UNIVERSITÀ DEGLI STUDI DI TORINO
ALMA UNIVERSITAS TAURINENSIS
Founded in 1404

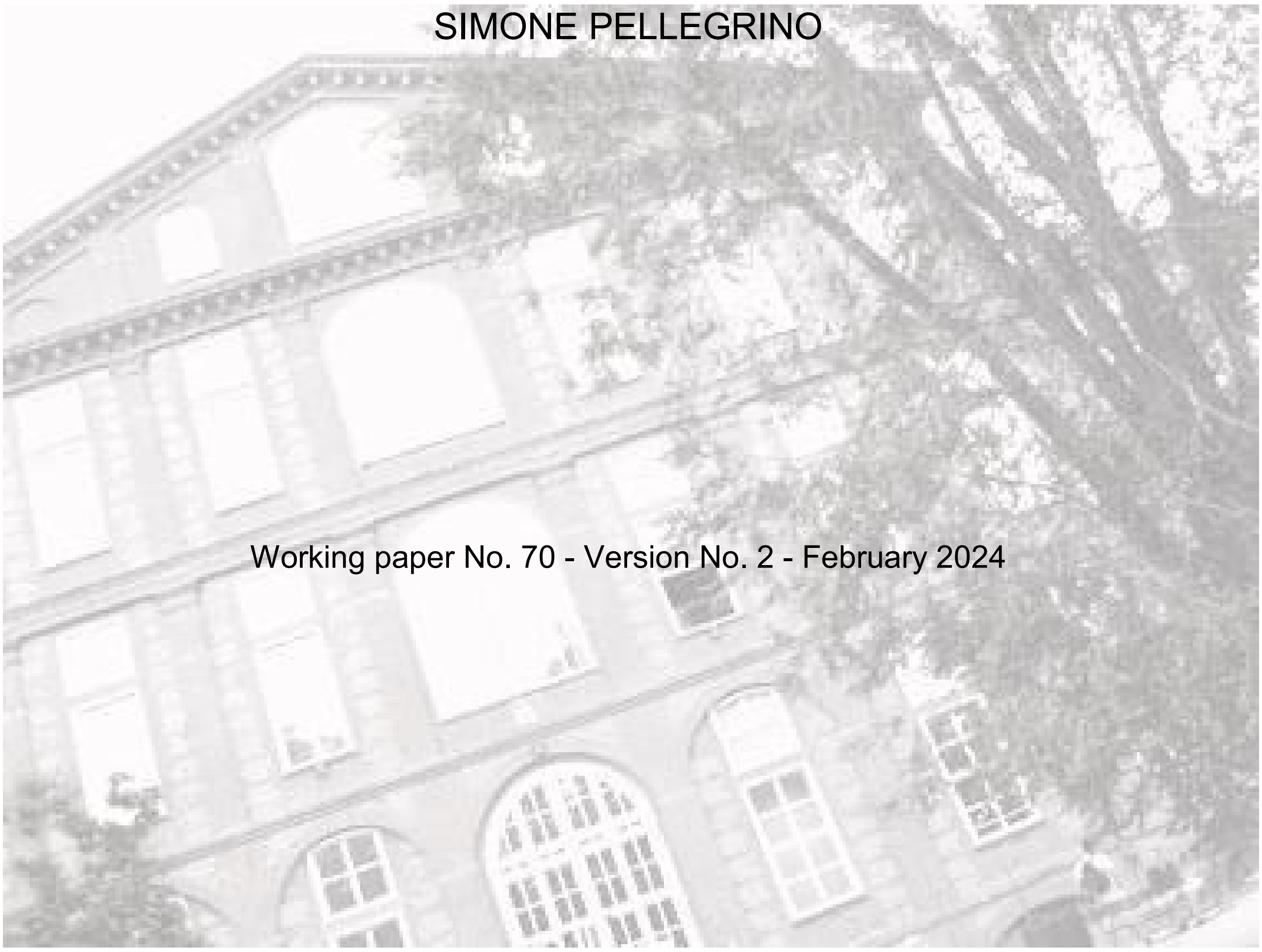
DEPARTMENT OF
ECONOMICS AND STATISTICS
WORKING PAPER SERIES

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

ISSN 2279-7114

THE GINI COEFFICIENT: ITS ORIGINS

SIMONE PELLEGRINO



Working paper No. 70 - Version No. 2 - February 2024

The Gini coefficient: its origins

Simone Pellegrino*

February 1, 2024

Abstract

This essay retraces the historical steps and analyses the theoretical motivation that influenced the definition of the Gini index G and its application today. In particular, it tries to link together the 'purely statistical' approach and the 'contextual' approach, related not only to the statistical methods discovered in the Gini's period but also to the succession of these discoveries. Having discussed the 'contextual' approach of these events, the remainder of the essay focuses on the 'purely statistical' approach, by presenting the statistical methods discovered by Corrado Gini and Gaetano Pietra as they chronologically appear in the years 1912, 1914 and 1915. The concept of mean difference, proposed by Corrado Gini in 1912 for applications in statistics and economics, is discussed. Then the difference between the concentration ratio R Gini advanced in 1914 and the Gini index G, 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, Social Studies, Applied Mathematics and Statistics – ESOMAS, Università degli Studi di Torino (IT). 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

The Gini index is undoubtedly the most popular and studied index used to assess income and wealth inequality. Its range of applications is very wide and endless literature has followed over the decades (Giorgi, 1990, 1992; Yitzhaki, 1998; Xu, 2003; Arnold, 2005; Basulto Santos and Busto Guerrero, 2010; Cowell, 2011; Yitzhaki and Schechtman, 2013a; Rogerson, 2015; Giorgi and Gigliarano, 2017; Giorgi, 2019; Mukhopadhyay and Sengupta, 2021).

After their discovery, the popularity of both the simple mean difference (1912) and the concentration ratio[1] (1914) started growing among Italian statisticians,[2] whilst among non-Italian scholars their diffusion was less rapid.[3] Dalton (1920)'s original statement according to which '*for the economist is primarily interested, not in the distribution of income as such, but in the effects of the distribution of income upon the distribution and total amount of economic welfare, which may be derived from income*' played instead no role in curbing the spread among scholars of inequality measures, such as those proposed by Gini,[4] based on incomes and not on the relation between incomes and social welfare levels (Monti et al., 2024). Some years later, Dalton's statement was also criticised by Yntema (1933) and '*went almost unnoticed for half a century*' (Brandolini, 2015; Atkinson and Brandolini, 2015), until the seminal paper by Atkinson (1970) from which the discussion between descriptive and ethical-founded indexes has started (Kolm, 1976a,b; Newbery, 1970; Sheshinski, 1972; Blackorby and Donaldson, 1977, 1978, 1980; Cowell, 1980; Kolm, 2011; Giorgi and Gubbiotti, 2017; Clementi et al., 2019; Pinto and Paidipaty, 2020; Gabbuti, 2020).

The less rapid diffusion, among international scholars, of Gini's discoveries can be also noted by analysing the temporal distribution of articles related to them. As a matter of fact, scientific interest in the Gini index exponentially grew, in terms of published scientific articles, precisely after Atkinson (1970)'s pioneering contribution (on these issues see Atkinson and Brandolini (2015) and Brandolini (2015)). By considering the hundreds of articles annotated by Giorgi (1990, 1992) until 1990, about three out of four have been published during the 70s and the 80s, and since then the interest in this inequality index has never stopped. One of the reasons is undoubtedly the new strand of literature related to the extension of the Gini coefficient to social judgements and to poverty (Dasgupta et al., 1973; Sen, 1973, 1974a,b, 1976a,b, 1978, 1979; Donaldson and Weymark, 1980; Weymark, 1981).

The main aim of this essay is to date back to the few seminal articles that let possible to reconstruct step by step the historical origins of what is today known as the Gini index. In so doing, two different perspectives are considered and linked together. Contributions about the history of the Gini index can indeed be classified into two approaches: the 'purely statistical' one, and a more 'contextual' one. These two approaches are not independent of each other, since there are several connections between them, primarily related not only to the statistical methods discovered in the Gini's period as well as in subsequent decades but also to the succession of these discoveries (Ogburn and Thomas, 1922; Merton, 1961; Niehans, 1995; Lamb and Easton, 1984).

This essay tries to fill this gap by focusing on four fundamental works: the first is the pioneering work by Lorenz (1905), the second is the work by Corrado Gini (1912), the third is the pioneering work by Corrado Gini in 1914, while the forth is the note published by Gaetano Pietra the following year. All of them, togheter, help us understand the steps and theoretical motivations that led to

[1] Also the debate on how to calculate it has always attracted scholars, yesterday as today; see, among others, Dorfman (1979), Milanovic (1997) and Furman et al. (2019).

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

[3] Dalton (1920) specifically underlines that in that time '*the problem of the measurement of the inequality of incomes has not been much considered by English economists . . . but it is in Italy that it has hitherto been most fully discussed.*' Furthermore, he notes that at that time the study of income and wealth inequality was slowed down by '*the inadequacy of the available statistics*'. See also Gabbuti (2020) on this issue.

[4] Dalton (1920), who knew some Italian, cites the work by Gini (1912) (see footnote 15) and by Ricci (1916).

the definition of the Gini index, G, and its application today. As a matter of fact, the two seminal articles by Corrado Gini cannot be separately analysed by the one which first defined the Lorenz curve; similarly, the article published by Gaetano Pietra cannot be separately analysed by the two articles by Corrado Gini and the one by Max Otto Lorenz.

In particular, we have to deal with four different discoveries: the discovery of the Lorenz (1905) curve; the discovery of the simple mean difference; the discovery of the concentration ratio R as defined by Gini (1914); the discovery of the exact corrispondence between the Lorenz curve and the Gini's way of measuring inequality, developed by Pietra (1915). The discovery of the Lorenz curve has been a multiple discovery (Derobert and Thieriot, 2003; Schneider, 2021). It is not possible to exactly infer whenever the discovery of the simple mean difference has been a multiple discovery or a chain multiple one (see Subparagraph 2.2); it is debated if the formula of the concentration ratio Gini proposed in 1914 has been an independent discovery or a chain multiple discovery (see Schneider (2021) and the discussion presented in this work); the discovery of the exact formula we employ today for the computation of the Gini coefficient has been undoubtedly a chain multiple discovery.

Basically, this essay covers the following points.

In 1912, Corrado Gini proposes the concept of simple mean difference (with and without repetition) as an index of variability for quantitative values (according to the last of the three definitions Gini (1914) comments, the simple mean difference (Gini, 1912) is the numerator of the concentration ratio R which he would discover two years later (see Paragraph 4)), which soon became a fundamental indicator for studies in statistics and economics (see Paragraph 2). The objective of his work was to highlight[5] '*...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[6] the application of the simple mean difference of observed quantities as an indicator that may be preferred over others in some areas of study.

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.[7]

First, he proposes the concentration ratio R (pp. 1203-1228), which would result in the Gini index G, as it is applied today. The concentration ratio presents some particular aspects (see Paragraph 3). 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 (Gini (1914) does not provide this interpretation, but it is useful to refer to it in explaining his reasoning) as a ratio of the sums of segments. Note that R, 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 (see Paragraph 6) to observe that[8] '*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),*

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

[6]Prior to this, Gini (1909, 1910) had already begun to examine concentration indexes; see also Maccabelli (2009) and Gabbuti (2020).

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

[8]'*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.*'

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 character becomes very large.

Finally, in the last part (pp. 1236-1240), Gini verifies (see Paragraph 4) that[9] '*the concentration ratio coincides with 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 order and the timing with which Corrado Gini published his innovations have raised, among the historians of economic thought, a discussion about the role that the Lorenz (1905) curve had in defining and refining Gini's contributions. In order to study the context in which the index was developed, Derobert and Thieriot (2003) and Schneider (2021) deal with this issue. In particular, Schneider (2021) offers an interpretation of the origin of the concentration coefficient R according to which the Lorenz curve played a role as 'catalyst'. He argues that '*... Gini might not have made his discovery had it not been for his becoming aware of the curve invemted by Lorenz*'. His reasoning is the following: Gini published different papers on this topic before 1914 (Gini, 1909, 1910, 1912) and he never used diagrams in explaining his discoveries; on the contrary, in 1914 he decided to compare his formula of the concentration ratio R with the diagram proposed by Lorenz. According to Schneider (2021), this '*seems to imply that Gini came across the curve in the course of reading works by* Lorenz (1905), Chatelain (1910) and Séailles (1910)', that is, logically, after 1912. This interpretation of events is reasonable, even more so by observing that '*if he had discovered the coefficient by 1912, he undoubtedly would have included it in the pamphlet he published in that year*' (Schneider, 2021).

To summarize, Schneider (2021) concludes '*that there are two ways in which Gini may have arrived to his discovery*': 1) by making progress with respect to his 1910 article; 2) '*by working out the algebraic formula that describes the Lorenz curve*'. The second hypothesis is unlikely to be true for the reasons explained in the remainder of this introduction and in the remaining part of the essay. The first one is likely to be true: Gini (1910)'s article focuses on the idea of considering the top quantities instead of the bottom ones. He could then have the idea of focus on the bottom quantities instead of the top ones by earnest making progress with respect to his 1910 article; or this idea may have been facilitated after reading the articles by Lorenz (1905), Chatelain (1910) and Séailles (1910) (see Eq. (15) of Paragraph 3). The same reasoning could also be applied to the formula evaluating the concentration ratio R as the ratio between the mean difference without repetition of the actual distribution and its maximum value always without repetition (see Eq. (17) of Paragraph 4), since the maximum value is somewhat related to the maximum area of the Lorenz's dyagram. What is perhaps most surprising is that Gini (1914) decided to immediately discuss the concentration ratio as the ratio of sums of segments and afterwards to discuss its identity by employing the mean differences; a greater importance would have been given to the simple mean difference, the indicator to which Gini dedicated an entire book two years earlier, were the order of presentation had been the opposite.

However, it has to be said that it is difficult to trace the exact timing of these discoveries. As the definition (and the paternity) of the simple mean difference (see Subparagraph 2.2) can be attributed to various authors (Jordan, 1869; von Andrae, 1869; Jordan, 1872; von Andrae, 1872; Helmert, 1876; Gini, 1912), it may also have happened that Gini defined by himself the concentration ratio R after 1912 (by solely making progress with respect his 1910 article) and even later discovered and read Lorenz's article. In this circumstance, the ordering in which he would have presented his discoveries is

[9]'*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.*'

reasonably the same observed in his 1914 article; moreover, always in this case, it sounds reasonable that he also employed graphical representations, with which he was previously unfamiliar, since the aim was to compare his concentration ratio and the area of the Lorenz curve.

Turning to the second hypothesis by Schneider (2021), according to which Gini worked out the algebraic formula that describes the Lorenz curve, it is questionable, since the concentration ratio R is not the algebraic version of the Lorenz curve. Nor it was possibile for Gini to state this identity by looking to the Lorenz curve: the concentration ratio R can be viewed (see Paragraph 6) as the ratio between sums of segments or the ratio between sums of rectangular areas, whilst the area of the Lorenz curve is the sum of areas of trapezoides (Pietra, 1915) (see Paragraph 7). As a consequence, focusing on R and the Lorenz curve, Gini was only able to skech the relationship between them by stating that the identity between R and the area of the Lorenz curve is only verified when the number of quantities is infinite. As a consequence, it is unlikely that Gini defined the concentration ratio R trying to determine the formulas underlying the Lorenz curve: in this case he should have obtained the formula we apply today to calculate the Gini index G, which was evaluated by Gaetano Pietra in 1915 precisely following this reasoning.[10]

In particular, Pietra (1915)[11] studied the link[12] between the concentration ratio R, proposed by Gini, and, by looking to the Lorenz curve, the ratio between the area of observed concentration and the area of maximum concentration, providing an elegant geometrical interpretation.[13] 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 – is the exact formula for the Gini index commonly applied today. Note that it does not vary between zero and one (as the original concentration ratio R), but between zero and $\frac{N-1}{N}$, in order for the exact link between the index and the Lorenz curve to be satisfied.

Also Pietra's contributions are illuminating in linking the 'purely statistical' approach as well as the 'contextual' one. As it is well known, a good inequality measure should respect four properties: anonymity, population principle, principle of transfers and scale invariance. Focusing on the population principle, it requires the inequality index not to vary whenever the distribution is replicated a finite number of times. It is easy to show that the concentration ratio R, as originally proposed by Gini, does not satisfy this principle, whilst the Gini index G as we apply it today does.[14] It is not possible to know, but perhaps the fortune of the concentration ratio R would have been different without Gaetano Pietra's theoretical contributions. And Gaetano Pietra's contributions would not have been possible without a comparison between the concentration ratio R and the formula defining the Lorenz curve. In this sense the Lorenz curve surely played a role as 'catalyst'.

There is still one point to be discussed: the role that language barriers played in allowing these theoretical innovations to develop in the international scientific community, then and now. Today again, many non-Italian scholars who work on income and wealth inequality have not read the original articles; it is therefore a good idea to propose the translation of these articles into English (Zanella and Leti, 1995; Arnold, 2005; Giorgi, 2011, 2014).[15]

[10] See also Dorfman (1979).

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

[12] The relationships between the simple mean difference and the concentration curve, first analysed in Pietra, were further highlighted and discussed in Gumbel (1928).

[13] The same article contains the first definition of the Gini index for the continuous case (more than 60 years later, Dorfman (1979) would propose an approach to unite the calculation of the continuous and discrete Gini indices) and the introduction of the concept of 'graduation', that is, the inverse of the distribution function.

[14] In particular, as we will discuss, $R > G$, and the difference $R - G$ decreases when the number of replication increases.

[15] 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*

Consider, for example, the complete and exaustive book by Yitzhaki and Schechtman (2013a). Chapter 2 discusses the alternative ways of computing the simple mean difference (Yitzhaki, 1998; Yitzhaki and Schechtman, 2013b), that is the numerator of the index. They cite the review of Gini (1912)'s simple mean difference by Ceriani and Verme (2012), appeared in an international journal, and in footnote No. 1 they state '*Unfortunately we are unable to survey the Italian literature which includes, among others, several papers by Gini, Galvani, and Castellano*', even if at least the English translation of Gini's 1914 article was already published in 2005 in Metron, an Italian journal (Giorgi, 2020). Similarly, Lambert (2012), in reviewing Ceriani and Verme (2012)'s article, states that Gini (1912) '*was published 100 years ago in Italian but has never been translated into English, and did not reach an international audience. ... Despite the numerous formulations of the Gini index that have appeared in the literature over the years, the original formulations are largely unknown. ... In Xu (2003)'s survey article, it is striking that none of the 14 different formulations of the Gini index reviewed by Xu corresponds to any of the 13 indexes elaborated by Gini himself.*' Moreover, in describing the discrete case, Yitzhaki and Schechtman (2005) say '*... in terms of areas, the usual practice is to obtain a Lorenz curve of a discrete distribution as a piecewise linear curve by connecting the points by straight lines*' and they cite Gastwirth (1971, 1972), whilst the original idea is due to Gini (1914) and Pietra (1915) (see Paragraph 5).

Over the decades, the language barrier had the consequence of curbing the diffusion, outside Italy, of other articles written in Italian by Italian authors.[16] To take an example related to the topics under discussion, what is today known as the Schutz (1951) coefficient (which '*measures that proportion of total income which would have to be transferred from incomes above the mean to incomes below it to achieve complete equality*' (Lambert, 2001)) was previously defined by Pietra (1915). Despite Rosenbluth (1951) noted[17] that the notion of relative mean deviation was also proposed by Ricci (1916) (and by Bresciani-Turroni (1916)), who was in turn cited by Yntema (1933), this coefficient is not yet attributed to Pietra (1915) and Ricci (1916) by the international scientific community. Of my knowledge, only the work by Costa and Pérez-Duarte (2019) underlines a multiple paternity of this coefficient, known as the Pietra-Ricci index by the Italian school of statistics (Frosini, 1996).[18]

The language barrier was also the cause of disputes. It is what happened in 1930 on the occasion of the conference of the International Statistical Institute in which von Bortkiewicz (1931a) presented a paper containing results obtained several year before by Gini (1912) and Pietra (1915). As a consequence, both Gini and Pietra replied (Gini, 1931b; Pietra, 1931b). In the counter-response (von Bortkiewicz, 1931b; Gini, 1931a; Pietra, 1931a) Bortkiewicz admitted that he did not read their articles (Forcina and Giorgi, 2005; Giorgi and Gubbiotti, 2017).

In order to continue the discussion on language barriers, a focus on some international papers published in the past can be done.[19] For example, consider the article by Nair (1936); he examines

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.*'

[16] On the contrary, the writings in English of Italian authors have had a different fortune. For example, Goodman and Kruskal (1959) cited the work by Pietra (1925).

[17] Also Dalton (1920) cites the work by Ricci (1916).

[18] On this issue the paper by Kondor (1971) provides some evidence as well.

[19] 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). For an account of his career from a historical perspective see Prévost (2016b).

the simple mean difference without repetition, whilst today we employ the one with repetition as the numerator of the Gini index. On the contrary, Pyatt (1976), in order to disaggregate the Gini coefficient by employing the discrete analysis, considers, as numerator, the simple mean difference with repetition, as in the Pietra (1915)'s seminal contribution.[20] In all likelihood, this alternate use of the simple mean difference with and without repetition is due to the fact that almost all the international literature focused on the study of continuous rather than discrete income and wealth distributions; as a conseguence, using the continuous case, this difference disappears (see Paragraph 7).[21]

Finally, the language barrier had another (obvious) consequence: the bibliography of many international scientific articles do not contain references to these original articles (Herzel and Leti, 1977).[22] Moreover, the note by Pietra (1915) is never cited in the literature (exceptions are very rare, see for example Giorgi's contributions, and often Gaetano Pietra is only credited with having defined the concentration ratio in the continuous case and not also with the improvement of the index itself), despite its original ideas and the fact that Gini (1921) himself cited Pietra (1915)'s contribution in his reply to Dalton (1920). There are also examples in which not only the original contributions by Gini (and by Pietra) are not cited, but also the name of Corrado Gini does not appear within the whole text of the article: this is the case of the paper by Rao (1969).

More generally, the analysis of Corrado Gini's role in the history of both statistics and econometrics would deserve a separate discussion. Corrado Gini played a crucial role on the development of both disciplines, although, for several reasons, today his contributions are not always fully recognized by the international economic community, even if Corrado Gini was an honorary fellow of the Royal Statistical Society since 1920 and also honorary member of the International Statistical Institute (Nixon, 1960; Bemmann, 2023) as well as he was visiting professor the University of Harvard and later held honorary degree from the same University, among others (Gini, 1956; Boldrini, 1966; Louçã, 2007; Parisi, 2011, 2013; Rogerson, 2015) and he was also a founding member of the Econometric Society (Gini, 1956).[23] For example, Parisi (2011) observes that even if the Italian tradition contributed to the birth of econometrics, its contribution '*has never considered crucial – despite the involment of Corrado Gini . . . who helped found the international society on economics, statistics and mathematics that became the Econometric Society. . . . It is in statistical sources that information on Gini is actually found.*'

This is due to the fact that at that time Italian statistics tradition (Pietra, 1939; Gini, 1965; Herzel and Leti, 1977; Zanella and Leti, 1995) was primarily involved to (descriptive) statistics, demography and economics (Benini, 1894; Gini, 1909; Frosini, 1996; Gabbuti, 2020) and less to mathematics (Gini, 1926).[24] Moreover, the Italian statisticians '*preferred to analyse the characteristics of entire population, making use of sampling as a secondary measure only or, as Gini asserted so synthetically, paying greater attention to complete populations rather than to samples*' (Herzel and Leti, 1977). Other circumstances played a crucial role on the scientific dialogue between the Italian Statistical Society

[20] See also Yitzhaki (2003).

[21] However, this does not mean that the use of continuous or discrete analysis is always equivalent. Consider for example the paper by Kakwani and Lambert (1998). By employing the continuos case, authors state three axioms which should be respected by an equitable tax system and propose a measurement system to evaluate the negative influences that axiom violations exert on the redistributive effect of taxes. In order to satisfy the theoretical achievements, a proper empirical analysis should be conducted by employing the discrete case (Pellegrino and Vernizzi, 2013).

[22] The paper by Goodman and Kruskal (1959) is a rare example.

[23] The good relationships among the Italian school and the international statisticians is also confirmed by the article by Fisher (1925) published in Gini's journal Metron.

[24] Gini (1926) underlines that the 'pure' statisticians are more frequent in Italy than elsewhere, and underlines: '*By 'pure Statistician' I mean the scholar who makes Statistics and its applications the principal aim of his scientific activity, and does not study it merely as ancillary to researches in Economics, Finance, Anthropology, Psychology, or Medicine. . . . Though several of the Italian statisticians come from Mathematics, I do not think that anybody attaches importance to the controversy whether Statistics should be treated as Mathematics of with Mathematics. It is universally admitted that, in Statistics, Mathematics are to be considered as a means for presenting the subject in a more or less elegant form, but this merit must not let us lose sight of the fundamental truth that for Statistics they are no more than a means.*'

and the International Statistical Institute as well as the Econometric Society in the following dacades, such as the scientific (Parisi, 2013), political and social atmosphere (Favero, 2010; Prévost, 2016a,b; Favero, 2017; Gabbuti, 2020); the Second World War also proved to be a watershed for the relations between the Societies, besides '*the excessive inflexibility shown by Gini against the inferential theories*' (Giorgi and Gubbiotti, 2017); following the wake of the statistics and mathematics of the Anglo-Saxon countries, after the Second World War also Italian economists and statisticians started using mathematical models more and more pervasively (Parisi, 2013).

Given '*his special gift for concreteness*' (Herzel and Leti, 1977), Gini's contributions were surely concerned with practical problems predominantly related to statistics and demography as well as economy.[25] He studied "*... those phenomena with manifest themselves in finite groups and which at any rate in theory could be completely observed... For such groups there are no sampling or generalization problems and therefore the problem of expressing as a general law , with limits of error, the descriptive properties through which they are manifested does not arise*" (Boldrini, 1966). This vision of statistics was certainly enhanced by the fact that he was (the first) president of the Italian National Statistical Institute (1926-1932) (Istituto Centrale di Statistica del Regno d'Italia, 1929; Alleva, 2015) and that '*Gini's view of probability was limited, being more related to the traditional ideas of classical authors... rather than on the neo-Bayesian approach introduced by de Finetti'* (Giorgi and Gubbiotti, 2017); see also Forcina (1982), Herzel and Leti (1977), Piccinato (2011) and Prévost (2016b) on this issue. Despite this approach to the study of statistics, Gini also dealt with the problem of balanced sampling (Gini and Galvani, 1929; Neyman, 1934; Langel and Tillé, 2011).

At the end of his career (criticisms related to the use of mathematical models can also be found earlier; see Gini (1926, pp. 707) on the laws of growth of a population by Knibbs (1926)), Gini rejected the use of (a priori) probability in statistics (Gini, 1939, 1941, 1943, 1947, 1949, 1956, 1964, 1965; Herzel and Leti, 1977; Frosini, 1996; Forcina, 1982; Frosini, 2005, 2008; Piccinato, 2011) and refused results reached by econometricians by employing mathematical models (Dagum, 1968) to explain reality,[26] since[27] *'... in most cases, the traditional methods in use in economic statistics, aimed at quantitatively expressing the state and movement of the observed phenomena, may be sufficient, resorting only exceptionally to methods of a higher nature, which always involve hypotheses, and often highly restrictive hypotheses, which not all economists realize and which are difficult to find confirmation in reality'* (Gini, 1956). Parisi (2011) well conludes this discussion by saying '*that Gini expected to handle mathematical tools when required by statistics; he did not consider mathematics as a basis to construct models.*'

Having discussed the 'contextual' approach of these events, the remainder of the essay focuses on the 'purely statistical' approach, by presenting the statistical methods discovered by Corrado Gini and Gaetano Pietra in 1914 and 1915, respectively. The only exception concerns the historical origins of the Gini's simple mean difference (1912), since this concept deserves a separate explanation. The essay is organised as follows: Subsection 2.1 introduces the concept of mean absolute difference, while

[25]It has not been that way his entire life. As Boldrini (1966) discusses, when he was young '*Gini's first interest was the statistical theory and application of the calculus of probability. ... Towards the end of his life his interest in this subject revived...*'. See also Forcina (1982), Parisi (2011), Giorgi (2005), Herzel and Leti (1977) and (Prévost, 2016b).

[26]He also observes that he accepted Irving Fisher and Ragnar Frisch's invite to be a member of the Econometric Society, '*while not promising myself exceptional results from the application of the refined methods of statistics and mathematics to the economy, which the new Society had planned*'. [Io sono stato uno dei soci fondatori di detta società, di cui Irving Fisher fu il patrono e Ragnar Frisch l'ideatore. Ho accettato il loro cortese invito, pur non ripromettendomi eccezionali risultati dall'applicazione all'Economia dei metodi raffinati della Statistica e della Matematica, che la nuova società aveva in programma.]

[27]*[... nella maggior parte dei casi, possono bastare i metodi tradizionali in uso nella statistica economica, diretti ad esprimere quantitativamente lo stato e il movimento dei fenomeni osservati, ricorrendo solo in via eccezionale a metodi di carattere più elevato, i quali implicano sempre ipotesi, e spesso ipotesi fortemente restrittive, di cui non tutti gli economisti si rendono conto e che difficilmente trovano riscontro nella realtà.]*

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 Max Otto 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 $x_1, x_2, \ldots, x_{N-1}, x_N$, 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

$$\begin{aligned}&(x_2 - x_1) + (x_3 - x_1) + \cdots + (x_{N-1} - x_1) + (x_N - x_1) = \\ &(x_2 + x_3 + \cdots + x_{N-1} + x_N) - (N-1)x_1 = \\ &x_1 + x_2 + x_3 + \cdots + x_{N-1} + x_N - Nx_1.\end{aligned} \tag{1}$$

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:

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

Adding up all the N equations and rearranging them, Gini (1912) gets[28] a first formulation able to express the arithmetic mean of the $N(N-1)$ possible differences between the N quantities, i.e. the

[28]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 (2013b)). 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 (Czuber, 1914; Pietra, 1915, 1925; De Gleria, 1929, 1930; 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.']* Today even faster methods are applied: as demonstrated by Stuart (1954) and applied by Pyatt et al. (1980) and Lerman and Yitzhaki (1984), the simple mean difference as well as the Gini index G can be evaluated by employing the covariance between the cumulative distribution function and the income distribution. See also Jenkins (1988) and Milanovic (1997) for empirical applications and Yitzhaki (2003) and Schechtman and Zitikis (2006).

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; today, most frequently, labeling Y the sum of differences in absolute value

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

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,[29] '*...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).

Corrado Gini (1914) claims that he 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),[30] that is, the mean absolute

[29] '*...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*'.

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

difference without repetition, and Δ_R, that is, the mean absolute difference with repetition (Formula 7).[31] 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,[32] '*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,[33] '*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 difference between multiple quantities, from an entirely different perspective.*'[34] 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.[35] 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[36] '*for the question of calculating probabilities relating to the theory of observational errors.*' The author concludes the discussion by stating,[37] '*As for the scope of this work mentioned above, it is clear that it lies rather*

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

[32] '*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.*'

[33] '*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.*'

[34] '*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.]

[35] '*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.']

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

[37] '*Quanto allo scopo, già accennato, di tali lavori, si comprende facilmente che esso è ben lontano da quello dello*

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.'

To conclude, it is difficult to reconstruct the origins of Gini's own work on the discovery of this formula, also because it is well known that Gini's archive is impossible to access. Corrado Gini may actually have discovered this formula independently of the work by German astronomers, or not. Some doubts may arise if only because we are asked to take for granted Gini's or his collaborators' own word. Moreover, it is known that Gini was very jealous of his discoveries (see footnotes 15 and 28), he was '*... by nature individualistic and exclusive*' (Boldrini, 1966)[38] and we also know from historians that, in more politically-heated issues, Gini can at least be suspected of not being precisely impartial in his science (Favero, 2010; Prévost and Beaud, 2015).

3. The concentration ratio

In the 1914 publication Corrado Gini immediately proposes the concentration ratio R. He considers N '*quantities that measure the intensity of a certain character in*' N '*different cases*', $x_1, x_2, \dots, x_i, \dots, x_l, \dots, x_{N-1}, x_N$, ordered in non decreasing order, so that $x_{k-1} \leq x_k \; \forall k$, with $k = 1, 2, \dots, N$. 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

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

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, also 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

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.'

[38]His '*authoritarian and detached behaviour*' observed during his working life faded during the retirement period (Giorgi and Gubbiotti, 2017).

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, $R = \frac{2}{9} = 0.\bar{2}$.

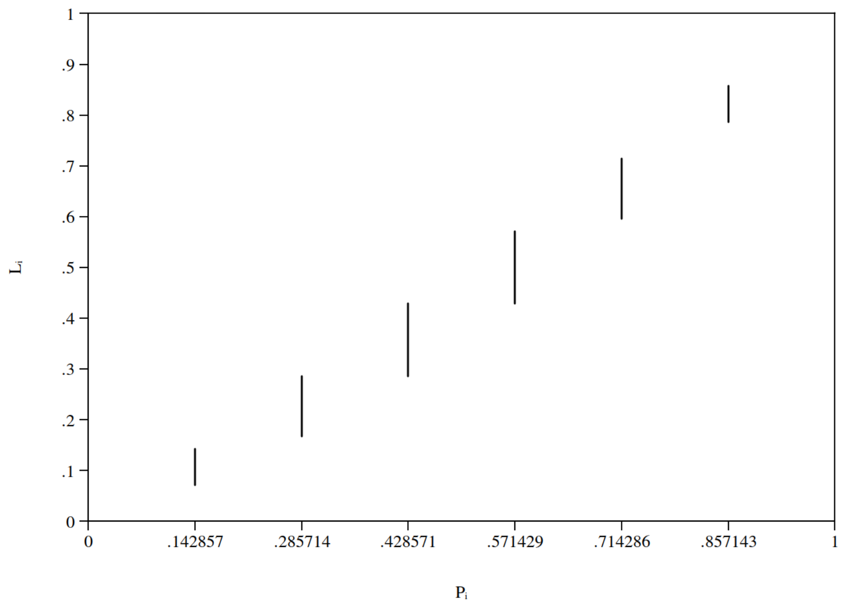
Table 1: The calculation of the concentration ratio R

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. Figure 1 is not mentioned in the original writings by Gini, but it is very useful in order to study in depth 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, p. 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.[39]

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

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

that he rewrites as[40]

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

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, p. 1238).

5. The Lorenz curve and piecewise linear function

Max Otto Lorenz (1905) devised[41] a graphical view[42] of inequality, proposing a comparison between the cumulative portion, L_i, of a quantitative variable[43] and the cumulative portion of frequencies P_i,

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] In his original article, Lorenz applies the reasoning indiscriminately to the distribution of income or wealth. See Derobert and Thieriot (2003) for a discussion of the historical origins of the Lorenz curve.

having ordered these frequencies from poorest to richest.[44] 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 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*' $\big(P_N = 1,\, L_N = 1\big)$.

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 coordinates of the Lorenz piecewise linear function are the pairs (P_i, L_i) connected by lines. Focusing on the same series discussed before, $(3, 4, 5, 6, 7, 8, 9)$, 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 series 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).

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

[44]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 Max Otto Lorenz the same year.

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

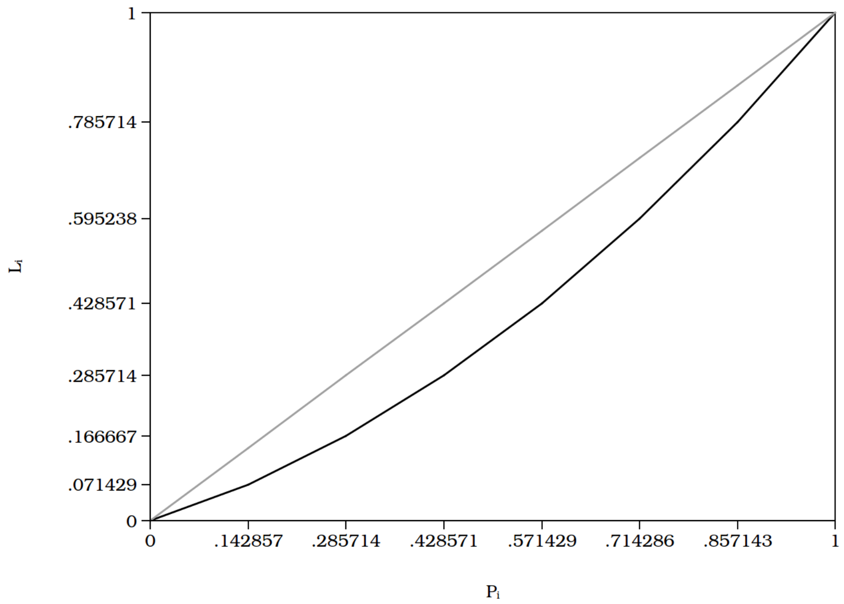
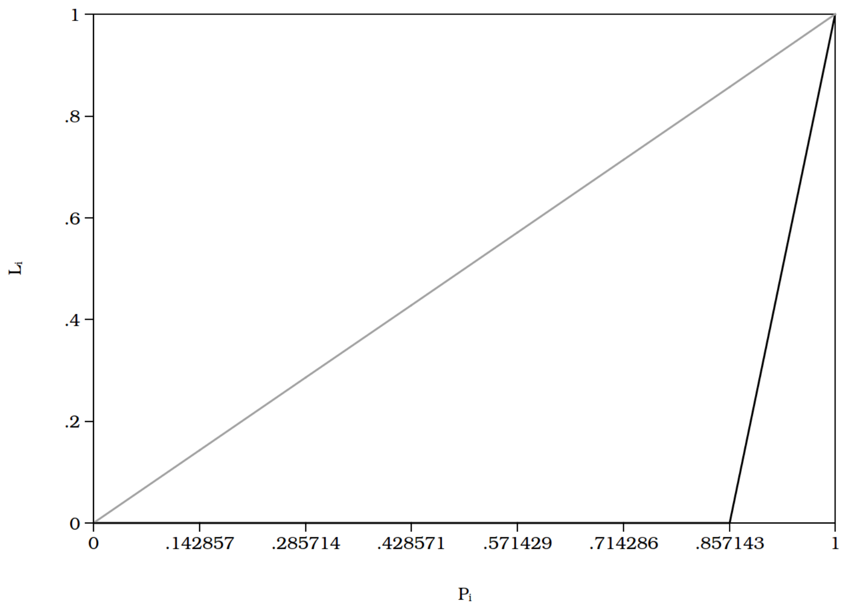


Figure 3: The Lorenz piecewise linear function with maximum inequality $(N = 7)$



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.[45] 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.[46] 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 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 $\frac{1}{N}\sum_{i=1}^{N-1} L_i$.

He then considers the $N-1$ rectangles limited by the x-axis and the equal distribution line (Figure

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

[46] As Dagum (1980) observes, to each Lorenz curve can be associated a Gini index G or a concentration ratio R; the reciprocal is not guaranteed, since to each value of G or R different Lorenz curves can be associated.

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.

Figure 4: Rectangles by considering L_i

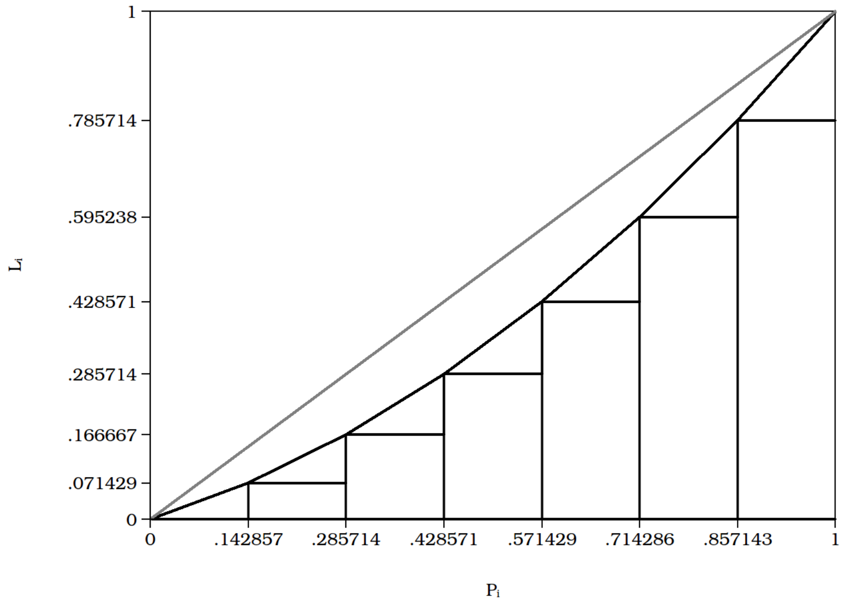
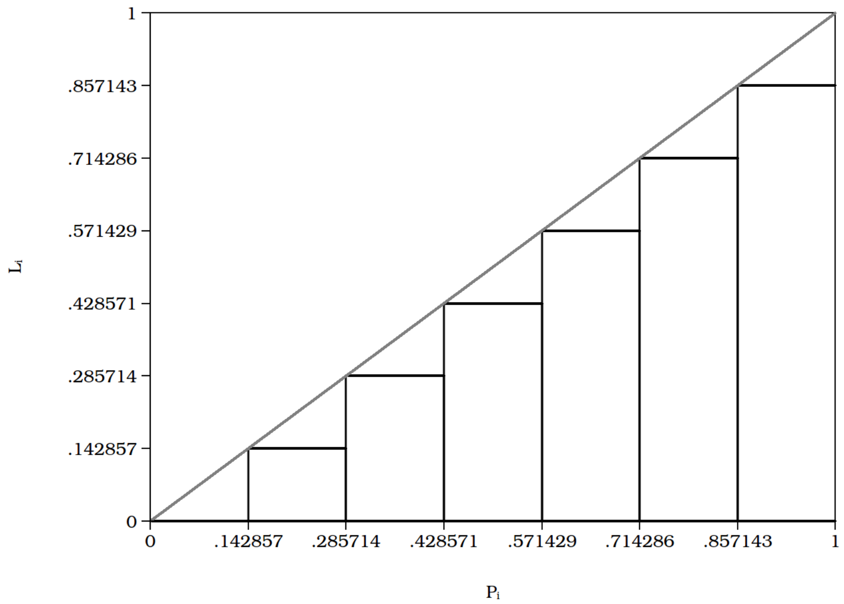


Figure 5: Rectangles by considering P_i



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 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

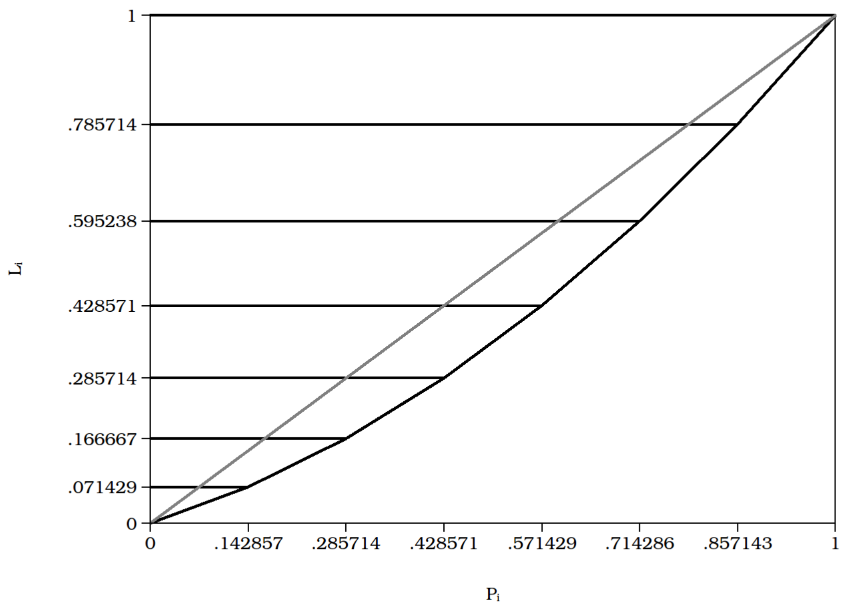
One year after the publication of the work by Gini (1914), Gaetano Pietra (1915), one of his collaborators, published a note regarding the relationships between the indices of variability. His work certainly did not reach the notoriety of Gini's; however, it contains results that were truly important 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:[47]

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

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

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

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.[48] 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 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.

Dividing it by N^2, he gets

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

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,

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

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

8. Concluding remarks

In retracing the historical steps that influenced the definition of the Gini index and its application today, this essay tries to link together the 'purely statistical' approach and the 'contextual' approach, related not only to the statistical methods discovered in the Gini's period but also to the succession of these discoveries.

The paper focuses, in particular, on the discovery of the simple mean difference, the discovery of the concentration ratio as defined by Corrado Gini in 1914 and the discovery of the exact corrispondence between the Lorenz curve and the Gini's way of measuring inequality, that is analysed in a paper by Gaetano Pietra in 1915.

It is argued that it is not possible to say if the discovery of the simple mean difference has been a multiple discovery or a chain multiple one: Corrado Gini claims that, as he finished writing his 1912 book, he became aware of the articles by the three German astonomers which forty year earlier proposed the same definition of the simple mean difference. Moreover, it is also debated if the formula of the concentration ratio Gini proposed in 1914 has been an independent discovery or a chain multiple discovery; finally, it is argued that the discovery of the exact formula we employ today for the computation of the Gini coefficient has been undoubtedly a chain multiple discovery, since Gaetano Pietra defined it by studying both the Lorenz curve and the concentration ratio.

The role as 'catalyst' that the Lorenz curve may have played in helping Gini to define the concentration ratio, as proposed by Schneider (2021), is also discussed. In particular, Schneider argues that there are two ways in which Gini may have arrived to his discovery: by making progress with respect to his 1910 article or by working out the algebraic formula that describes the Lorenz curve. The first hypothesis is likely to be true, since Gini started studying income and wealth inequality a few years earlier. The second one is instead unlikely to be true: it is possible that Gini may have drawn inspiration from reading Lorenz's article, but it is unlikely that he derived his formula by looking at the Lorenz curve, since the concentration ratio he proposed is not the algebraic version of the Lorenz curve (and indeed he was only able to skech the relationship between the concentration ratio and the area of the Lorenz curve when the number of quantities is infinite). If this were the case, he should have obtained the formula we apply today to calculate the Gini index, which was evaluated by Gaetano Pietra in 1915 by unequivocally looking at the Lorenz curve and the concentratio ratio.

In order to underline these issues, the essay presents the statistical methods discovered by Corrado Gini and Gaetano Pietra as they chronologically appear in the years 1912, 1914 and 1915. In particular, having defined the concept of mean difference Gini proposed in 1912, the difference between the concentration ratio Gini advanced 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.

The scientific community preferred and prefers to employ Pietra's revised version of the index since this formula can be perfectly interpreted as the ratio of areas under the Lorenz piecewise linear function, although Gini's original proposal was slightly different.

Finally, a few considerations on the role the language barrier had in allowing these theoretical innovations to spread into the international scientific community are also discussed. It is well known that many non-Italian scholars who work on income and wealth inequality have not read the original articles of the Italian statistics tradition; consequently, peculiarities as well as details of the articles by Corrado Gini and Gaetano Pietra are not well known, and the bibliography of many scientific articles do not contain references to these articles. Finally, Gaetano Pietra's decisive contribution to the most popular inequality index we use today is even less known.

Acknowledgements

I would like to thank Maria Giovanna Monti for numerous useful discussions and her exhaustive comments that have helped to improve this essay. I am also very grateful to Achille Vernizzi and three anonymous referees. Any error is the author's sole responsibility.

References

ALLEVA, G. (2015): "Corrado Gini e la fondazione dell'Istituto nazionale di statistica," *Il pensiero e l'opera di Corrado Gini, Accademia dei Lincei, Roma, 4 giugno.*

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

ARNOLD, B. (2005): "Inequality measures for multivariate distributions," *Metron*, 63, 317–327.

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

ATKINSON, A. B. AND A. BRANDOLINI (2015): "Unveiling the Ethics behind Inequality Measurement: Dalton's Contribution to Economics," *The Economic Journal*, 125, 209–234.

BASULTO SANTOS, J. AND J. J. BUSTO GUERRERO (2010): "Gini's concentration ratio (1908-1914)," *Electronic Journal for History of Probability and Statistics*, 16, 1–42.

BEMMANN, M. (2023): "Introduction: exploring the International Statistical Institute, 1885–1938," *European Review of History: Revue européenne d'histoire*, 30, 1–12.

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

BENINI, R. (1894): "Distribuzione probabile della ricchezza privata in Italia per classi di popolazione," *La Riforma Sociale*, 1, 862–869.

BLACKORBY, C. AND D. DONALDSON (1977): "Utility vs equity: Some plausible quasi-orderings," *Journal of Public Economics*, 7, 365–381.

——— (1978): "Measures of relative equality and their meaning in terms of social welfare," *Journal of Economic Theory*, 18, 59–80.

——— (1980): "A Theoretical Treatment of Indices of Absolute Inequality," *International Economic Review*, 21, 107–136.

BOLDRINI, M. (1966): "Obituary: Corrado Gini," *Journal of the Royal Statistical Society, Series A (General)*, 129, 148–150.

BRANDOLINI, A. (2015): "Gini, his contemporaries, and the modern measurement of income inequality," *Statistics and demography, the legacy of Corrado Gini.*

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.

Clementi, F., M. Gallegati, L. Gianmoena, S. Landini, and J. E. Stiglitz (2019): "Mismeasurement of inequality: a critical reflection and new insights," *Journal of Economic Interaction and Coordination*, 14, 891–921.

Costa, R. N. and S. Pérez-Duarte (2019): "Not all inequality measures were created equal," European Central Bank, Statistics Paper Series, No. 31.

Cowell, F. A. (1980): "On the Structure of Additive Inequality Measures," *The Review of Economic Studies*, 47, 521–531.

——— (2011): *Measuring Inequality*, Oxford University Press.

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

Dagum, C. (1968): "On Methods and Purposes in Econometric Model Building," *Zeitschrift für Nationalökonomie – Journal of Economics*, 28, 381–398.

——— (1980): "The generation and distribution of income, the Lorenz curve and the Gini ratio," *Economie appliquée*, 33, 327–367.

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

Dardanoni, V. and P. J. 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.

Dasgupta, P., A. Sen, and D. Starrett (1973): "Notes on the Measurement of Inequality," *Journal of Economic Theory*, 6, 180–187.

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 Gleria, A. (1929): "Una abbreviazione nel calcolo della differenza media," *Rivista italiana di statistica*, I, 381–384.

——— (1930): "Sul calcolo della differenza media," *Rivista italiana di statistica*, II, 153–161.

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.

Derobert, L. and G. Thieriot (2003): "The Lorenz curve as an archetype: A historico-epistemological study," *The European Journal of the History of Economic Thought*, 10, 573–585.

Donaldson, D. and J. A. Weymark (1980): "A single-parameter generalization of the Gini indices of inequality," *Journal of Economic Theory*, 22, 67–86.

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

Favero, G. (2010): "Le statistiche dei salari industriali in periodo fascista," *Quaderni storici*, 319–358.

——— (2017): "A reciprocal legitimation: Corrado Gini and statistics in fascist Italy," *Management & Organizational History*, 12, 261–284.

Fisher, R. A. (1925): "Applications of "Student"'s distribution," *Metron*, 5, 90–120.

Forcina, A. (1982): "Gini's Contributions to the Theory of Inference," *International Statistical Review / Revue Internationale de Statistique*, 50, 65–70.

Forcina, A. and G. M. Giorgi (2005): "Early Gini's contributions to inequality measurement and statistical inference." *Journal Électronique d'Histoire des Probabilités 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.

Frosini, B. V. (1996): "Recenti contributi italiani alla statistica descrittiva," *Statistica Applicata*, 8, 361–396.

——— (2005): "Objective Bayesian intervals: some remarks on Gini's approach," *Metron*, 63, 435–450.

——— (2008): "On some counter-counter-examples about classical inference," *Statistica*, 68, 135–152.

Furman, E., Y. Kye, and J. Su (2019): "Computing the Gini index: A note," *Economics Letters*, 185, 1–4.

Gabbuti, G. (2020): "Between Pareto and Gini: The Origins of Personal Inequality Measurement in Italy, 1894–1939," *History of Political Economy*, 52, 435–454.

Gastwirth, J. L. (1971): "A General Definition of the Lorenz Curve," *Econometrica*, 39, 1037–1039.

——— (1972): "The Estimation of the Lorenz Curve and Gini Index," *The Review of Economics and Statistics*, 54, 306–16.

Gini, C. (1909): "Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza," *Giornale degli economisti*, 38, 27–83.

——— (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.

——— (1926): "The Contributions of Italy to Modern Statistical Methods," *Journal of the Royal Statistical Society*, 89, 703–724.

——— (1931a): "Note," *Bulletin de l'Institut International de Statistique*, 25, 317–320.

——— (1931b): "Observations à la communication du Prof. L. von Bortkiewicz," *Bulletin de l'Institut International de Statistique*, 25, 299–306.

——— (1939): "I pericoli della statistica," Discorso inaugurale letto alla prima riunione scientifica della Società Italiana di statistica (S.I.S.) tenutasi a Pisa il 9 ottobre 1939.

——— (1941): "Alle basi del metodo statistico: il principio della compensazione degli errori accidentali e la legge dei grandi numeri," *Metron*, 14, 173–240.

——— (1943): "I testi di significatività," *Atti della VII Riunione Scientifica della Società Italiana di Statistica, Roma, 27-30 giugno.*

——— (1947): "Statistical Relations and Their Inversions," *Review of the International Statistical Institute*, 15, 24–42.

——— (1949): "Concept et mesure de la probabilité," *Dialectica*, 3, 36–54.

——— (1956): "Delusioni dell'econometria," *Giornale degli Economisti e Annali di Economia*, 15, 174–177.

——— (1964): "Le applicazioni induttive del calcolo delle probabilità," *Rivista di Politica Economica*, 12, 1059–1083.

——— (1965): "On the Characteristics of Italian Statistics," *Journal of the Royal Statistical Society. Series A (General)*, 128, 89–109.

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

Gini, C. and L. Galvani (1929): "Di un'applicazione del metodo rappresentativo all'ultimo censimento italiano della popolazione (1 dicembre 1921)," *Annali di Statistica*, VI, 1–107.

Giorgi, G. M. (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.

——— (2019): "The Gini Concentration Ratio: Back to the Future," *Rivista Italiana di Economia Demografia e Statistica*, LXXIII, 5–13.

——— (2020): "Metron, a century old, but it does not show it," *Metron*, 78, 99–101.

Giorgi, G. M. and C. Gigliarano (2017): "The Gini concentration index: a review of the inference literature," *Journal of Economic Surveys*, 31, 1130–1148.

Giorgi, G. M. and S. Gubbiotti (2017): "Celebrating the Memory of Corrado Gini: a Personality Out of the Ordinary," *International Statistical Review*, 85, 325–339.

Goodman, L. and W. Kruskal (1959): "Measures of Association for Cross Classifications. II: Further Discussion and References," *Journal of the American Statistical Association*, 54, 123–163.

Gumbel, E. (1928): "Das Konzentrationsmass,," *Matematicheskii Sbornik*, 35, 65–86.

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

Herzel, A. and G. Leti (1977): "Italian Contributions to Statistical Inference," *Metron*, 35, 3–48.

Istituto Centrale di Statistica del Regno d'Italia (1929): "L'attività dell'Istituto Centrale di Statistica nel suo primo biennio di vita," Annali di statistica, Serie VI – Vol. II – Anno VII.

Jenkins, S. (1988): "Calculating Income Distribution Indices from Micro-data," *National Tax Journal*, 41, 139–142.

Jordan, W. (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. C. (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.

Kakwani, N. C. and P. J. Lambert (1998): "On measuring inequity in taxation: a new approach," *European Journal of Political Economy*, 14, 369–380.

Knibbs, G. H. (1926): "The Laws of Growth of a Population," *Journal of the American Statistical Association*, 21, 381–398.

Kolm, S.-C. (1976a): "Unequal inequalities. I," *Journal of Economic Theory*, 12, 416–442.

——— (1976b): "Unequal inequalities. II," *Journal of Economic Theory*, 13, 82–111.

——— (2011): "Inequality, new directions: full ethical foundation and consequences," *The Journal of Economic Inequality*, 9, 329–352.

Kondor, Y. (1971): "An Old-New Measure of Income Inequality," *Econometrica*, 39, 1041–1042.

Lamb, D. and S. M. Easton (1984): *Multiple Discovery: the Pattern of Scientific Progress*, Avebury, Amersham.

Lambert, P. J. (2001): *The distribution and redistribution of income*, Manchester University Press.

——— (2012): "Lidia Ceriani and Paolo Verme's paper 'The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini'," *The Journal of Economic Inequality*, 10, 419–420.

Langel, M. and Y. Tillé (2011): "Corrado Gini, a pioneer in balanced sampling and inequality theory," *Metron*, 69, 45–65.

Lerman, R. I. and S. Yitzhaki (1984): "A note on the calculation and interpretation of the Gini index," *Economics Letters*, 15, 363–368.

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

Louçã, F. (2007): *The Years of High Econometrics: A short history of the generation that reinvented economics*, Routledge.

Maccabelli, T. (2009): "Measuring Inequality: Pareto's Ambiguous Contribution," *History of Political Economy*, 41, 183–208.

Merton, R. K. (1961): "Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science," *Proceedings of the American Philosophical Society*, 105, 470–486.

Milanovic, B. (1997): "A simple way to calculate the Gini coefficient, and some implications," *Economics Letters*, 56, 45–49.

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

Monti, M., S. Pellegrino, and A. Vernizzi (2024): "The Zenga index reveals more than the Gini and the Bonferroni indexes: an analysis of distributional changes and social welfare levels," Working Paper No. 84, University of Turin, Department of Economics, Social Studies, Applied Mathematics and Statistics.

Mukhopadhyay, N. and P. Sengupta (2021): *Gini Inequality Index*, CRC Press.

Nair, U. S. (1936): "The Standard Error of Gini's Mean Difference," *Biometrika*, 28, 428–436.

Newbery, D. (1970): "A theorem on the measurement of inequality," *Journal of Economic Theory*, 2, 264–266.

Neyman, J. (1934): "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection," *Journal of the Royal Statistical Society*, 97, 558–625.

Niehans, J. (1995): "Multiple discoveries in economic theory," *The European Journal of the History of Economic Thought*, 2, 1–28.

Nixon, J. W. (1960): *A history of the International Statistical Institute, 1885-1960*, International Statistical Institute.

Ogburn, W. F. and D. Thomas (1922): "Are Inventions Inevitable? A Note on Social Evolution," *Political Science Quarterly*, 37, 83–98.

Parisi, D. (2011): "Corrado Gini (1884–1965): The Leading Figure of the Italian Group in the Econometric Society," *History of Political Economy*, 43, 211–232.

——— (2013): "A turning point in the history of Econometrics in Italy 1950. Luigi Pasinetti and Siro Lombardini, his mentor, at the Catholic University of Milan," *Il pensiero economico italiano*, 21, 47–68.

Pellegrino, S. and A. Vernizzi (2013): "On measuring violations of the progressive principle in income tax systems," *Empirical Economics*, 45, 239–245.

Piccinato, L. (2011): "Gini's criticisms to the theory of inference: a missed oppotunity," *Metron*, 69, 101–117.

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.

——— (1925): "The theory of statistical relations with special reference to cyclical series," *Metron*, IV, 383–557.

——— (1931a): "Note," *Bulletin de l'Institut International de Statistique*, 25, 320L–320M.

——— (1931b): "Observations à la communication du Prof. L.von Bortkiewicz," *Bulletin de l'Institut International de Statistique*, 25, 310.

——— (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.

——— (1939): "La società italiana di statistica," *Supplemento statistico*, 4, 1–7.

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

Pinto, P. R. and P. Paidipaty (2020): "Introduction: Measuring Matters," *History of Political Economy*, 52, 413–434.

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.

Prévost, J.-G. (2016a): "Statistical Theory, Scientific Rivalry and War Politics in Fascist Italy (1939-1943)," in *Science Policies and Twentieth-Century Dictatorships. Spain, Italy and Argentina*, ed. by A. Gómez, A. Canales, and B. Balmer (Eds.), Routledge, 159–177.

——— (2016b): "Statistics, expertise and politics: Corrado Gini and the course of Italian history," *Metron*, 74, 145–165.

Prévost, J.-G. and J.-P. Beaud (2015): *Statistics, Public Debate and the State, 1800-1945: A Social, Political and Intellectual History of Numbers*, Routledge.

Pyatt, G. (1976): "On the Interpretation and Disaggregation of Gini Coefficients," *The Economic Journal*, 86, 243.

Pyatt, G., C. nan Chen, and J. Fei (1980): "The Distribution of Income by Factor Components," *The Quarterly Journal of Economics*, 95, 451.

Rao, V. M. (1969): "Two Decompositions of Concentration Ratio," *Journal of the Royal Statistical Society. Series A (General)*, 132, 418.

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

Rogerson, P. (2015): "Corrado Gini (1884-1965): Versatile Originator of Measures of Variability," in *Great Minds in Regional Science, Vol. 2*, ed. by P. Batey and D. Plane, Springer, chap. 4, 109–124.

Rosenbluth, G. (1951): "Note on Mr. Schutz's Measure of Income Inequality," *The American Economic Review*, 41, 935–937.

Schechtman, E. and R. Zitikis (2006): "Gini indices as areas and covariances: what is the difference between the two representations?" *Metron*, 64, 385–497.

Schneider, M. (2021): "The Discovery of the Gini Coefficient: Was the Lorenz Curve the Catalyst?" *History of Political Economy*, 53, 115–141.

Schutz, R. (1951): "On the Measurement of Income Inequality," *The American Economic Review*, 41, 107–122.

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

Sen, A. (1973): "Poverty, Inequality and Unemployment: Some Conceptual Issues in Measurement," *Economic and Political Weekly*, 8, 1457–1464.

——— (1974a): "Informational bases of alternative welfare approaches," *Journal of Public Economics*, 3, 387–403.

——— (1974b): "Informational bases of alternative welfare approaches: aggregation and income distribution," *Journal of Public Economics*, 3, 387–403.

——— (1976a): "Poverty: An Ordinal Approach to Measurement," *Econometrica*, 44, 219–231.

——— (1976b): "Real National Income," *The Review of Economic Studies*, 43, 19–39.

——— (1978): "Ethical measurement of inequality: some difficulties," in *Personal Income Distribution*, ed. by W. Krelle and A. F. E. Shorrocks, North-Holland, Amsterdam., 416–431.

——— (1979): "The Welfare Basis of Real Income Comparisons: A Survey," *Journal of Economic Literature*, 17, 1–45.

Sheshinski, E. (1972): "Relation between a social welfare function and the gini index of income inequality," *Journal of Economic Theory*, 4, 98–100.

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.

Stuart, A. (1954): "The correlation between variate-values and ranks in samples from a continuos distribution," *British Journal of Statistical Psychology*, 7, 37–44.

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

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

von Bortkiewicz, L. (1931a): "Die Disparitätsmasse der Einkommensstatistiks," *Bulletin de l'Institut International de Statistique*, 25, 189–298.

——— (1931b): "Erwiderung," *Bulletin de l'Institut International de Statistique*, 25, 311–316.

Weymark, J. A. (1981): "Generalized Gini inequality indices," *Mathematical Social Sciences*, 1, 409–430.

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

Yitzhaki, S. (1998): "More than a Dozen Alternative Ways of Spelling Gini," *Research in Economic Inequality*, 8, 13–30.

——— (2003): "Gini's Mean difference: a superior measure of variability for non-normal distributions," *Metron*, LXI, 285–316.

Yitzhaki, S. and E. Schechtman (2005): "The properties of the extended Gini measures of variability and inequality," *Metron*, 63, 401–433.

——— (2013a): *The Gini Methodology. A Primer on a Statistical Methodology*, Springer New York.

——— (2013b): "More Than a Dozen Alternative Ways of Spelling Gini," in *The Gini Methodology. A Primer on a Statistical Methodology*, Springer New York, 11–31.

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

Zanella, A. and G. Leti (1995): "Italian Society of Statistics," *Alma Mater Studiorum - Università di Bologna*.