from the statistical study of variability, for which, as was stated, Gini is therefore given the merit of having introduced the mean difference as a useful index.*'

3. The concentration ratio

In the 1914 publication Corrado Gini immediately proposes the concentration ratio. He considers N '*quantities that measure the intensity of a certain character in*' N '*different cases*', $x_1, x_2, \ldots, x_i, \ldots, x_l, \ldots, x_{N-1}, x_N$, ordered in non decreasing order, so that $x_{k-1} \leq x_k \ \forall k$, with $k = 1, 2, \ldots, N$. He then observes that, by considering two values, i and l, of k, with $i < l$, $x_i \leq x_l$ is obtained and also

$$\frac{\sum_{k=1}^{i} x_k}{\sum_{k=1}^{l} x_k} \leq \frac{i}{l}. \tag{12}$$

[17]'*Scopo delle loro indagini era, non già di esaminare se la misura della variabilità, eseguita in base alla differenza media, può condurre a risultati diversi da quelli ottenuti in base allo scostamento quadratico medio o allo scostamento semplice medio, e di decidere in quali casi la misura appropriata della variabilità dei fenomeni è fornita dall'una, in quali dalle altre costanti; ma di esaminare, nel caso particolare in cui le quantità osservate sono il risultato di rilevazioni ugualmente plausibili di una grandezza incognita, se lo scostamento probabile, determinato indirettamente mediante la differenza media fra le quantità osservate, risente l'influenza del numero limitato delle osservazioni più o meno dello scostamento probabile determinato indirettamente mediante lo scostamento quadratico medio, e di decidere quindi se, per caratterizzare la precisione delle rilevazioni, è preferibile attenersi all'uno o all'altro procedimento.*' ['Their investigations were not to examine if the variability measured using the mean difference would lead to different results from those obtained using the standard deviation or mean deviation or decide in which cases the appropriate measure of variability of a phenomenon is provided by one or the other. Rather, it was to examine, for the particular case in which the observed quantities are the result of equally plausible detections of an unknown quantity, if the probable deviation determined indirectly through the mean difference of the observed quantities is influenced by the more or less limited number of observations of the probable deviation determined indirectly through the standard deviation. Therefore, it was to decide if it is preferable to use one procedure or the other to characterise the precision of the detections.

[18]'*Il Jordan aveva ritenuto che, col crescere del numero*' N '*delle osservazioni, il valore della differenza media tendesse al suo limite per*' N '*infinito più rapidamente che il valore dello scostamento quadratico medio; il von Andrae invece dimostrò che la rapidità è, per il valore della differenza media, minore che per lo scostamento quadratico medio, ma maggiore che per lo scostamento semplice medio. Credo doveroso avvertire che in questi articoli si trova già qualcuna delle formule, a cui, del tutto indipendentemente e per vie differenti, io sono giunto in questo studio. Il von Andrae perviene, con una dimostrazione diversa della mia, ad una formula equivalente alla 5 per la determinazione della differenza media fra più quantità e dimostra pure la relazione 31 fra differenza quadratica media e scostamento quadratico medio. Jordan perviene, in base a una dimostrazione non rigorosa, alla relazione 80 fra differenza media e scostamento quadratico medio, nell'ipotesi che le quantità tendano a disporsi secondo la legge di Gauss; Andrae dà la dimostrazione rigorosa di tale relazione ed Helmert la dimostrazione rigorosa della 79.*' [Jordan had maintained that by increasing the number N of observations, the value of the mean difference tends to its limit for infinite N more quickly than the value of standard deviation. Von Andrae instead showed that for the value of mean difference, this occurs more slowly than the standard deviation, but more quickly than the mean absolute deviation. I feel it necessary to warn [the reader] that these articles already contain some of the formulas that I found entirely independently and via different routes in this study. With a different demonstration, von Andrae arrives at a formula equivalent to 5 to determine the mean difference between multiple quantities, and shows relationship 31 between the root mean square deviation and the standard deviation. Based on a non-rigorous demonstration, Jordan arrives at relationship 80 between the mean difference and standard deviation under the assumption that the quantities tend to be arranged according to Gauss' law. Andrae provides the rigorous demonstration of this relationship and Helmert provides the rigorous demonstration of 79.']

[19]'*a proposito di una questione di calcolo delle probabilità relativa alla teoria degli errori di osservazione.*'

[20]'*Quanto allo scopo, già accennato, di tali lavori, si comprende facilmente che esso è ben lontano da quello dello studio statistico della variabilità, nel quale dunque, come s'è asserito, spetta al Gini il merito d'aver introdotto la differenza media come un utile indice.*'

For the special case $l = N$

$$\frac{\sum_{k=1}^{i} x_k}{\sum_{k=1}^{N} x_k} \leq \frac{i}{N} \tag{13}$$

is obtained. At this point Gini defines P_i as the ratio between the rank of the i-th quantity and the overall number of observed cases

$$P_i = \frac{i}{N} \tag{14}$$

and L_i the ratio between the amount of character accruing to the portion of cases occupying a position equal to or less than the i-th position and the total amount of the observed character:

$$L_i = \frac{\sum_{k=1}^{i} x_k}{\sum_{k=1}^{N} x_k}. \tag{15}$$

Lastly, he underlines '*We say that the stricter the inequality*' $P_i > L_i$ '*for the*' $N-1$ '*values of i, the stronger the concentration of the character.*'

Gini (1914) shows the concetration ratio R as follows:

$$R = \frac{\sum_{i=1}^{N-1}(P_i - L_i)}{\sum_{i=1}^{N-1} P_i} = 1 - \frac{\sum_{i=1}^{N-1} L_i}{\sum_{i=1}^{N-1} P_i} \tag{16}$$

specifying '*the smaller the part of the total amount of the character owned by those cases whose intensity of the character itself is below a certain level, the stronger the concentration of the character.*'

Even if every possibile series gets N values, the numerator of Eq. (16) can be suitably computed up to $N-1$, since $P_N = 1$, $L_N = 1$, and $P_N - L_N = 0$. Similarly, even if the denominator of R can be computed up to $N-1$, since $P_N = 1$ and $L_N = 1$. The denominator $\sum_{i=1}^{N-1} P_i$ evaluates the summation of differences in the case of maximum concentration, that is a situation in which $N-1$ statistical units get a nil share of the overall sum of the character and only one statistical unit gets the overall amount of the character. The numerator $\sum_{i=1}^{N-1}(P_i - L_i)$ evaluates the summation of differences in case of the observed concentration. The concentration ratio R is then equal to zero whenever quantities are equally distributed and is exactly equal to 1 whenever $N-1$ quantities are equal to zero and only one is positive.

Table 1 shows the steps for the calculation of R for an ordered series of non-negative quantities $(3, 4, 5, 6, 7, 8, 9)$. In particular, the numerator of Eq. (16) is equal to $\sum_{i=1}^{N-1}(P_i - L_i) = \frac{2}{3}$, while the denominator to $\sum_{i=1}^{N-1} P_i = 3$. As a consequence, $R = \frac{2}{9} = 0.\bar{2}$.

Table 1: The calculation of the concentration ratio R

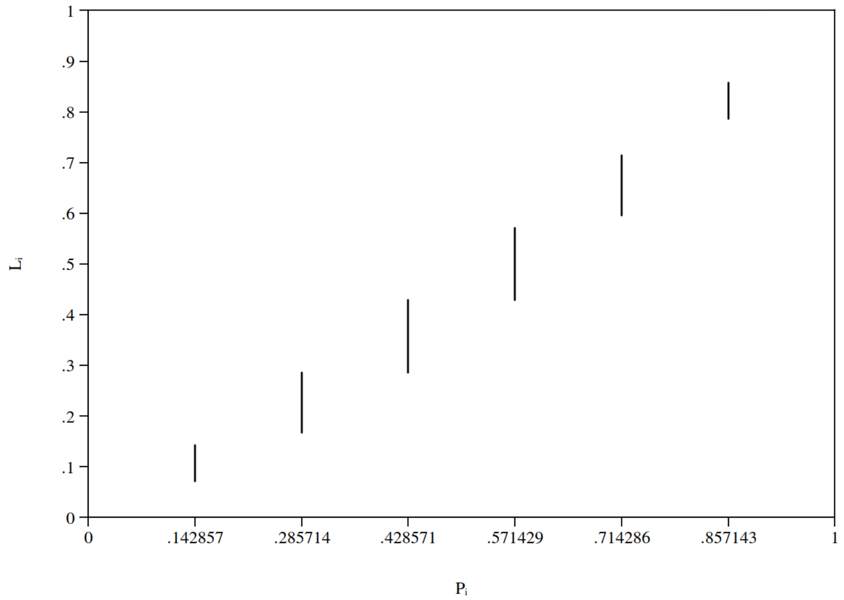
i	x_i	P_i	L_i	$P_i - L_i$
1	3	0.142857	0.071429	0.071429
2	4	0.285714	0.166667	0.119048
3	5	0.428571	0.285714	0.142857
4	6	0.571429	0.428571	0.142857
5	7	0.714286	0.595238	0.119048
6	8	0.857143	0.785714	0.071429
7	9	1.000000	1.000000	0.000000

Source: Own elaborations.

Considering the same series, Figure 1 offers a graphic vision in which every difference $P_i - L_i$ of the summation in the numerator of Eq. (16) is represented with a segment. Considering the i-th segment, its upper-extreme point is P_i, while the lower one is L_i. Segments considered in the denominator of

Eq. (16) are the same segments of Figure 1, but longer, up to a zero value of the ordinate. Note that Eq. (16) is then defined as a ratio between sums of segments. Segments $P_i - L_i$ are, for every i, the distance between the situation that would be observed with perfect equality, that is P_i, and the observed situation, that is L_i. Figure1 is not mentioned in the original writings by Gini, but it is very useful in order to study in deep the reasoning of Gini in proposing the comparison between his concentration ratio and the Lorenz curve.

Figure 1: Segments to the numerator of concetration ratio R



4. The concentration ratio and the mean difference

Gini (1914, pag. 1236-1238) studies '*the relationship between the concentration ratio and the indices of variability, used to characterize the distribution of the variables investigated.*' In particular, he demonstrates '*that the concentration ratio coincides with the ratio between the mean difference and its maximum value, or in other words, with the ratio of the mean difference with twice the arithmetic mean of the character.*' In math terms, Gini, focusing on the mean difference without repetition, Δ (as defined by Eq. (5)), verifies that

$$R = \frac{\Delta}{2\mu} = \frac{\sum_{i=1}^{\frac{N+1}{2}} (N+1-2i)(x_{N-i+1} - x_i)}{(N-1)\sum_{i=1}^{N} x_i} \tag{17}$$

where $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $2\mu = \Delta^{MAX}$ shows the mean difference without repetition of the maximizing series, that is a series in which every quantities are equal to zero but one, to which are transferred all the quantities of remaining $N-1$ quantities.[21]

More precisely, by replacing Eq. (14) and Eq. (15) into Eq. (16), Gini (1914, pag. 1208) obtains

$$R = 1 - \frac{2}{(N-1)} \frac{\sum_{i=1}^{N-1}(N-i)x_i}{\sum_{i=1}^{N} x_i} \tag{18}$$

[21]It can be also showns that $R = \frac{\Delta}{\Delta^{MAX}} = \frac{\Delta_R}{\Delta_R^{MAX}}$, where Δ_R^{MAX}, similarly shows the mean difference with repetition of the maximising series.

that he rewrites as[22]

$$R = \frac{(N-1)\sum_{i=1}^{N} x_i - 2\sum_{i=1}^{N-1}(N-i)x_i}{(N-1)\sum_{i=1}^{N} x_i}. \tag{19}$$

To proof the equality between R and $\frac{\Delta}{2\mu}$, Gini observes that it is sufficient to demonstrate that the numerator of Eq. (17) and the numerator of Eq. (19) are equal, since the corresponding denominators are equal. This proof is promptly provided (Gini, 1914, pag. 1238).

5. The Lorenz curve and piecewise linear function

Otto Max Lorenz (1905) devised[23] a graphical view[24] of inequality, proposing a comparison between the cumulative portion, L_i, of a quantitative variable[25] and the cumulative portion of frequencies P_i, having ordered these frequencies from poorest to richest.[26] Although from a practical perspective, he underlines that information about income and wealth is often available only for aggregate data, he assumes point-like data on income or wealth of a population and presents the first two Lorenz curves for two specific cases. The first refers to Prussian incomes in 1892 and 1901, where the two curves do not intersect, highlighting a greater concentration of incomes in 1901 with respect to those in 1892. The second refers to a theoretical example of a distribution of 10 incomes that instead determine two intersecting curves (he states that in this situation, some conclusions can also be drawn from the variation of the observed inequality).

Even if the number of quantities N considered in the distributions applied in the original essay by Lorenz is relatively small (10 incomes in the teoretical example), he graphically represents his examples by employing curves, therefore reasoning in the continuous case. Also Gini (1914) always represents concentration curves (in his original essay synonymous with Lorenz curve). Moreover, Gini observes that, when N is large enough '*If in a Cartesian diagram, we report the values*' P_i '*on the abscissa and the values*' L_i '*on the ordinate and we connect the points*' (P_i, L_i), '*the resulting curve is the concentration curve, which is increasing and convex.*' The concentration curve tends to be more convex the larger the inequality in the distribution, while it flattens with less inequality.

However, concluding his observations on the relationships between the Lorenz curve and its concentration ratio R, the author examines the Lorenz curve for values grouped into classes. Here, he presents a 'piecewise linear function' composed of as many segments as there are classes and he tries to approximate the area enclosed by the Lorenz curve by calculating the corresponding area enclosed in what today we indicate with the Lorenz 'piecewise linear function'. Finally, Gini observes that '*the*

[22]Note that $\sum_{i=1}^{N} i = \frac{N(N+1)}{2}$; as a consequence, $\sum_{i=1}^{N-1} i = \frac{N(N-1)}{2}$ and $\sum_{i=1}^{N-1} P_i = \sum_{i=1}^{N-1} \frac{i}{N} = \frac{1}{N}\sum_{i=1}^{N-1} i = \frac{N(N-1)}{2N} = \frac{N-1}{2}$. On the basis of Eq. (16), Eq. (18) can be rewritten as

$$R = 1 - \frac{2}{N-1}\sum_{i=1}^{N-1} L_i.$$

Similarly , since $\sum_{i=1}^{N-1} P_i = \frac{N-1}{2}$, from Eq. (16) we get

$$R = 2\frac{\sum_{i=1}^{N-1}(P_i - L_i)}{N-1}.$$

The last equation underlines that the concentration ratio R is equal to the double of the aritmetic mean of the $N-1$ differences between P_i and L_i, that is the double of the aritmetic mean of cumulated percentage shares that should be added to every cumulated intensity in order to obtain perfect equidistribution.

[23]In the first part of his article, Lorenz criticises the methods used up to then to assess inequality, which usually consisted of a class-based comparison between the amount of income or wealth and the corresponding shares of population.

[24]The cumulative portion of the quantitative variable is represented on the x-axis while the cumulative portion of the population is shown on the y-axis, in contrast to how it is usually plotted today.

[25]In his original article, Lorenz applies the reasoning indiscriminately to the distribution of income or wealth. See Schneider (2004) for a discussion of the historical origins of the Lorenz curve.

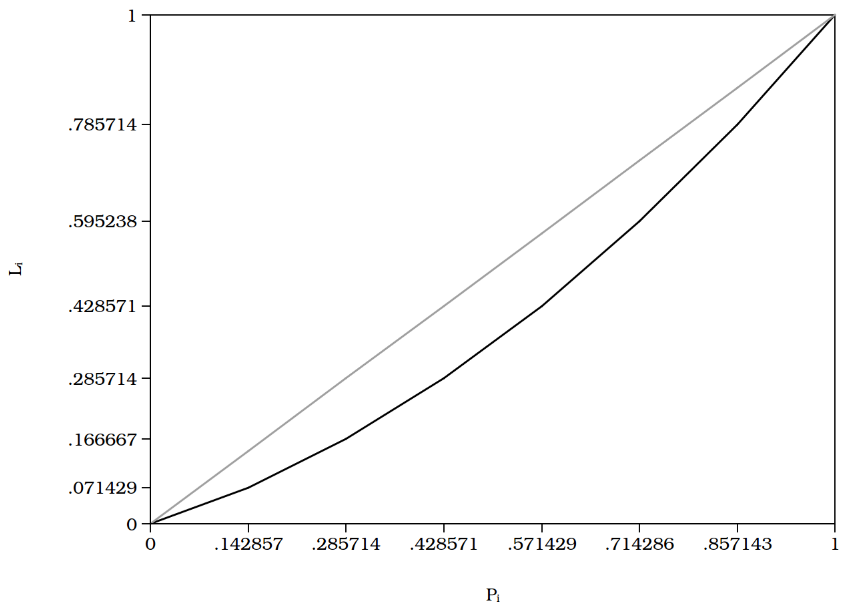
[26]The same year, Money (1905) discusses (third chapter) the inequality of income and wealth in the United Kingdom, using a stylised and non-rigorous approach similar to what was proposed by Otto Max Lorenz the same year.

piecewise line is inscribed inside the concentration curve', specifying that the difference between the concentration area delimited by the piecewise linear function and the concentration curve grows with increasing observed concentration.

In his original essay, immediately Pietra (1915) explicitly considers a piecewise linear function of concentration: '*In the general case in which the values*' x_i '*are not all equal, a polygonal chain is obtained by joining the points of coordinates*' P_i e L_i. '*The lowest point of such polygonal has coordinates*' $\left(P_1 = \frac{1}{N},\, L_1 = \frac{x_1}{\sum_{i=1}^{N} x_i}\right)$ '*while its highest point is in*' $\left(P_N = 1,\, L_N = 1\right)$.

Following the interpretation given by Pietra (1915), when N is not particularly large, the Lorenz curve may be approximated through the corresponding piecewise linear function. The same series discussed before $(3, 4, 5, 6, 7, 8, 9)$ is considered. The coordinates of the Lorenz piecewise linear function are the pairs (P_i, L_i) connected by lines. The black line in Figure 2 represents the Lorenz piecewise linear function, while the grey line is the Lorenz curve obtained when L_i is equal to P_i $\forall i$, that is, when all quantities are equal (line of equal distribution). The area under the equal distribution line and above the Lorenz piecewise linear function (indicated with A) gives a graphical view of how far a generic series is from a situation in which all values of the property are equal. B indicates the area under the Lorenz piecewise linear function. One can see that $A + B = \frac{1}{2}$. Hereafter, if not specified otherwise, reference is made to the Lorenz piecewise linear function (discrete case) and not to the corresponding curve (continuous case).

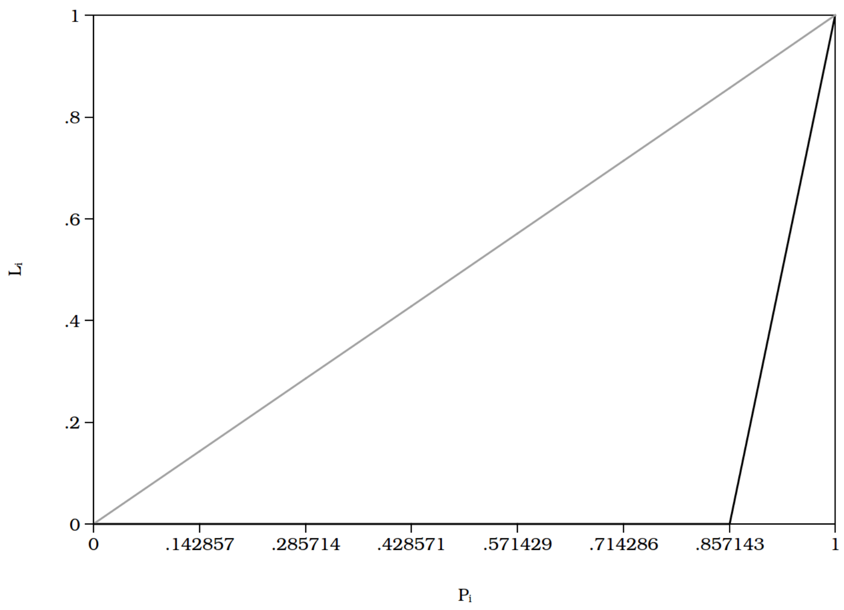
Figure 2: The Lorenz piecewise linear function $(N = 7)$



Both Lorenz and Gini discuss the case of maximum equality; neither directly addresses the situation of maximum inequality. We now pause on this extreme case, which is useful for understanding the remainder of the essay.

While the case of maximum equality presents no interpretative problems, since the area A is trivially equal to zero, so that the area B consequently is equal to $\frac{1}{2}$, the case of maximum inequality shows a peculiarity when quantities are not sufficiently numerous. In the discrete case, by considering a generic maximising series, $X^{MAX} = (0, 0, 0, 0, \ldots, 0, 0, x_N)$, where x_N is clearly positive, the maximum area A of the Lorenz piecewise linear function is not equal to $\frac{1}{2}$ but it is lower. Specularly, the minimum area B is not equal to zero, since it is positive. See Figure 3 for $N = 7$. It can be shown that $B = \frac{1}{2}\frac{1}{N}$, $A = \frac{1}{2} - B = \frac{1}{2} - \frac{1}{2}\frac{1}{N} = \frac{1}{2}\left(1 - \frac{1}{N}\right) = \frac{1}{2}\frac{N-1}{N}$; from which $A + B = \frac{1}{2}$.

Figure 3: The Lorenz piecewise linear function with maximum inequality ($N = 7$)



6. The Lorenz curve, concentration ratio, and Gini's interpretation

To confirm the validity of the concentration ratio R, Gini (1914, p. 1230-1231) underlines two problems with the graphical view of inequality proposed by Lorenz, as follows: if two or more distributions are compared and the corresponding concentration curves (or piecewise linear functions) intersect, it is not possible to order these concentrations.[27] Furthermore, the graphical representation in itself does not allow the dimension of the concentration observed in a distribution to be summarised or quantified as a single scalar. Gini goes on to state '*Both the above drawbacks disappear if, as measurement of the concentration, we consider the ratio between the area limited by the concentration curve and the egalitarian line (concentration area) and the ... concentration area in the case of maximum concentration.*' These areas are, respectively, the area A and the sum of the areas A and B defined in Section 5 in reference to the discrete case. Gini finds it convenient to assume the following ratio as a measure of concentration:

$$\frac{A}{A+B} \tag{20}$$

about which he states that '*It is now straightforward to show that this ratio is the limit the concentration ratio R tends to, when the number*' N '*of cases increases and the distribution of the character is unchanged.*'

This observation merits careful discussion. As observed in Section 3, the concentration ratio R proposed by Gini is, in fact, a ratio between the sums of segments. The segments in the numerator represent the degree of inequality observed (Figure 1). The upper extreme of these segments pertains to the equal distribution line and the lower extreme to the Lorenz piecewise linear function. The segments in the denominator represent the maximum observable inequality. In this case too, the upper extreme lies on the equal distribution line, while the lower extreme lies on the x-axis. Increasing N increases the number of segments, and if the number of observations increases indefinitely, they fall

[27] These limitations were the object of study decades later byAtkinson (1970), Shorrocks (1983), Shorrocks and Foster (1987), Kakwani (1984), and Dardanoni and Lambert (1988).

into the following two areas: those representing the observed inequality, enclosed between the diagonal equal distribution line and the Lorenz curve (area A), and the area representing maximum inequality (the sum of area A and area B, i.e., $\frac{1}{2}$).

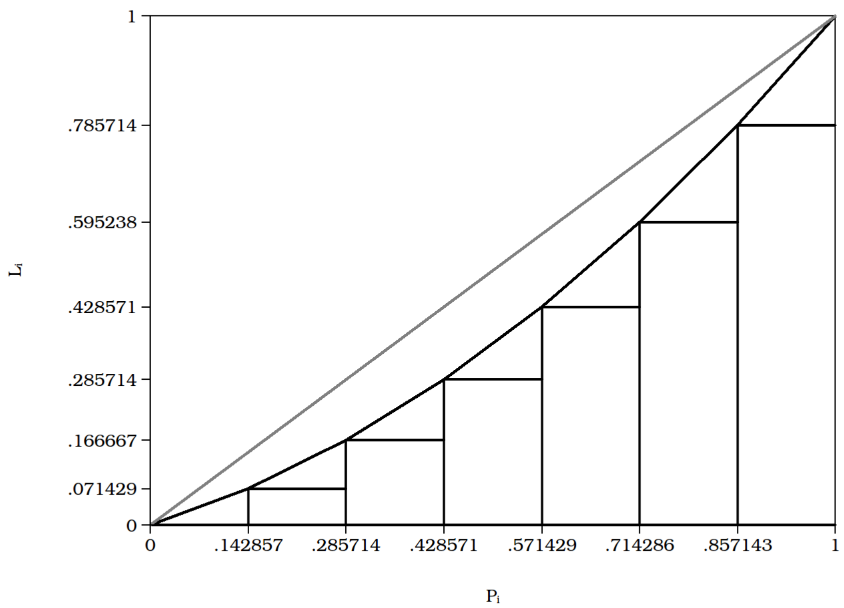
Gini does not provide a rigorous demonstration of his statement, rather, he supports its robustness with purely geometrical reasoning, which may be reinterpreted as follows: We consider Eq. (16). The original concentration ratio R, expressed as the ratio of sums of segments, may be appropriately defined as the ratio between sums of rectangular areas:

$$R^* = \frac{\sum_{i=1}^{N-1}(P_i - L_i)\frac{(i+1)-i}{N}}{\sum_{i=1}^{N-1} P_i \frac{(i+1)-i}{N}} = \frac{\frac{1}{N}\sum_{i=1}^{N-1}(P_i - L_i)}{\frac{1}{N}\sum_{i=1}^{N-1} P_i}. \tag{21}$$

Gini does not directly explain Eq. (21); however, it is useful for interpreting the graphical view he proposes about this point. As underlined, he considers a figure showing the concentration curve; here, it is preferable to present the argument considering the piecewise linear concentration function, as illustrated in Figure 4 for $N = 7$.

Specifically, Gini first considers $N - 1$ rectangles with base $\frac{1}{N}$ limited by the x-axis and the concentration piecewise linear function. The height of each of these rectangles is equal to L_i and the sum of their areas is equal to $\frac{1}{N}\sum_{i=1}^{N-1} L_i$.

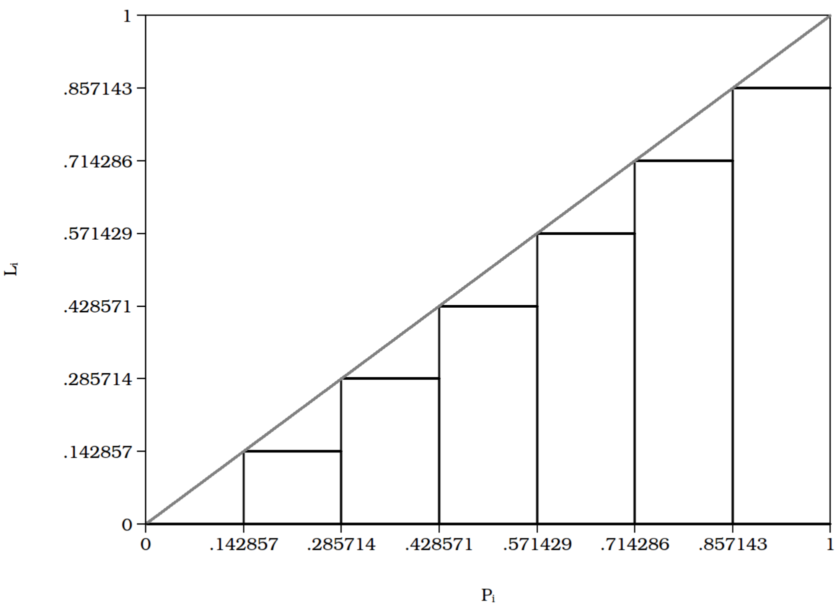
Figure 4: Rectangles by considering L_i



He then considers the $N-1$ rectangles limited by the x-axis and the equal distribution line (Figure 5). These rectangles always have a base equal to $\frac{1}{N}$ and a height equal to P_i. The sum of their areas is equal to $\frac{1}{N}\sum_{i=1}^{N-1} P_i$. The difference between the two sums, that is, $\frac{1}{N}\sum_{i=1}^{N-1}(P_i - L_i)$, is equal to the sum of the areas of rectangles partially inscribed in the concentration area.

It does not seem that Gini adds anything else that is relevant to this point. He limits himself to concluding that with increasing N, the two areas described with the sum of the rectangles tend to coincide with the observed concentration area and the maximum concentration area, respectively, and that their ratio coincides with the concentration ratio R. The sum of the areas of the triangles, that is, the area not calculated in the sum of the rectangles, grows progressively smaller with increasing N. When N becomes sufficiently large, a substantial correspondence is seen between the area calculated

Figure 5: Rectangles by considering P_i



with Eq. (16), or Eq. (21), and the effectively observed area.

To conclude this discussion, the following statements can be considered. In case of maximum inequality (see Figure 5), the area under the line of perfect equality, the maximum area B, which we label B^{MAX}, is equal to $\frac{1}{2}$. Label instead $\frac{1}{N}\sum_{i=1}^{N-1} P_i$ with B_R^{MAX}. B_R^{MAX} does not consider N triangles; the base of each triangle is $\frac{1}{N}$ and the height is similarly equal to $\frac{1}{N}$. By applying the Gini's approach, the total not computed area is then $\frac{N}{2}\left(\frac{1}{N}\right)^2 = \frac{1}{2N}$.

It follows that B_R^{MAX}, evaluated by Eq. (21), that is the sum of the areas of the $N-1$ rectangles, is equal to $\frac{1}{2} - \frac{1}{2}\frac{1}{N} = \frac{1}{2}\frac{N-1}{N} < \frac{1}{2}$. The area B^{MAX} is instead equal to $\frac{1}{2}$. In case of non-egalitarian distributions, the reasoning is the same (see Figure 4). When N increases, the sum of the areas of the rectangles tends to be equal with the concentration area, that is

$$\lim_{N\to+\infty} B_R^{MAX} = B^{MAX}. \tag{22}$$

When N increases, '*the areas of the small surfaces limited by the egalitarian line and the upper side of the rectangles of height*' P_i '*decreases and, analogously, the area of the small surfaces limited by the concentration curve and the upper sides of the rectangles of height*' L_i '*decreases as well (Gini, 1914).*' Gini's statements, according to which R tends to be equal to $\frac{A}{A+B}$ derives from these considerations.

7. The Lorenz piecewise linear function, Pietra's intuition, and the Gini index today

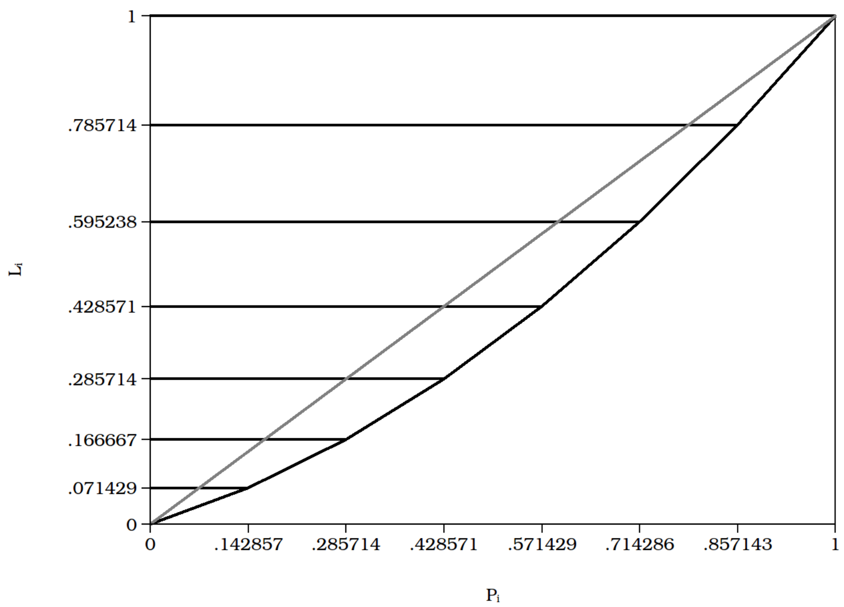
One year after the publication of the work by Gini (1914), Gaetano Pietra (1915), one of his students, published a note regarding the relationships between the indices of variability. The student's work certainly did not reach the notoriety of his master's; however, it contains results that were truly important for the future of descriptive statistics and its applications to the study of economic inequality.

In the first part of the note, Pietra (1915) considers Lorenz's piecewise linear function for a generic series, as described in reference to the comment to Figure 2, and proposes an elegant method, known today as the trapezoidal rule, to calculate the concentration area, A, in the discrete case. Having done so, he considers it natural to relate the area A to its maximum theoretical value, which is obtained in the continuous case, that is, the sum of areas A and B, which equals $\frac{1}{2}$. The strength of his intuition lies in the fact that he ties the ratio $\frac{A}{A+B}$ to the ratio of mean differences (the mean difference with repetition in the observed series over the mean difference without repetition in the maximising series), thereby, yielding an alternative formula to precisely calculate the concentration ratio in the discrete case. Therefore, the synthetic measure of inequality may be expressed precisely by $\frac{A}{A+B} = 2A$, which is equal to zero for maximum equality and equal to $\frac{N-1}{N}$ for maximum inequality (see the considerations in Section 5).

In the course of his demonstration, Pietra suggests the expression used today to calculate what we usually indicate as the Gini index, G, which differs from the concentration ratio R originally proposed by Gini, when the modalities of the characteristic are not sufficiently numerous. He also defines, for the first time, the concentration ratio in the continuous case, indicating that R is the limit that G tends to for very large N.

The author begins by representing Lorenz's piecewise linear function with a generic series and observes that the area A, that is, the area under the equal distribution line and above the piecewise linear concentration function, can be calculated as the excess with respect to $\frac{1}{2}$ of the sum of areas of a triangle and $N-1$ trapezoids, as illustrated in Figure 6.

Figure 6: The area above the Lorenz piecewise linear function



The sides of a triangle are P_1 (since $P_1 - P_0 = P_1$) and L_1 (since $L_1 - L_0 = L_1$); the overall length of the parallel sides of the i-th trapezoid is $P_i + P_{i-1}$, while the distance between them is $L_i - L_{i-1}$.

The area A can be then evaluated as

$$A = \sum_{i=1}^{N} \frac{(P_i + P_{i-1})(L_i - L_{i-1})}{2} - \frac{1}{2}. \tag{23}$$

From Eq. (23), Pietra (1915) observes that $P_i = \frac{i}{N}$ and $P_{i-1} = \frac{i-1}{N}$, from which $P_i + P_{i-1} = \frac{2i-1}{N}$. Moreover, he notes that $L_i - L_{i-1} = \frac{x_i}{W}$, where $W = \sum_{i=1}^{N} x_i = N\mu$, and $L_i - L_{i-1} = \frac{x_i}{\sum_{i=1}^{N} x_i}$. The area of the i-th trapezoid can be rewritten as

$$\frac{(P_i + P_{i-1})(L_i - L_{i-1})}{2} = \frac{1}{2}\frac{2i-1}{N}\frac{x_i}{\sum_{i=1}^{N} x_i} \tag{24}$$

so that he defines the area A as

$$A = \frac{1}{2N\sum_{i=1}^{N} x_i}\sum_{i=1}^{N}(2i-1)x_i - \frac{1}{2}. \tag{25}$$

At this point he divides the area A (Eq. (25)) by $B^{MAX} = \frac{1}{2}$, so that

$$2A = \frac{1}{N\sum_{i=1}^{N} x_i}\sum_{i=1}^{N}(2i-1)x_i - 1. \tag{26}$$

Expanding Eq. (26) and remembering that $\sum_{i=1}^{N} x_i = N\mu$, Pietra (1915) comes to the definition of the Gini index as we apply today[28]

$$G = \frac{\Delta_R}{2\mu} = \frac{\Delta_R}{\Delta^{MAX}} = \frac{1}{2\mu N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|x_i - x_j|. \tag{27}$$

The Gini index today G is equal to the ratio between the mean difference with repetition of the observed series and the mean difference without repetition of the maximising series. Its minimum value is zero, and its maximum one is $\frac{N-1}{N}$, which asymptotically tends to 1 when N increases.[29] Note that G and R are related by $\frac{N-1}{N}$:

$$G = \frac{N-1}{N}R. \tag{28}$$

The reason is intuitive: in evaluating the areas of the rectangles as interpreted with Eq. (21), one does not calculate the sum of areas of N triangles, which is argued to be equal to $\frac{1}{2N}$. To maintain the original interpretation while providing a precise graphical representation in the discrete case, the following expedient is necessary: $G = \frac{\Delta_R}{2\mu}$.

Finally, Pietra (1915) underlines: '*When the number of observed values is very high, the concentration polygonal chain becomes indistinguishable from the continuous smooth curve passing*

[28]Focusing on Eq. (10) and (11), Y can be rewritten as

$$\begin{aligned}
&Y = 2\sum_{i=1}^{N}\sum_{j=1}^{i}(x_i - x_j) = \\
&= 2(x_2 - x_1) + [(x_3 - x_1) + (x_3 - x_2)] + \cdots + [(x_N - x_1) + (x_N - x_2) + \cdots + (x_N - x_{N-1})] = \\
&= 2\sum_{i=1}^{N} ix_i - (N+1-i)x_i = 2\sum_{i=1}^{N}[2ix_i - (N+1)x_i].
\end{aligned}$$

Dividing it by N^2, he gets

$$\begin{aligned}
&\Delta_R = \tfrac{Y}{N^2} = \tfrac{2}{N^2}\sum_{i=1}^{N}[2ix_i - (N+1)x_i] = \\
&= \tfrac{2}{N^2}\sum_{i=1}^{N} 2ix_i - \tfrac{2}{N^2}\sum_{i=1}^{N} Nx_i - \tfrac{2}{N^2}\sum_{i=1}^{N} x_i = \\
&= \tfrac{2}{N^2}\sum_{i=1}^{N}(2i-1)x_i - \tfrac{2}{N^2}NW = \\
&= \tfrac{2}{N^2}\sum_{i=1}^{N}(2i-1)x_i - \tfrac{2W}{N}.
\end{aligned}$$

where $W = \sum_{i=1}^{N} x_i = N\mu$. By gathering $\frac{2W}{N}$ and noting that $\frac{W}{N}$ is equal to μ, he gets

$$\Delta_R = \frac{2}{N^2}\sum_{i=1}^{N}(2i-1)x_i - \frac{2W}{N} = \frac{2W}{N}\left[\frac{1}{NW}\sum_{i=1}^{N}(2i-1)x_i - 1\right]$$

from which

$$\Delta_R = 2\mu 2A$$

and, consequently,

$$2A = \frac{\Delta_R}{2\mu}.$$

[29]For our series $(3,4,5,6,7,8,9)$ of seven quantities $G = \frac{\frac{112}{49}}{12} = 0.190476$.

from its vertices. Such curve will be called the concentration curve.' Analytically representing the concentration curve with $y = \varphi(x)$, Pietra defines the Gini index in the continous case, expressing it as the ratio between the area A and the sum of the areas A and B:

$$\frac{A}{A+B} = 1 - 2\int_0^1 \varphi(x)dx. \tag{29}$$

From Eq. (7), observing that

$$\lim_{N\to+\infty} \Delta_R = \Delta \tag{30}$$

it is immediately clear that, in the continuos case,

$$G = \frac{\Delta_R}{2\mu} = \frac{\Delta}{2\mu} = R \tag{31}$$

validating the result obtained by Gini. Thanks to this last equation, Pietra is able to demonstrate, in the continuos case, that the concentration ratio R is exactly equal to the ratio between the concentration area and the area of maximum inequality, relationship that Gini affirmed but not rigorously demonstrated.

8. Conclusion

This essay retraced the two pioneering works that led to the definition of the Gini index as we apply it today. These works have not often been cited in the related literature and the contribution of Gaetano Pietra in particular has not been sufficiently supported.

While the Italian school has always focused its analysis on the discrete case, international literature immediately began to focus on analysing the continuous case (often not recognising Gaetano Pietra's work), due to which $\frac{N-1}{N}$ in Eq. (30) loses relevance, as does the distinction between the mean difference with and without repetition.

However, the primordial studies analysed in this essay enabled the application of the Gini index as we know it and apply it today. Historically, the preference has been to calculate the value of the index as in Eq. (27), because it can be interpreted perfectly as the ratio of areas under the Lorenz piecewise linear function, although Gini's original proposal was slightly different. Therefore, it was Corrado Gini that specified the validity of his index in light of Lorenz's curve and not vice versa.

Acknowledgements

I would like to thank Maria Giovanna Monti for numerous useful discussions and her exhaustive comments that have helped to improve this essay. Any error is the author's sole responsibility.

References

Amato, V. (1947): "Sulla differenza media," *Statistica*, VII, 31–37.

Atkinson, A. B. (1970): "On the Measure of Inequality," *Journal of Economic Theory*, 2, 244–263.

Benedetti, C. (1980): "Contro-esempi in statistica e probabilità," in *Studi in onore di Paolo Fortunati*, Clueb, Bologna, 23–39.

Bresciani-Turroni, C. (1916): "Recenti pubblicazioni: Gini C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri," *Giornale degli Economisti e Rivista di Statistica*, XXVII (serie III), 68–70.

Castellano, V. (1935): "Recente letteratura sugli indici di variabilità," *Metron*, XII, 101–131.

——— (1937): "Sugli indici relativi di variabilità e sulla concentrazione dei caratteri con segno," *Metron*, XIII, 31–49.

——— (1965): "Corrado Gini: a memoir," *Metron*, 24, 3–84.

Ceriani, L. and P. Verme (2012): "The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini," *The Journal of Economic Inequality*, 10, 421–443.

Chatelain, E. (1910): "Le tracé de la courbe des successions en France," *Journal de la Société Statistique de Paris*, 51, 352–356.

Czuber, E. (1914): "Beitrag zur Theorie statistischer Reihen," *Versicherungswissenschaftlichen Mitteilungen*, IX, 101.

Dalton, H. (1920): "The measurement of the inequality of incomes," *The Economic Journal*, 30, 348–361.

Dardanoni, V. and P. Lambert (1988): "Welfare rankings of income distributions: A role for the variance and some insights for tax reform," *Social Choice and Welfare*, 5, 1–17.

De Finetti, B. (1931): "Sui metodi proposti per il calcolo della differenza media," *Metron*, IX, 47–52.

De Finetti, B. and U. Paciello (1930): "Calcolo della differenza media," *Metron*, VIII, 1–6.

de Vergottini, M. (1940): "Sul significato di alcuni indici di concentrazione," *Giornale degli Economisti e Annali di economia*, 2, 317–347.

——— (1950): "Sugli indici di concentrazione," *Statistica*, X, 445–454.

Dorfman, R. (1979): "A Formula for the Gini Coefficient," *The Review of Economics and Statistics*, 61, 146.

Forcina, A. and G. Giorgi (2005): "Early Gini's contributions to inequality measurement and statistical inference." *Journal lectronique d'Histoire des Probabilits et de la Statistique*, 1, 3–15.

Fortunati, P. (1955): "Schema di graduazione, misure di variabilità e valori medi," in *Studi in onore di Gaetano Pietra*, Cappelli, Bologna, 403–430.

——— (1957): "Rapporto di concentrazione, valori medi e schemi teorici di distribuzione massimante e minimante della variabilità," *Statistica*, XVII, 143–157.

Gini, C. (1910): "Indici di concentrazione e di dipendenza," *Atti della riunione della società italiana per il progresso delle scienze*, 453–469.

——— (1912): *Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche*, C. Cuppini, Bologna.

——— (1914): "Sulla misura della concentrazione e della variabilità dei caratteri," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, 73, 1203–1248.

——— (1921): "Measurement of Inequality of Incomes," *The Economic Journal*, 31, 124–126.

——— (2005): "On the measurement of concentration and variability of characters," *Metron*, 63, 3–38.

Giorgi, G. (1990): "Bibliographic portrait of the Gini concentration ratio," *Metron*, XLVIII, 183–221.

——— (1992): *Il rapporto di concentrazione di Gini. Genesi, evoluzione ed una bibliografia commentata*, Università degli Studi di Siena.

——— (2005): "Gini's scientific work: an evergreen," *Metron*, LXIII, 299–315.

——— (2011): "Corrado Gini: the man and the scientist," *Metron*, 69, 1–28.

——— (2014): "A couple of good reasons to translate papers of the Italian statistical tradition," *Metron*, 72, 1–3.

Helmert, F. (1876): "Die Berechnung des wahrscheinlichen Beobachtungsfehlers aus der ersten Potenzen der Differenzen gleichgenauer direkter Beobachtungen," *Astronomische Nachrichten*, 88, 127–132.

Jordan, v. (1869): "Ueber die Bestimmung der Genauigkeit mehrfach wiederholter Beobachtungen einer Unbekannten," *Astronomische Nachrichten*, 74, 209–226.

——— (1872): "Ueber die Bestimmung des mittleren Fehlers durch Wiederholung der Beobachtungen," *Astronomische Nachrichten*, 79, 219–222.

Kakwani, N. (1984): "On the measurement of tax progressivity and redistributive effects of taxes with applications to horizontal and vertical equity," in *Economics inequality, measurement and policy*, ed. by G. Rhodes and R. Basmann, 149–168.

Lorenz, O. (1905): "Methods of Measuring the Concentration of Wealth," *Publications of the American Statistical Association*, 9, 209–219.

Money, L. G. C. (1905): *Riches and Poverty*, London: Methuen & Co. Ltd.

Pietra, G. (1915): "Delle relazioni tra gli indici di variabilità," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, 74, 775–804.

——— (1917): "Recenti pubblicazioni di metodologia statistica," *Rivista italiana di sociologia*, XXI, 308–322.

——— (1932): "Nuovi contributi alla metodologia degli indici di variabilità e di concentrazione," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, a.a. 1931-32*, XCI, 989–1008.

——— (1935): "A proposito della misura della variabilità e della concentrazione dei caratteri," *Supplemento Statistico ai Nuovi Problemi di Ferrara*, I, 3–12.

——— (1937): "Ancora sulle discordanze fra gli indici di variabilità e di concentrazione," *Giornale degli Economisti e Rivista di Statistica*, LII(serie IV), 666–673.

——— (2013): "On the relations between variability indices," *Statistica & Applicazioni*, 11, 7–22.

Pizzetti, E. (1955): "Osservazioni sul calcolo aritmetico del rapporto di concentrazione," 580–593, studi in Onore di Gaetano Pietra, Biblioteca di Statistica, n. 1, Cappelli, Bologna.

Ricci, U. (1916): "L'indice di variabilità e la curva dei redditi," *Giornale degli Economisti e Rivista di Statistica*, 53, 177–228.

Schneider, M. (2004): "Measuring Inequality: The Origins of the Lorenz Curve and the Gini Coefficient," Mimeo.

Séailles, J. (1910): *La repartition des fortunes en France.*, Alcan.

Shorrocks, A. (1983): "Ranking Income Distributions," *Economica*, 50, 3–17.

Shorrocks, A. and J. Foster (1987): "Transfer Sensitive Inequality Measures," *The Review of Economic Studies*, 54, 485.

von Andrae, H. (1869): "Schreiben des Herrn Geheimen Etatsrath von Andrä an der Herausgeber," *Astronomische Nachrichten*, 74, 283–284.

von Andrae, V. (1872): "Ueber die Bestimmung des wahrscheinlichen Fehlers durch die gegebenen Differenzen von m gleich genauen Beobachtungen einer Unbekannten," *Astronomische Nachrichten*, 79, 257–272.

Xu, K. (2003): "How has the Literature on Gini's Index Evolved in the Past 80 Years?" *SSRN Electronic Journal*.

Yitzhaki, S. and E. Schechtman (2013): *The Gini Methodology. A Primer on a Statistical Methodology*, Springer New York.

Yntema, D. (1933): "Measures of the Inequality in the Personal Distribution of Wealth or Income," *Journal of the American Statistical Association*, 28, 423–433.

DEPARTMENT OF ECONOMICS AND STATISTICS
UNIVERSITY OF TORINO
Corso Unione Sovietica 218 bis - 10134 Torino (ITALY)
Web page: http://esomas.econ.unito.it/