Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 6, n°1; Juin/June 2010

www.jehps.net

GINI's CONCENTRATION RATIO (1908-1914)

Jesús Basulto Santos and J. Javier Busto Guerrero[1]

Abstract

The books and research papers that Corrado Gini published from 1908 to 1914 about measures of concentration and dispersion are: the paper "*Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza*" in 1909, of which Gini presented a summary at the second meeting of the Italian Society for the Progress of Sciences (ISPS), Florence, 1908; the book "*Indice di Concentrazione e di Dipendenza*" in 1910, from which Gini presented a paper at the third meeting of the ISPS, Padua, 1909, and that was to be published by the above-mentioned Society in 1910; the book "*Variabilità e Mutabilità: contributo allo Studio delle distribuzioni e delle relazioni statistiche*" in 1912, and, finally, the research paper "*Sulla misura della concentrazione e della variabilità dei caratteri*" in 1914.
Over the period 1908-1912, Gini proposed two functional measures of concentration and several numerical measures of concentration. It was in 1914 that Gini proposed his R ratio of concentration, which is applicable to a distribution function of a nonnegative random variable with a finite expected value. Gini carried out his concentration measures as a review to the functional criterion that Vilfredo Pareto had proposed to measure the concentration of the income distribution. While Gini was comparing Lorenz curves to measure the concentration of the income distribution, Pareto compared distribution functions or generalized Lorenz curves. This would explain the disagreement that Gini had with Pareto about the behaviour of the parameter α in Pareto's income distribution model.

Resumen

Los libros y trabajos de investigación que Corrado Gini publicó sobre medidas de concentración y dispersión, desde 1908 hasta 1914 son: el artículo de 1909, "*Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza*", del que Gini había presentado un resumen en la II Reunión de la Sociedad Italiana para el Progreso de las Ciencias (SIPS) en Firenze en 1908; el trabajo del 1910, "*Índice di Concentrazione e di Dipendenza*", del que presentó una comunicación en la III Reunión de SIPS en Padova en 1909 y que sería publicada por dicha Sociedad en 1910; el libro de 1912, "*Variabilità e Mutabilità: contributo allo Studio delle distribuzioni e delle relazioni statistiche*" y, finalmente, el trabajo de investigación en 1914, "*Sulla misura della concentrazione e della variabilità dei caratteri*".
Durante los años 1908-1912, Gini propuso dos medidas funcionales de concentración y, por su interés práctico, varias medidas numéricas de concentración. Será en 1914, cuando Gini proponga su razón R de concentración que es aplicable a una función de distribución de una variable aleatoria no negativa con esperanza matemática finita. Gini desarrolló sus medidas de concentración como una critica al criterio funcional que Wilfredo Pareto había propuesto para medir la concentración de la distribución de la renta. Si Gini comparaba curvas de Lorenz para medir la concentración, Pareto comparaba funciones de distribución o, su implicación, curvas de Lorenz generalizadas, lo que explicaría el desacuerdo que Gini mantenía sobre el parámetro α del modelo de distribución de la renta de Pareto.

[1] Universidad de Sevilla, Facultad de Ciencias Económicas y Empresariales, Avenida de Ramón y Cajal nº 1; 41018 Sevilla, Spain. basulto@us.es; jjbusto@us.es.

1. Introduction

Between 1908 and 1913, Gini proposed several criteria and concentration indexes before obtaining his R concentration ratio in 1914.

It was in 1908 that Gini published his thesis "*Il sesso dal punto di vista statistico*" and that he presented his paper "*Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza*" in the second meeting of the Italian Society for the Progress of Sciences, in Florence. This paper was published in 1909 in the Giornale degli Economisti.

In the second part of this paper, Gini proposed an inequality criterion different from the Pareto criterion of "*diminution de l'inégalité des revenus*" and his index of concentration δ . This δ - index was obtained by Gini from Pareto's income distribution model.

In 1909 Gini presented at the third meeting of of the Italian Society for the Progress of Sciences in Padua, the paper "*Indice di Concentrazione e di Dipendenza*", which was pusblished by the above-mentioned Society in 1910.

In the part of this paper called "*Indice di Concentrazione*", Gini proposed a new inequality criterion more general than the one of 1908. He related this new criterion with his δ - index and proposed what he called the mean concentration index. Gini related his δ - index with Pareto's parameter α and he observed that the harmonic mean of ʃ and α was equal to one for $\alpha > 1$. From this final result, Gini affirmed, against Pareto, that a decrease of α implied an increase of the inequality, whenever the inequality was measured by the δ - index.

Gini's book, "*Variabilità e Mutabilità: contributo allo Studio delle distribuzioni e delle relazioni statistiche*" was published in 1912 by the Università de Cagliari.

Gini devoted the greatest part of his 1912 book to study a dispersion measure called mean difference which nowadays is called mean difference of Gini. The book gathered several formulas of this dispersion measure, including one for continuous variables. Gini proved that when the parameter α of Pareto decreased, then the relativized mean difference increased and the other way round. This led Gini to demonstrate the erroneous interpretation that Pareto made with the parameter α . Another result was to prove in a rigorous way that if supposing the incomes follow the first Model of Pareto, then the harmonic mean of the parameters α and δ is equal to one for $\alpha > 1$. In consequence, Gini admitted that his ʃ-index was dependent on the Model of Pareto. This led Gini to begin a new research in order to find an inequality measure valid to a set of distribution functions.

Gini reached his goal in 1914 with his reserch paper "*Sulla misura della concentrazione e della variabilità dei caratteri*" which was published in the *Atti del R. Istituto Veneto di Science, Lettere ed Arti*, in 1913-1914.

Until 1913 Gini carried out his research spurred by Pareto's ideas. But the restriction of his ʃ-index led him to the article that M. O. Lorenz published in 1905 under the title

"*Methods of measuring the concetration of Wealth*". This paper led Gini to propose the double of the concentration area (area between the Lorenz curve and the equality line) as a concentration measure that Gini named R[2] concentration ratio.

The remainder of this article is organized as follows: we have gathered in section 2 a summary of the research that Pareto studied about the income distribution and his income inequality criterion; the research of Gini during the period from 1908 to 1910 will be studied in section 3; the book that Gini published in 1912 in section 4 and finally in section 5 we will analyze the research that Gini did in 1914.

2. Pareto's law of income distribution and income inequality

2.1 Pareto type I of income distribution

Wilfredo Pareto (1848-1923) began his statistic investigations on income distribution in 1893 when he was in Lausanne. Towards the end of the 19th century, some agencies in England and other industrialized countries began releasing income distribution statistics by giving the numbers of taxpayers in different income brackets. A discussion of this data by Paul Leroy-Beaulieu (1881) greatly influenced Pareto. During three years, Pareto studied the income statistics of several countries such as France, England, Belgium, Germany, Switzerland, Austria and the United States of America over different periods of time.

To facilitate comparisons among societies with different population sizes and currencies, Pareto applied an interpolation method to his income tax data, from which he proposed "a simple enough empirical law" which seemed to explain the phenomenon of the income distribution. The results of his research first appeared in [Pareto, 1895], and, two extensions of Pareto's law appeared in [Pareto, 1896a].

The articles cited above and other research papers by Pareto on the income distribution were published in [Busino,1965].

In this section we are interested in Pareto's law type I and his relation with the delta-index of Gini.

Having defined x as a given income, and N_x as the number of taxpayers with personal income greater or equal to x, if one plots the logarithm of N_x against the logarithm of x in a Cartesian coordinate system, the points $(\log x, \log N_x)$ approximately trace a straight line with negative slope. Empirical linearity of the $(\log x, \log N_x)$ relationship corresponds to a relationship between N_x and x given by

$$\log N_x = \log A - \alpha \log x ,$$

or equivalently

[2] R is the ratio between the concentration area (area between the Lorenz curve and the equality line) and the maximum value of the concentration area. When n is large then the maximum of the concentration area tends to 0.5.

$$N_x = \frac{A}{x^{\alpha}}, \tag{2.1}$$

where the slope α and the intercept $\log A$ can be estimated from the distribution data by the Method of Least Squares. The equation (2.1) defines the Pareto type I.

If you suppose that the income is continuous and non-negative, the minimum income h is a positive parameter and the maximum income is infinite, then the continuous approximation of the equation (2.1) corresponds to a random variable X with a cumulative distribution function $F(x)$ defined by

$$F(x) = \begin{cases} 1 - \left(\frac{h}{x}\right)^{\alpha} & x \geq h > 0 \\ 0 \ en[0,h) \end{cases}, \tag{2.2}$$

where $\alpha > 0$[3], $A = N_h h^{\alpha}$, $\frac{N_x}{N_h} = 1 - F(x)$, h is the minimum income and N_h is the total of taxpayers. The expression (2.2) is the Pareto Model I.

The fact that empirically the values of parameter α remain stable (with Pareto's data, the estimated value of α varies relatively little between the minimum, 1.13, and the maximum, 1.89, with mean=1.51 and CV=2.34%), led Pareto to formulate the following statement

"*Ces résultats sont très remarquables. Il est absolument impossible d'admettre qu'ils sont dus seulement au hasard. Il y a bien certainement une cause qui produit la tendance des revenus à se disposer suivant une certaine courbe. La forme de cette courbe paraît ne dépendre que faiblement des différentes conditions économiques des pays considérés, puisque les effets sont à peu près les mêmes pour des pays dont les conditions économiques sont aussi différentes que celles de l'Angleterre, de l'Irlande, de l'Allemagne, des villes italiennes, et même du Pérou*". (§959 *Cours*, 1897, Tomo II, p. 312).

Thus Pareto can conclude

"*Enfin, si la répartition de la richesse varie peu pour des contrées, des époques, des organisations différentes, il nous faudra conclure que, sans vouloir négliger les autres causes, nous devons chercher dans la nature de l'homme la cause principale qui détermine le phénomène".* (§957 *Cours*, 1897, Tomo II, p. 304).

Pareto always insisted that his formula for the distribution of income was only a first approximation; that the fit was good only for the right tail of the distribution -the only part from which he had solid empirical evidence- and even then not perfect.

In his Manuel of 1909, p.391, Pareto wrote:

[3] When Pareto and Gini applied the Model I to the income data, they found that $\alpha > 1$, and therefore the mean income was finite for this Model.

"*On a voulu en tirer une loi générale, d'après laquelle l'inégalité des revenus devait continuer à diminuer. Cette conclusion dépasse de beaucoup ce qu'on peut tirer des prémisses. Les lois empiriques, comme celle-ci, n'ont que peu de valeur, ou même n'en ont aucune, en dehors des limites dans lesquelles elles ont été reconnues vraies*".

2.2 Pareto's inequality criterion

In this section we are interested in the definition and the mathematical formalization of Pareto's criterion of income inequality, which is gathered in his *Cours* [1897, volume II, §964-§965].

The first definition of inequality:

"*La diminuion de cette inégalite sera donc définie par le fait que le nombre des pauvres va en diminuant par rapport au nombre des riches ou, ce qui est la même chose, par rapport au nombre total des membres de la société. C'est le sens qui paraît avoir prévalu, et c'est donc celui que nous adopterons*".

Pareto illustrated this first definition with an example in his *Manuel d'économie politique* [1909,VII,24,p.389]. Pareto considered a first population (A) with ten people, where nine people had 1.000 francs each (poor people), and only one had 10.000 francs, which was considered rich; next he considered a second population (B) with ten people, where eight people that were poor in (A) now have 10.000 francs each and the others two people one were poor and the other one rich in (A), the other two people from (A) that go to (B) remain the same (one rich and one poor), being nine people rich and only person poor in (B). When passing from (A) to (B) it can be observed than the poor people had decreased compared to the rich ones. So, according to the first definition of inequality, the population (B) has less inequality than the population (A) because (B) has fewer poor people than (A).

Four lines futher down, Pareto proposed a second definition:

"*En général, lorque le nombre des personnes ayant un revenu inférieur à* x *augmente*[4] *par rapport au nombre des personnes ayant un revenu supérieur à* x*, nous dirons que l'inégalité des revenus diminue*".

At the bottom of page 320, Pareto formalized this second definition.

Pareto first defined the function u_x by

$$u_x = \frac{N_x}{N_h}, \tag{2.3}$$

which calculates the proportion of individuals whose incomes are greater than x. Then, the second definition Pareto's inequality is equivalent to: the income inequality decreases when u_x increases, for all $x > h$.

[4] The text of the Cours reads "increase" instead of "decreases".

Another expression of (2.3) is the relation $u_x = 1 - F(x)$, where $F(x)$ is the cumulative distribution function, and u_x is the survival function. A mathematical formalization of the second definition of inequality is: $F_2(x)$ <u>is less inequal than</u> $F_1(x)$ when

$$F_2(x) \leq F_1(x)^5, \tag{2.4}$$

for all $x \geq 0$.

The expression "diminution de l'inégalité des revenus" was criticized by subsequent authors as being "erroneous", instead of simply accepting it as a definition. Pareto, in response to such reactions, changed that terminology by "diminution de l'inégalité de la proportion des revenus"in his Manuel [VII, 24, p.389].

Bortkiewicz showed that if (2.4) was true, then $\mu_2 \geq \mu_1$, where μ_k was the mean income of $F_k(x)$, for k=1,2. When $F_1(x) \neq F_2(x)$, then $\mu_2 > \mu_1$.

The consequence that the mean income increases when the income inequality decreases must be understood in the context of the political position of Pareto. Pareto declared that "*Pour amener une répartition plus favorable aux pauvres, il n'y a qu'un moyen: améliorer la production et, por là, faire croître la richesse plus vite que ne croît la population*" [Pareto, 1896b]. In the above example of Pareto, the difference between the income of (B) and income of (A), 72.000 francs, has been transferred to eight poor people from (A) so that they become rich in (B).

In relation to (2.4), it can be shown that: if $F_2(x) \leq F_1(x)$ for all x, then $G_2(p) \geq G_1(p)$ for all p in $[0,1]$, where $G_k(p)$ is the generalized Lorenz curve[6] of $F_k(x)$, for k=1,2. Also, if $G_2(p) \geq G_1(p)$, for all p in $[0,1]$, then $\mu_2(1-R_2) \geq \mu_1(1-R_1)^7$, where R_k is the Gini concentration ratio, for k=1,2. The formula $\mu\,(1-R)$, which combines growth and inequality, is interpreted as "the mean income, μ, modified downward by the Gini inequality R. The formula $\mu(1-R\,)$ fits as an intuitive and usable welfare indicator" [Sen, p.137].

2.3 Examples in Pareto Model I

In this section we apply the criterion of inequality of Pareto to Model I.

Example I

Models	α	h	μ	R of Gini	$\mu(1-R)$
A	1,8	24	54	0,38	33,2

[5] This inequality was obtained by [Bortkiewicz, 1931, p. 221-222]. This criterion of income inequality of Pareto is called today First Order Stochastic Dominance [Basulto et al, 2009].

[6] The generalized Lorenz curve is equal to $G(p) = \mu L(p)$, where μ is the mean income and $L(p)$ is the Lorenz curve.

[7] Sen proposed this criterion in 1976. The value $\mu\,(1-R)$ is twice the area below of the generalized Lorenz curve. [Sen and Foster, 1997, p. 137].

B	1,6	34	90.6	0,45	49,4

In this example α decreases and h increases, which is equivalent to saying that $F_B(x) \le F_A(x)$ for all $x \ge 0$, therefore we can apply the criterion of inequality of Pareto. The main consequences are: $\mu_B > \mu_A$; $\frac{x_B(p)}{h_B} \ge \frac{x_A(p)}{h_A}$, i.e., the ratio between the quantile function, $x(p)$, and the minimum income, h, increases for $0 < p < 1$; $G_B(p) \ge G_A(p)$, for $0 < p < 1$; $\mu_B(x_B(p)) \ge \mu_A(x_A(p))$, i.e., the mean income of the incomes below $x(p)$ increases for $0 < p < 1$; $M_B(x_B(p)) \ge M_A(x_A(p))$, i.e., the mean income of the incomes above $x(p)$ increases for $0 < p < 1$; $R_B > R_A$ and $\mu_B(1 - R_B) > \mu_A(1 - R_A)$.

Example II

Models	α	h	μ	R of Gini	$\mu(1-R)$
A	1,3	20	86,6	0,62	32,4
B	1,5	39	117	0,5	58,5

In this example α increases and μ increases. Now $F_B(x) < F_A(x)$ for $x < 2{,}994.26$ and $F_B(x) > F_A(x)$ for $x > 2{,}994.26$, being $x_c = 2{,}994.26$ the intersection point of the cumulative distribution functions. In this example we can't apply the criterion of inequality of Pareto. As $1 - F_A(x_c) = 1 - F_B(x_c) = 0.00148$, i.e., the percentage of population with income above $x_c = 2{,}994.26$ is 0.148%, then we can approximately apply the criterion of inequality of Pareto[8]. The main consequences are: $h_B \ge h_A$; $\frac{x_B(p)}{h_B} \le \frac{x_A(p)}{h_A}$, for $0 < p < 1$; $G_B(p) \ge G_A(p)$ for $0 < p < 1$; $\mu_B(x_B(p)) \ge \mu_A(x_A(p))$, for $0 < p < 1$; $R_B < R_A$ and $\mu_B(1 - R_B) > \mu_A(1 - R_A)$. Now, the curves $M_A(x_A(p))$ and $M_B(x_B(p))$ intersect in the value of $p_c = 0{,}946$, where $x_A(p_c) = 190.04$ and $x_B(p_c) = 274.47$. Being $M_B(x_B(p)) > M_A(x_A(p))$ for $p < 0{,}946$ and $M_B(x_B(p)) < M_A(x_A(p))$ for $p > 0{,}946$.

Exampe III

Models	α	h	μ	R of Gini	$\mu(1-R)$
A	4/3	20	80	0,6	32
B	2.0	40	80	0,33	53,3

In this example α increases and $\mu_A = \mu_B$. Now $F_B(x) < F_A(x)$ for $x < 160$ and $F_B(x) > F_A(x)$ for $x > 2{,}994.26$, being $x_c = 160$ the intersection point of the cumulative distribution functions. In this example we can't apply the criterion of inequality of Pareto. The main consequences are the same as in example III, except that now $M_B(x_B(p)) \le M_A(x_A(p))$ for $0 \le p \le 1$.

[8] This type of approximation was used by Pareto in his *Cours* [1897, §965, p. 320-326].

Example IV

Models	α	h	μ	R of Gini	$\mu(1-R)$
A	1,3	20	86,6	0,62	32,4
B	1,5	24	72	0,5	36

In this example α increases, h increases and μ decreases. Now $F_B(x) < F_A(x)$ for $x < 78.5$ and $F_B(x) > F_A(x)$ for $x > 78.5$, being $x_c = 78.5$ the intersection point of the cumulative distribution functions. In this example we can't apply the criterion of inequality of Pareto. Now, $G_B(p) \geq G_A(p)$ for $p < 0{,}9886$ and $G_B(p) \leq G_A(p)$ for $p > 0{,}9886$. The main consequences are: $\frac{x_B(p)}{h_B} \leq \frac{x_A(p)}{h_A}$ for $0 < p < 1$; $M_B(x_B(p)) \leq M_B(x_B(p))$ for $0 < p < 1$ and $R_B < R_A$. The measure statistics $\mu(1-R)$ increases in this example, but if $\alpha = 1{,}9$ and $h = 23$ in Model B, then $\mu(1-R)$ decreases. Also, $\mu_B(x_B(p)) > \mu_A(x_A(p))$ for $p < 0{,}9886$ and $\mu_B(x_B(p)) < \mu_A(x_A(p))$ for $p > 0{,}9886$.

Next we sum up the results of the Model I of Pareto[9].

(1) If $\alpha_B \leq \alpha_A$ and $h_B \geq h_A$[10], with some strict inequality, then: $\mu_B \geq \mu_A$, $\frac{x_B(p)}{h_B} \geq \frac{x_A(p)}{h_A}$, $M_B(x_B(p)) \geq M_A(x_A(p))$, $\mu_B(x_B(p)) \geq \mu_A(x_A(p))$ and $G_B(p) \geq G_A(p)$, for $0 \leq p \leq 1$. Also, $R_B \geq R_A$ and $\mu_B(1-R_B) \geq \mu_A(1-R_A)$.

(2) If $\alpha_B \geq \alpha_A$ and $\mu_B \geq \mu_A$[11], with some strict inequality, then: $h_B \geq h_A$, $\frac{x_B(p)}{h_B} \leq \frac{x_A(p)}{h_A}$, $\mu_2(x_2(p)) \geq \mu_1(x_1(p))$ and $G_B(p) \geq G_A(p)$, for $0 \leq p \leq 1$. Also, $R_B \leq R_A$ and $\mu_B(1-R_B) \geq \mu_A(1-R_A)$. If $\mu_A = \mu_B$, then $M_B(x_B(p)) \leq M_A(x_A(p))$ for $0 \leq p \leq 1$.

(3) If $\alpha_B \geq \alpha_A$ and $\mu_B \leq \mu_A$, with some strict inequality, then $\frac{x_B(p)}{h_B} \leq \frac{x_A(p)}{h_A}$ and $M_B(x_B(p)) \leq M_A(x_A(p))$ for $0 \leq p \leq 1$. Also, $R_B \leq R_A$.

3. Inequality criterion and concentration index of Gini

In this section we have proposed to study the second part of the paper "*Il diverso accrescimiento delle classi sociali e la concentrazione della reccezza*[12]" of 1909 and

---

[9] The consequences: of (1), (2) and (3) are necessary conditions.
[10] This hypothesis is equivalently to Pareto's inequality criterion.
[11] This hypothesis is a weakness of Pareto's inequality criterion. This is the case of singly intersecting of cumulative distribution functions, where $F_B(x) \leq F_A(x)$ for $x < x_c$ and $F_B(x) > F_A(x)$ for $x > x_c$, and $\mu_B > \mu_A$, where the point of intersect is x_c. *Cours* [1897, §965, p. 323-324].
[12] A study on the first part of this paper can be seen in [G. Levi Della Vida; La teoria Della circolazione delle aristocrazie del Pareto e la teoria del recambio sociale del Gini; 1935; Comitato Italiano per lo Studio dei Problemi Della Popolazione].

the part on "*Concentrazione*" of the paper "*Indici di Concentrazione e di Dependeza*" of 1910.

3.1 The inequality criterion of Gini

Two inequality criteria can be distinguished to measure the concentration of the income of Gini (1909,XV, p. 69):

(1) "*Some authors supposed that the income distribution was as unequal as the number of rich people decreased with respect to the poor people. These authors have probably based their analysis on the fact that the wealth is deeply felt when only a few people are fortunate." With this, Gini was referring to the first inequality definition of Pareto that we gathered in section* **2.2**[13].

(2) "*On the contrary, other authors said that the wealth was as unequal as the number of rich people was* greater wit *respect to the poor people. The nearby condition to an absolute equality would be the population that had only one rich person and the remaining people had the same incomes*"

Both criteria seem fallible because they consider just one income distribution factor (the population factor), i.e., the number of rich or poor people, or i.e. the total wealth or incomes; both never at the same time. So Gini considered that a good definition for the inequality or the concentration must take into account both factors[14]: the relation between rich and poor people and the relation between the total income of the rich people and the total income of the poor people. Considering this, Gini proposed the following definition of wealth concentration (Gini, 1909, pp.69-70):

"*Pare a me che la concentrazione della ricchezza debba dirsi, in un paese o in un'epoca A, maggiore che in un paese o in un'epoca B, quando la parte della popolazione che possiede una parte* $=\frac{1}{x}$ *della ricchezza nazionale sia in A minore che in B o, viceversa, quando la parte della ricchezza posseduta da un aparte* $=\frac{1}{y}$ *della popolazione sia in B minore che in A*"

That is to say that if we call p_A a proportion or a part of the population of A and q_A its corresponding income proportion, and if we also call p_B a proportion or a part of the population B and q_B its corresponding income proportion, then we'll be able to affirm that the income concentration in A is greater than in B when the two following conditions are fulfilled:

(1) If $q_A = q_B = \frac{1}{x}$, then $p_A < p_B$,

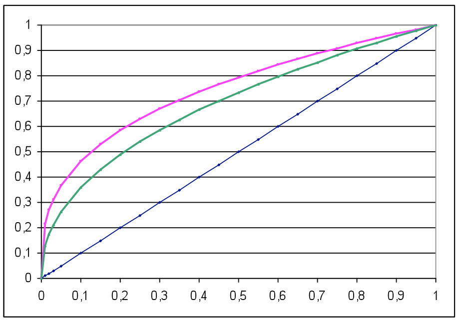
[13] Pareto affirmed that the inequality decreased when the proportion of poor people with respect to the rich people decreased. In consequence, when the proportion of poor people increase with respect to the rich people, the inequality must increase too as Gini affirmed.

[14] M.O. Lorenz affirmed: "It is apparent that we need to take into account simultaneously the changes in wealth and the changes in population [Lorenz, 1905, p.213].

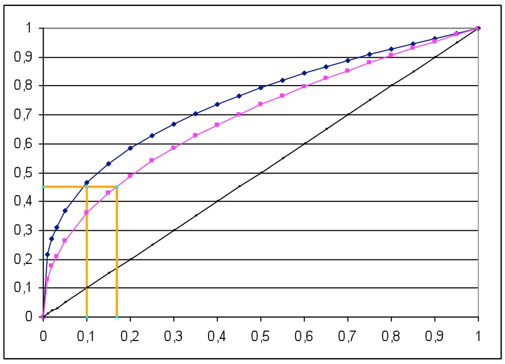
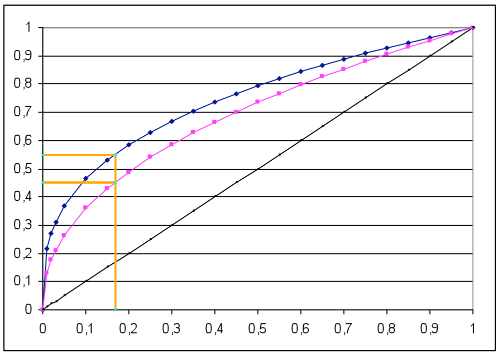
and

(2) If $p_A = p_B = \frac{1}{y}$, then $q_A > q_B$.

To interpret this inequality criterion of Gini we have supposed that we have ordered the income of the people of each population from the lower to the greater value and next we have calculated the accumulated proportion of the incomes and the population by accumulating the people from the richest to the poorest[15] ones. For example, p_A could be the 3% of the rich people and q_A the 5% of the total income corresponding to the 3% of the rich people. With this, Gini compared the following two continuous dual[16] Lorenz curves



and, according to the definition, they shouldn't cut. The following examples illustrate the definition of the concentration of Gini. The upper curve is *A* and the one below is *B*.

(1) A value of q = 0.45, corresponds to two values of p, as, p1=0,1 and p2=0,17	(2) A value of p=0,17, corresponds to two values of q, as, q1=0,55 and q2=0,45
	

---

[15] He did this clearly in his book in 1910.

[16] The dual Lorenz curve is symmetrical to the Lorenz curve when we consider $\left(\frac{1}{2},\frac{1}{2}\right)$ as the symmetrical point of the square unit. Also, if $L(p)$ is the Lorenz curve, then the dual Lorenz curve is $L^*(p) = 1 - L(1-p)$, for $0 \le p \le 1$.

In the first example we have considered the 45% of the income in A and B (axis of ordinate) and it can be observed that the curve with more concentration has less population. In the second example we have considered the 17% of the population in A and B (axis of abscissa) and we can observe that the curve with more concentration has more part of income.

If we hadn't taken the proportions with the income order showed, the definition of Gini would have provoked some contraindications, as showed in the following example: $A=\{0,0,8,12\}$ and $B=\{1,4,5,10\}$, where $p_A(0,8)=p_B(4,5)=\frac{2}{5}$, but $q_A(0,8)=\frac{8}{20}$ is lower than $q_B(4,5)=\frac{9}{20}$, which is against the previous condition (2) of Gini[17].

Although Gini didn't used in his inequality criterion that the mean income could increase, we have gathered some comments that Gini showed about this question.

"*Ognuno si domanderà a questo punto: La progresiva concentrazione dei reddito e di patrimoni rappresenta un pericolo sociale ? Non è, credo, una domanda, a cui si possa dare una risposta pacifica. Certo, la risposta dovrebbe seere afermativa, qualora la ricchezza media del paese restasse costante, poichè il concentrarsi della ricchezza acuirebbe allora il malcontento dei poveri, che vedrebbero diminuito assolutamente il loro avere, e renderebbe eccessiva la potenza dei ricchi. Ma la risposta diviene dubbia quando, come è il caso normales, la ricchezza media del paese aumenta. Non vi ha dubbio infatti, che la disuguaglianza della ricchezza è meno sentita quando più è alta la ricchezza media. E potrebbe anche darsi (solo estesissime ricerche potrebbero risolvere la questione) che la concentrazione della ricchezza rappresentasse nella evoluzione della socitè un fenomeno natural parallelo all'aumentare della ricchezza media, nello steso modo che, nella scala biologica, la supremazia del sistema nervoso diviene tanto più spiccata, quanto più l'organismo è vasto e complesso*" [1909, 81-82].

In this text from section XXI at the end of the paper, Gini gathered his ideas about the increase of the mean income, where he showed that the inequality is "*meno sentita quando più è alta la ricchezza media*" and, also, that it must increase its concentration. On the contrary, Gini didn't agree with not changing the mean income because it could exaggerate "*la potenza dei ricchi*".

3.2 The concentration index δ

From the concentration criterion of section **3.2**, Gini showed that,
"*Resta a determinare, mediante un indice appropriato, la relacizione tra* $\frac{1}{x}$ *e* $\frac{1}{y}$".
That means that for Gini, a concentration index must connect the income proportions and its corresponding population proportions.

In section XVI of his paper in 1909 Gini gathered the construction of the delta-index δ.

[17] To prevent this problem, we should choose in the whole population, any part with a proportion of people p that had the greater proportion of income q.

With the notation of section **2.1**, Gini considered the following Model I of Pareto

$$N_x = \frac{K}{x^{\alpha}}, \tag{3.1}$$

where N_x are the total of taxpayer whose incomes are greater or equal to the income x; being K and α parameters. When $x = h$, where h is the minimum income, then N_h will be the total of taxpayers, so formula (3.1) is written as

$$\frac{N_x}{N_h} = \left(\frac{h}{x}\right)^{\alpha}. \tag{3.2}$$

Next, Gini proposed the following model

$$A_x = \frac{L}{x^{\beta}}\ ^{18}, \tag{3.3}$$

where A_x is the proportion of income for the people with incomes greater or equal to the income x, the parameters L and β must be estimated by logarithms A_x and x. This generates a line as in the Model I of Pareto. Applying (3.3) to the minimum income h, the following is obtained

$$\frac{A_x}{A_h} = \left(\frac{h}{x}\right)^{\beta}. \tag{3.4}$$

Now, as we have gathered in the beginning of this section, we have connected the proportion of income of (3.4) with the corresponding proportion of population of (3.2). This operation leads us to the expression

$$\frac{N_x}{N_h} = \left(\frac{A_x}{A_h}\right)^{\delta},\ ^{19} \tag{3.5}$$

where the parameter δ is equal to $\delta = \frac{\alpha}{\beta}$.

As the proportion of incomes and population are calculated from the richest to the poorest people, it's easy to deduce the following equality

$$\frac{N_x}{N_h} < \frac{A_x}{A_h},$$

[18] If we take logarithms in both parts of the equation we'll obtain a linear model in the parameters L and β. Gini adjusted this model by the method of Cauchy, as"*non ci pare che valga la pena di ricorrere a metodi di interpolazione più complidati (metodo dei momenti, metodo dei minimi quadrati)*". In a paper of [Pollastri, 1990], the author concluded that:: (1) The method of Cauchy obtained better results than in the method of least squares, (2) The model of Pareto got more efficient estimated parameters and good adjusts of the proportion of population. Gini only got good adjusts in the proportion of the income. The author affirmed that the comparison of both models depended on the criteria used of valuation.

[19] This result justified the interpretation we've made of the concentration criterion of Gini.

for every income x, except when $x = h$ which an inequality. So, if we want the identity (3.5) to be true, the parameter δ must verify that $\delta \geq 1$. The value of $\delta = 1$ only will be true when the proportions of the incomes are equal to the corresponding proportions of populations, for every income x. In this situation we'll have the maximum equality or minimum concentration.
If the proportion of income increases in proportion of the population, the parameter δ should increase in order to fulfil the identity (3.5). This will increase the concentration.
The parameter δ was the first concentration index that was proposed by Gini. If we express formula (3.5) depending on the proportion of the income

$$\frac{A_x}{A_h} = \left(\frac{N_x}{N_h}\right)^{\frac{1}{\delta}}, \tag{3.6}$$

then, formula (3.6) defines the dual Lorenz of the Model I of Pareto[20], that has been used to interpret the inequality or concentration definition proposed by Gini[21].

To calculate δ, Gini did two adjusts, the Model of Pareto that gave him an estimation of α and the one based in the formula (3.3) of Gini that gave him an estimation of β. This let him estimate δ by the quotient of the estimations of α and β.

Gini applied this index to several countries and over different periods of time in sections XVIII and XIX of his paper, let's see some of these consequences:

"*i dati precedenti relativi a redditi, patrimonî censito e patrimonî ereditarî ci mostrano como la ricchezza sia andata concentrandosi in Inghilterra, Prussia, Amburgo, Sassonia, Norvegia, Massachussetts. Un netto processo di concentrazione non si manifesta invece in Austria. Sarà interessante estendere la ricerca ad altri paesi e ad altri tempi. Ma, anche entro questi limiti, i resultati ottenuti ci sembrano di una certa importanza. Finora infatti le opinioni sulla concentrazione della ricchezza attraverso il tempo erano molte divise. Il Pareto, fondandosi sulla costanza approssimativa del coefficiente α, riteneva che la distribuzione della ricchezza fosse costante. Altri invece (Warner) in base ai dati delle statistiche prussiane, sostiene che la ricchezza si concentra; e ad altri infine (Giffen, Huncke) pareva di poter dedurre, dalle statistiche inglesi, che la distribuzione della ricchezza va facendosi più uguale.*
Le conclusioni sono diverse quando i dati vengono elaborati matematicamente con un metodo che tiene conto, non solo del numero dei censiti nelle varie classi di reddito o di patrimonio, ma anche dell'ammontare del loro reddito o del loro patromonio".

This is a criticism of the result that Pareto defended, that was that the parameter α remained nearly constant in any country or period of time. At the end of this text, Gini reminds again that the inequality must not only consider the number of people for every type of income but also the amount of income of the people.

[20] We know that in Model I of Pareto the dual Lorenz curves don't get crossed, so that the concentration index δ is compatible with Gini's general definition of concentration.
[21] Gini didn't see in that moment that the dual Lorenz curve (10) was related with the Model I of Pareto. An explanation to this could be ought due to the process that Gini used to get that curve, using first the Model of Pareto and next his Model (3.3).

Gini reviewed that Pareto used only the population factor in the inequality criterion this means that Pareto didn't take into account the total of individuals for each one of the incomes but didn't take into account the corresponding amounts of income. We have seen in section **2.2** that the inequality criterion of Pareto compared generalized Lorenz curves. This is the same that comparing for each proportion of population[22], p, the mean incomes that were below the quartile $x(p)$.

3.3 *Indici di Concentrazione e di Dipendenza*

This paper was presented by Gini at the third meeting of the Italian Society for the Progress of Sciences, Padua, 1909, and it was published by the above-mentioned Society in 1910. We are interested in the part called "*Indici di Concentrazione*" of this paper of Gini.

In this paper, Gini introduced a new inequality criterion that he connected to his δ index of concentration of section **3.2** Gini presented a new index called arithmetic mean index.

Gini arranged the n values of a quantitative variable X, which was used to measure the income, from the lower to greater value

$$x_1 \le x_2 \le x_3 \le \ldots \le x_n .$$

Gini calculated the arithmetic mean

$$\frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{B_n}{n},$$

and next, he chose the m greater values to calculate the arithmetic mean

$$\frac{x_{n-m+1} + x_{n-m+2} + \ldots + x_n}{m} = \frac{B_m}{m},$$

for every m that $1 \le m < n$.

Next, Gini proposed the following inequality criterion: the income concentration will increase (decrease) if the following inequality is true

$$\frac{B_n}{n} \le \frac{B_m}{m}, \qquad (3.7)$$

[23]

that increases (decreases) for every m that $1 \le m < n$.

Let's see how the inequalities of (3.7) are connected with the values of the variable X.

If we take $m = n - j$, with $j = 1,2,\ldots,n-1$, then

[22] Population arranged from the greater to lower income.

[23] We include the equality in the inequality (3.7) to take into account the possibility of having equal values of X.

$$\frac{B_n}{n} = \left(\frac{j}{n}\right)\frac{\sum_{k=1}^{j} x_k}{j} + \left(\frac{n-j}{n}\right)\frac{B_{n-j}}{n-j}, \tag{3.8}$$

which shows (3.7).

If now we define the following distances

$$d_m = \frac{B_m}{m} - \frac{B_n}{n}, \tag{3.9}$$

for $m = 1,2,...,n-1$, then we can show that

$$\frac{B_n}{n} - x_k = (n-k)d_{n-k} - (n-k+1)d_{n-k+1}. \tag{3.10}$$

This formula (3.10) shows that the distance between the arithmetic mean, $\frac{B_n}{n}$, and the value x_k is a combination of two distances called d_{n-k} and d_{n-k+1}, for $m = 1,2,...,n-1$.

Now, if we suppose that n and $\frac{B_n}{n}$ are constants, then we can study the behaviour of the inequalities (3.7).

(a) If the inequalities (3.7) are equalities, then, $d_m = 0$ for $m = 1,2,...,n-1$, and from (3.10) we obtain that $x_k = \frac{B_n}{n}$, thus the concentration is minimum.

(b) If the inequalities (3.7) are maximum, then $d_m = \frac{B_n}{n} - \frac{B_n}{m}$, and from (3.10) the values of the variable X are: $x_k = 0$, for $m = 1,2,...,n-1$, and $x_n = B_n$, i.e. the concentration is maximum.

Another expression of (3.7) is

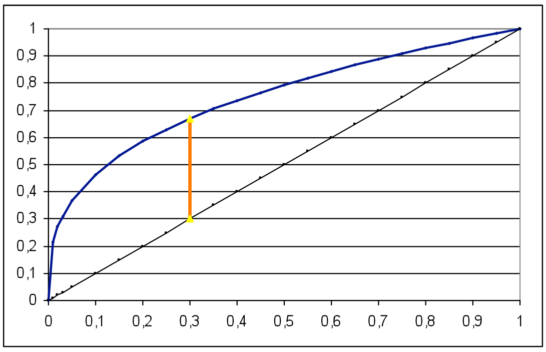
$$\frac{m}{n} \leq \frac{B_m}{B_n}, \tag{3.11}$$[24]

which compares the cumulative proportion of population with the corresponding cumulative proportion of income, for $m = 1,2,...,n-1$.

[24] This inequality functional criterion is more general that the inequality criterion of section **3.1**. This functional criterion doesn't imply that the Lorenz curves cut when comparing the two income populations, contrary to the inequality criterion in section **3.1**.

As the cumulative proportions, $\frac{m}{n}$ and $\frac{B_m}{B_n}$ are the same as the cumulative proportions of (3.11), then we can interpret theses proportions as the points $\left(\frac{m}{n};\frac{B_m}{B_n}\right)$, for $m=1,2,...,n-1$, of a dual Lorenz curve, where the differences $\frac{B_m}{B_n}-\frac{m}{n}$, for $m=1,2,...,n-1$, corresponds to the distances between the dual Lorenz curve and the equality line.

The following graph shows a continuous dual Lorenz curve, where we show a distance between the proportions $(0{,}3;0{,}67)$.



Next, we are going to see how Gini measured the distances $\frac{B_m}{B_n}-\frac{m}{n}$, for $m=1,2,...,n-1$.

In the paper of 1910, Gini measured the inequalities (3.11) by his δ index of concentration of 1909, i.e.

$$\left(\frac{B_m}{B_n}\right)^{\delta}=\frac{m}{n}, \tag{3.12}$$

for $m=1,2,...,n-1$.

If in (3.12) we take logarithms in both terms and we consider the index $\delta'=\delta-1$, then,

$$\delta'=\frac{\log\left(\frac{m}{n}\right)-\log\left(\frac{B_m}{B_n}\right)}{\log\left(\frac{B_m}{B_n}\right)}, \tag{3.13}$$

for $m=1,2,...,n-1$. We see that (3.12) measures the distance $\frac{B_m}{B_n}-\frac{m}{n}$ in a reative logarithmic scale. If (3.13) is approximately constant for $m=1,2,...,n-1$, then we can use the δ index of Gini. And also if the points $\left(\frac{m}{n};\frac{B_m}{B_n}\right)$, for $m=1,2,...,n-1$, approximately to straight line $\log(m) = \delta \log(B_m) - \log(W)$, where $W = \log\left(\frac{n}{B_n^\delta}\right)$, then δ index of Gini can be calculated from the Cauchy interpolation method.

For the cases where it wasn't possible to apply the formula (3.12) or the model $\log(m) = \delta \log(B_m) - \log(W)$, Gini proposed the following arithmetic mean index:

$$\delta_m = \frac{\log\left(\frac{m}{n}\right)}{\log\left(\frac{B_m}{B_n}\right)} \text{[25]}, \tag{3.14}$$

for $m = 1,2,...,n-1$, Gini proposed the index mean $\bar{\delta}$ [26] as the arithmetic mean of the indexes δ_m, for $m = 1,2,...,n-1$.

Gini applied his δ index of concentration to the total income distributions for several countries and different periods of time. At the end of section **13**, Gini came to the following conclusion:

"*Si ricordi ora che, per i redditi delle sole persone fisiche, δ varia tra 1,6 e 4, e, per i redditi delle persone fisiche e giuridiche, da 2,5 a più di 6. Per quante riserve si facciano sull'esattezza delle statistiche che ci servirono a calcolare gli indici di concentrazione, non pare dubbia la conclusione che: la distribuzione dei redditi globali è da Stato a Stato enormemente diversa*".

At the beginning of section **14**, Gini said that

"Questa conclusione è in netto contrasto con quello costitui per molti il più notevole e per tutti, credo, il più inaspettato fra i risultati delle moderne ricerche di economia inductiva.

Il Pareto, infatti, trattando le seriazioni dei redditi con metodo matematico, giunse alla conclusione che la distribuzione dei redditi globali è prssochè identica per tutti gli State e per tutti i tempi. Di qui egli traeva conseguenze di capitale importanza, come quella che la distribuzione della ricchezza di uno Stato è indipendente o quasi dalla sua

[25] Another expression is: $\frac{\delta_m - 1}{\delta_m} = \frac{\log\left(\frac{m}{n}\right) - \log\left(\frac{B_m}{B_n}\right)}{\log\left(\frac{m}{n}\right)}$, which is bounded in $[0,1]$. We'll be able to see how Gini defined in his 1914 paper his concentration ratio from relative distances of the cumulative proportions of population and income.

[26] An equivalent expression is $\bar{\delta} = 1 + \frac{1}{(n-1)}\sum_{k=1}^{n-1} \frac{\log\left(\frac{k}{n}\right) - \log\left(\frac{B_k}{B_n}\right)}{\log\left(\frac{B_k}{B_n}\right)}$.

costituzione economica. Tal conclusione era basata sul fatto che i valori dell'indice di distribuzione dei redditi globale, α *, determinato dal Pareto, differiscono poco da luego a luego e da tempo a tempo, oscillando intorno a 1,50, tra 1,89 (Prussia, 1852) e 1,13 (Amburgo, 1891)".*

In these texts, Gini criticized that the parameter α of Pareto changed relatively little among the different data sets. Gini obtained more different values of his δ index of concentration.

This different behaviour of δ index and the parameter α led Gini to say that

"...di esaminare la relazione que passa tra l'ndice di distribuzione dei redditi del Pareto e il nostro indice di concentrazione".

It was in section **14** that Gini obtained with approximate methods, that $\delta = \frac{\alpha}{\alpha - 1}$, for $\alpha > 1$. Gini said that

"*La relazione teorica* $\delta = \frac{\alpha}{\alpha - 1}$ *tra l'indice di distribuzione dei redditi del Pareto e l'indice di concentrazione nostro ci permette di esaminare se i risultati del Pareto autorizzano realmente quelle conclusión sulla uniformità della distribuzione della ricchezza a cui egli e molti altri sulle sue orme sono venuto*".

Again, in the beginning of section **16**, Gini criticized the behaviour of parameter α of Pareto

"*Ci pare che il nostro indice di concentrazione* δ *presenti alcuni vantaggi di fronte all'indice di distribuzione* α *del Pareto*".

In this section **16** Gini compared the δ index with the parameter α of Pareto. We are interested in the following points:

(a) L'indice δ è molto più sensibile dell'indice α.

"Notevoli differenze di distribuzione dei redditi rimangono appena avvertite dai valori di α *, specialmente se i valori di* α *sono bassi. Di qui era sorta presso molti l'idea, come vedemmo infondata, che la distrubuzione della ricchezza fosse pressochè uguale in tutti i paisi e in tutti tempi".*

From the relation $\delta = \frac{\alpha}{\alpha - 1}$, when α varies in the interval $(1,2)$, then δ varies in the interval $(2,+\infty)$.

(b) L'indice δ ha un significato preciso…Altrattanto non si può dire dell'indice α :

"Mentre il Pareto infatti ritiene che il crescere di α *indichi aumento di disuguaglianza nella distribuzione, il Benini ritiene al contrario che esso indichi disminuzione di disuguaglianza. E da avvertire che il dissenso dipende dal*

In consequence, Gini's inequality criterion is not comparable with Pareto's inequality criterion because Gini's criterion is equivalent to $\alpha_2 \geq \alpha_1$, which increases the parameter α, and Pareto's criterion is equivalent to $\alpha_2 \leq \alpha_1$ and $h_2 \geq h_1$, which decreases the parameter α.

When the cumulative distribution functions intersect, we can't apply Pareto's inequality criterion. An alternative could be to consider Pareto's weaker inequality criterion. According to the last criterion: $F_2(x)$ is a weaker inequal than $F_1(x)$ when $G_2(p) \geq G_1(p), 0 \leq p \leq 1$. Thus, Pareto's weaker inequal criterion is equivalent to $\alpha_2 \geq \alpha_1$ and $\mu_2 \geq \mu_1$[31].

In consequence, Gini's inequality criterion and Pareto's weaker inequal criterion are equivalent when $\alpha_2 \geq \alpha_1$ and $\mu_2 \geq \mu_1$. Thus, the parameter α increases in Gini's inequality criterion and Pareto's weaker inequality criterion[32] .

4. The mean difference of Gini (1912)

In this section we'll see the first part, *Variabilità e Mutabilità*, of the book whose title is *Indici di Variabilità*. This part has 111 pages that are grouped in 63 sections.

The sections we're interested in are: sections **11-13**, where Gini justified the study of the mean difference; section **14**, where different formulas of the mean difference were gathered for ungrouped data; section **19**, that extended those formulas in the case of having grouped data; section **24**, where Gini obtained a very useful result, a new formula of the mean difference that led him to extend it to the continuous case. In section **39** Gini showed the connection between the mean difference and the parameter α in the Model I of Pareto. Finally, in sections **43** and **44**, Gini obtained a rigorous proof of the connection between the δ of Gini and the $\in$ of Pareto.

A possible explanation of why Gini was interested in the study of the "*Variabilità e Mutabilità*" would be his inequality measure δ, that we know is built by certain accumulations of proportions. He didn't specifically gather specifically the differences of the income among the individuals as the variability measures do when used by contemporary statistics.

Gini wanted to show if the inequality of the incomes had increased or not, so it was reasonable that he looked for variability measures.

[31] To remember that one supposition of Pareto was that increasing the mean income is a necessary condition by to transfer income to poor persons.

[32] We have introduced Pareto's weaker inequality criterion because it implies an increase of the mean income. This criterion has been applied in section **2.3,** examples II and III; but Pareto' weaker inequality criterion has failed in example IV.

diverso significato che i due illustri statistici dànno alle espressioni: maggiore o minore disuguaglianza della distribuzione. Pochè, per il Pareto, la disuguaglianza aumenta quando diminuisce la percentuale dei censiti con redditi superiore ad x, mentre, per il Benini, in tal caso, la disuguaglianza diminuisce".

In this text, we can see that Gini knew the inequality criterion of Pareto. Gini said that the different conclusions of Vilfredo Pareto and Rodolfo Benini were due to different inequality criteria. Moreover, Gini said that

"*Prendendo le parola nel loro significato etimologico e corrente, dobbiamo dire che la concentrazione alla ricchezza aumenta e la sua disuguaglianza si fa più forte quando diminuisce la frazione dei censiti al di sopra di un dato reddito che possiedono una data parte dei redditi accertati, o viceversa quando aumenta la parte dei redditi accertati posseduta da una data frazione di censiti al di sopra di un dato reddito. Ora, in tal caso, aumenta il valore di δ e diminuiste corrispondentemente, in teoria, el valore di $\alpha = \frac{\delta}{\delta - 1}$".*

Gini defended that his δ index of concentration described better the concept of inequality than the inequality criterion of Pareto did, thus, he obtained that if the concentration increased, measured by the δ index, then parameter α decreased.

"*É necesario dunque concluyere che l'interpretazione del Benini, secondo il quale il diminuire di α indica un aumento di disuguaglianza nella distribuzione, corrisponde meglio che quella del Pareto al significato che comunemente si attribuisce alla espressione <<disuguaglianza nella distribuzione>>".*

3.4 Inequality criteria of Gini and Pareto, and the parameter α of Pareto Model I[27]

According to Gini's inequality criterion[28]: $F_2(x)$ is <u>less inequal</u> than $F_1(x)$ when $L_2(p) \geq L_1(p)$[29], for $0 \leq p \leq 1$. This definition is equivalent to $\frac{\mu_2(x_2(p))}{\mu_2} \geq \frac{\mu_1(x_1(p))}{\mu_1}$ for all $0 \leq p \leq 1$, where $\mu_k(x_k(p))$ is the mean income of incomes below the quantile $x_k(p)$, for $k = 1,2$. As Lorenz curves never intersect in the Model I of Pareto, then Gini's criterion is equivalent to delta-index Gini, thus $L_2(p) \geq L_1(p)$ is equivalent to $\delta_2 \leq \delta_1$ for all $0 \leq p \leq 1$. As $\alpha = \frac{\delta}{\delta - 1}$, then Gini's criterion is equivalent to $\alpha_2 \geq \alpha_1$[30].

According to Pareto's inequality criterion: $F_2(x)$ is <u>less inequal</u> than $F_1(x)$, when $F_2(x) \leq F_1(x)$ for all $x \geq 0$. Pareto's criterion is equivalent to $\alpha_2 \leq \alpha_1$ and $h_2 \geq h_1$, where some inequality is strict.

---

[27] [Barbut, 2007, chapter 7], for a discussion on behaviour of the parameter α.

[28] This is the interpretation that we have considered in section 3.1 on Gini's inequality criterion.

[29] Gini considered dual Lorenz curves, we consider Lorenz curves.

[30] We consider $\alpha > 1$.

4.1 Two groups of quantitative variables

In section **11** of the book, Gini distinguished two groups of quantitative variables: (1) the variable X defined by $X = I + J$, where I was a real value and J an error. If the error was null, then in the course of repetitive observations X would be equal to the real value, I. As normally is a non-null error, X appears with different quantitative modalities due to accidental or systematic errors that are produced by the observer, the measure instruments and other unexpected circumstances; and (2) the variable that, on the contrary, during the course of repetitive observations appears with different quantitative modalities that are a real[33] value, for example, the income values in a sample of individuals. Next, Gini said that the object of a variability examination was different for these two variable categories. In the first type, we are interested in: "*Di quanto le quantità rilevate differiscono dalla grandeza effectiva del carattere?*"[34]. In the second type, on the contrary, for every observed quantity there is a real objectivey size, so we have to set out the following problem: "*Di quanto le varie grandezze effettive differescono tra di loro?*"[35].

In section **12** of the book, Gini pointed out that the research about variability was limited to the first-type variables, whose studies had been carried out by astronomers[36]. In this case the observation of the variable breaks up in a real value plus an error due to the observation process, estimating the real value by the mean of the observations, i.e., that the mean "*rappresenta il valore probabile della grandezza effectiva del carattere*". A quantity that measures the difference between the observed value and the mean "*costituisce pertanto in indice appropriato di variabilità*". The following family gathers different variability indexes $^m S_X = \sqrt[m]{\frac{\sum_{i=1}^{n} |e_i|^m}{m}}$, where the values $e_i = x_i - \bar{x}$, $i = 1,2,...,n$, are the observation errors and X is a statistic quantitative variable. For $m = 2$ we obtain the variance[37] that "*é meno sensibile di ogni altro all'influenza del numero n delle osservazioni*". Less sensitive is the mean deviation that it's the mean of the absolute values of the errors that is has obtained for $m = 1$.

In section **13** of the book, Gini considered that the study of the variability in demography, anthropology, biology and economy, must approximate from the second type of variable. We are interested in: how different are the real values of the

[33] In this case the values of variable X will be defined by $x_i = \mu_i + \varepsilon_i$, i=1,2,… n, where μ_i is the real value of the observation ith and ε_i is its error. Gini considered that the errors are null when supposing that the real absolute values are much higher than its respective errors.
[34] How do the measured quantities differ from the real size of the variable?
[35] How do the real sizes differ among them?
[36] Gini gathered in pages 58 and 59 of this book the quotes from von Andrae (1869,1972), F. R. Helmert (1976) and W. Jordan (1869) who published several papers about the mean difference in the journal Astronom Nachr, see [David, 1968, 1998].
[37] Let's remember that the variance can be also expressed by $\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{2n^2}$, which is also valid for the study of the second type variables.

variable X ?[38] Gini said that while the mean of the first type of the variable was a real value, in the second type, the mean was a subjective value. Gini concluded that for the second type indexes must be used that could gather the differences between the real values of the variables that were studied. An important example of a variability index is the mean difference that Gini studied in the following sections of his book.

4.2 Mean difference formulas

In sections **14** to **24,** Gini brought in some formulas for the mean difference to simplify his calculus, and showed some of its proprieties.
Although the mean difference can be calculated for n values of a quantitative variable X without any order, the formulas proposed by Gini could be applied when the values are ordered from the lowest to the highest[39].

From the following order

$$x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n ,$$

Gini obtained the following expression for the mean difference

$$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{n(n-1)} = \frac{2}{n(n-1)}\sum_{k=1}^{k^*}(n+1-2k)(x_{n+1-k} - x_k)^{40}. \qquad (4.1)$$

When the left side of (4.1) is the definition of the mean difference without repetition and the right expression k^* is a whole number equal to $\frac{n}{2}$ for n even and $\frac{n+1}{2}$ for n odd. In the formula (4.1) it can be observed, for example, that the difference between x_n and x_1 is weighted with the quantity $n-1$; for the following differences, x_{n-1} and x_2, the weight is $n-3$, and so until the nearest values, x_{n+1-k^*} and x_{k^*}, whose weights are $n+1-2k^*$. From (4.1) we obtain the following formula

$$\Delta = \frac{1}{n(n-1)}\sum_{k=1}^{n}(n+1-2k)(x_{n+1-k} - x_k)^{41}, \qquad (4.2)$$

where we sum n terms.

---

[38] This is another explanation of why Gini got interested in the study of "*Variabilità e Mutabilità*".
[39] Another alternative that wasn't used by Gini would be to order the data from the highest to the lowest.
[40] This formula corresponds to formula (5) of section 14 of Gini and it's based on calculating each value of x_l, with $l = 1,2,\ldots,k^*$, the sum

$$\sum_{k=l+1}^{n-l}(x_k - x_l) + \sum_{k=l+1}^{n-l}(x_{n+1-l} - x_k) + (x_{n+1-l} - x_l) = (n+1-2l)(x_{n+1-l} - x_l).$$

[41] From this formula is easy to obtain the following one: $\Delta = \frac{2}{n(n-1)}\sum_{k=1}^{n}(2k-n-1)x_k$, that will be deduced in the research paper in 1914 as (12bis).

That the formula (4.1) was the double of (4.2) is due to, for example, that for $k = n$, the products $(n+1-2n)(x_{n+1-n} - x_n) = (-n)(x_1 - x_n)$ match up with the products $(n+1-1)(x_{n+1-1} - x_1) = (n)(x_n - x_1)$, when $k = 1$, where it is observed that although the terms of the multiplications change the sign when passing from $k = 1$ to $k = n$, the product doesn't change. This calculation has been made with two values of k, k_1 and k_2 so $k_1 + k_2 = n + 1$[42].

A similar expression to (4.2) is the following

$$\Delta = \frac{1}{n(n-1)} \sum_{k=1}^{n} |n+1-2k| |x_{n+1-k} - x_k|, \qquad (4.3)$$

where we have considered absolute values in each term of the formula (4.2). Gini named the first absolute value as the "distanza graduale" which would be represented by $d_{k,n+1-k}$, i.e. $d_{k,n+1-k} = |k - (n+1-k)| = |2k - n - 1| = 2\left|k - \frac{n+1}{2}\right|$, seeing that the above distance was the double of the distance between the rank k and the median of the ranks $1,2,...,n$ that will be named M_r [43]by us. With this notation, Gini wrote the formula (4.3) as

$$\Delta = \frac{1}{n(n-1)} \sum_{k=1}^{n} d_{k,n+1-k} |x_{n+1-k} - x_k|. \qquad (4.4)$$

Which coincides with formula (10) of section **16** of Gini.

If we calculate the median of the values of variable X, that when n is odd is equal to $M_x = x_{\frac{n+1}{2}}$, and when n is even we will choose the value $M_x = \frac{x_{\frac{n}{2}} + x_{\frac{n+2}{2}}}{2}$. It can be verified with the definition of the median that $|x_{n+1-k} - x_k| = |x_{n+1-k} - M_x| + |x_k - M_x|$. If we name d_{k,M_r} the distance between the rank k and the median M_r of the ranks, the expression (4.4) can be written by

---

[42] If n is odd, the value with rank $\frac{n+1}{2}$ also it fulfills the condition $\frac{n+1}{2} + \frac{n+1}{2} = n + 1$.

[43] When n is even, we know that there are infinite medians between $\frac{n}{2}$ and $\frac{n+2}{2}$. From all these medians, we are interested in the median $M_r = \frac{\frac{n}{2} + \frac{n+2}{2}}{2} = \frac{n+1}{2}$, that will be the one that will lead us to explain the formulas.

$$\sum_{k=1}^{n} d_{k,n+1-k} \left| x_{n+1-k} - x_k \right| = 2\sum_{k=1}^{n} d_{k,M_r} \left[\left| x_{n+1-k} - M_x \right| + \left| x_k - M_x \right| \right] =$$

$$= 2\sum_{k=1}^{n} d_{k,M_r} \left| x_{n+1-k} - M_x \right| + 2\sum_{K=1}^{n} d_{k,M_r} \left| x_k - M_x \right| =$$

$$= 4\sum_{k=1}^{n} d_{k,M_r} \left| x_k - M_x \right|,$$

where $d_{k,M_r} = d_{n+1-k,M_r}$. In consequence,

$$\Delta = \frac{4}{n(n-1)} \sum_{k=1}^{n} d_{k,M_r} \left| x_k - M_x \right|, \tag{4.5}$$

that coincides with formula (14) of section **18** of Gini.

The expression (4.5) led Gini to obtain a valid formula for grouped data in absolute frequencies.

Gini grouped the n observations of variable X in the s different values, i.e. in $\{(x_j, n_j)\ j = 1,2,...,s\}$, where the data are still ordered from the lowest to the highest. Now formula (4.5) is written as

$$\Delta = \frac{4}{n(n-1)} \sum_{j=1}^{s} \bar{d}_{j,M_r} n_j \left| x_j - M_x \right|^{44}, \tag{4.6}$$

where the expression $\bar{d}_{j,M_r}$ is the mean of the distances of the rank of the individuals that have the same value x_j to the median of the ranks M_r, i.e.

$$\bar{d}_{j,M_r} = \frac{(N_{j-1} + 1 - M_r) + (N_{j-1} + 2 - M_r) + ... + (N_{j-1} + n_j - M_r)}{n_j},$$

where N_j are the cumulative absolute frequency, from the lowest to the highest till the value x_j.

Gini gave a last expression of the mean difference in section **20** of his book. Let’s see Gini’s procedure.

For n even, it’s easy to prove that $\sum_{k=1}^{n} \left| k - \frac{n+1}{2} \right| = \frac{n^2}{n}$, on the contrary, when n is odd, then $\sum_{k=1}^{n} \left| k - \frac{n+1}{2} \right| = \frac{(n+1)(n-1)}{4} = \frac{n^2 - 1}{4}$. From here, we obtain the following formulas.

[44] This formula corresponds to formula (16) of section **19** of Gini. In his 1914 paper, Gini gave us another formula of the mean difference for grouped data that was more operative.

The expression (4.5) is also written as

$$\Delta = \frac{n}{(n-1)} \frac{\sum_{k=1}^{n} d_{k,M_r} |x_k - M_x|}{\frac{n^2}{4}},$$

and for n even, we have

$$\Delta = \frac{n}{(n-1)} \frac{\sum_{k=1}^{n} d_{k,M_r} |x_k - M_x|}{\sum_{k=1}^{n} d_{k,M_r}}^{45}, \tag{4.7}$$

that corresponds to expression (21) gathered in section **21** of Gini.

For n odd, the following formula is obtained

$$\Delta = \frac{n+1}{n} \frac{\sum_{k=1}^{n} d_{k,M_r} |x_k - M_x|}{\sum_{k=1}^{n} d_{k,M_r}}.^{46}$$

All the previous formulas can be applied to the mean difference with repetition as showed in the following expression

$$\Delta_r = \frac{n-1}{n}\Delta,$$

where Δ_r is the mean difference with repetition.

The last formula we are gathering is the one that is obtained by Gini in section **24** of his book. Gini extended the formula of Δ_r to the continuous case in order to apply the formula Δ_r to the model I of Pareto. Let's see how Gini proceeded.
For each value x_h of variable X, $h = 1,2,...,n$; Gini calculated n times the mean deviation with respect to this value, i.e.

$$\sum_{k=1}^{n} |x_k - x_h| = \sum_{k=h}^{n} (x_k - x_h) + \sum_{k=1}^{h-1} (x_h - x_k),$$

and solving the equation

$$\sum_{k=1}^{n} |x_k - x_h| = 2\sum_{k=h}^{n} x_k - \sum_{k=1}^{n} x_k - nx_h + 2(h-1)x_h,$$

it can be written $|x_k - M_x|$

[45] This formula shows that the weights, $\frac{d_{k,M_r}}{\sum_{j=1}^{n} d_{j,M_r}}$, are symmetries and they increase when $|x_k - M_x|$ increase.

[46] This formula and formula (4.7) have been used in [Berrebi and Silber, 1987].

$$\sum_{k=1}^{n}|x_k - x_h| = 2\sum_{k=h}^{n} x_k - \sum_{k=1}^{n} x_k + nx_h - 2(n-h+1)x_h .$$

If we sum for every h

$$\sum_{h=1}^{n}\sum_{k=1}^{n}|x_k - x_h| = 2\sum_{h=1}^{n}\sum_{k=h}^{n} x_k - 2\sum_{h=1}^{n}(n-h+1)x_h . \tag{4.8}$$

Gini wrote the formula (4.8) by the following expression,

$$\sum_{h=1}^{n}\sum_{k=1}^{n}|x_k - x_h| = 2\sum_{h=1}^{n}(T_h - F_h x_h), \tag{4.9}$$

where

$$T_h = \sum_{k=h}^{n} x_k \quad \text{and} \quad F_h = n - h + 1 .$$

Formula (4.9) corresponds to formula (26) of the book of Gini.

From (4.9), Gini obtained the following formula of the mean difference with repetition in the continuous case.

$$\Delta_r = 2\int_{x_0}^{\infty}\left[\int_{x}^{\infty} y f(y)dy - x(1-F(x))\right] f(x)dx \; ^{47}, \tag{4.10}$$

where $f(x)$ is the density function of a random variable X with its values in $[x_0, \infty)$ and F(x) is its cumulative distribution function with a finite expected value. This formula (4.10) was used by Gini in section **39** to calculate the mean difference in the Model I of Pareto.

Another expression that can be deduced from (4.8) is

$$\sum_{h=1}^{n}\sum_{k=1}^{n}|x_k - x_h| = 2\sum_{h=1}^{n}(2h - 1 - n)x_h , \tag{4.11}$$

That is more operative than expression (4.9).

This formula can be written also as,

$$\sum_{h=1}^{n}\sum_{k=1}^{n}|x_k - x_h| = 4\sum_{h=1}^{n}\left(h - \frac{(n+1)}{2}\right)x_h .$$

Now, the expression of the mean difference with repetition is

[47] This formula is valid when the mathematical expectation is finite. It is also obtained from the following formula: $\Delta_r = \int_{x_0}^{\infty}\int_{x_0}^{\infty}|x-y| f(x)f(y)dxdy = 2\int_{x_0}^{\infty}\left[\int_{x}^{\infty}(y-x)f(y)dy\right]f(x)dx$.

$$\Delta_r = \frac{\sum_{h=1}^{n}\sum_{k=1}^{n}\left|x_i - x_j\right|}{n^2} = \frac{4}{n}\sum_{h=1}^{n}\left(\frac{h}{n} - \frac{(n+1)}{2n}\right)x_h, \tag{4.12}$$

that can be written also as

$$\frac{\Delta_r}{2} = 2Cov(X,F), \tag{4.13}$$

where $F_h = \frac{h}{n}$, for $h = 1,2,...,n$.

Formula (4.13) is written for the continous case as

$$\frac{\Delta_r}{2} = 2\int_{x_0}^{\infty} x\left(F(x) - \frac{1}{2}\right)f(x)dx^{48}. \tag{4.14}$$

This formula (4.14) is more operative than formula (4.10) obtained by Gini.

The expression of the mean difference without repetition for (4.12) is

$$\Delta = \frac{\sum_{h=1}^{n}\sum_{k=1}^{n}\left|x_i - x_j\right|}{n(n-1)} = \frac{4}{(n-1)}\sum_{h=1}^{n}\left(\frac{h}{n} - \frac{(n+1)}{2n}\right)x_h, \tag{4.15}$$

that can be written as

$$\frac{\Delta}{2} = \frac{Cov(X,h)}{\overline{h}}, \tag{4.16}$$

where $h = 0,1,2,...,n-1$, and $\overline{h}$ is the mean.

4.3 Relationship between the mean difference and the parameter α of Pareto

In section **39** of his book, Gini obtained for the Model I of Pareto a theoretical expression of the mean difference. For this calculation he used formula (4.10) so Gini was able to connect a variability measure as the mean difference with repetition, with the parameter α of Pareto, which was useful to show that when the relative mean difference increases (decreases), then the parameter α must increase (decrease) and the other way around. Let's see this result of Gini.

Gini began with the Model I of Pareto which was defined in the following way

[48] Formula (4.14) defines the mean difference as four times the covariance between the random variables X y F(X), as $\int_{x_0}^{\infty} F(x)f(x)dx = \int_0^1 udu = \frac{1}{2}$. Also the covariance between the random variables X y F(X) is the area of the segment that connects the points (0,0) and (0, μ) in a Cartesian coordinate System, equality line, and the generalized Lorenz curve. Let's remember that $\mu(1-R)$ is the double of the area below the generalized Lorenz curve when the variable is non-negative.

$$f(x) = V x^{-h}, \tag{4.17}$$

where $h = \alpha + 1$ and V is a constant that depends on h and the minimal income x_0. To calculate the mean difference with repetition, Δ_r, we will use formula (4.14) that is more operative than formula (4.10) of Gini [49].

$$\Delta_r = 4\int_{x_0}^{\infty} x\left(F(x) - \frac{1}{2}\right) f(x)dx = 4\int_{x_0}^{\infty} x\left[1 - \left(\frac{x}{x_0}\right)^{-\alpha} - \frac{1}{2}\right]\frac{\alpha x_0^{\alpha}}{x^{\alpha+1}}dx,$$

obtaining

$$\Delta_r = \frac{2(h-1)x_0}{(h-2)(2h-3)}, \tag{4.18}$$

That corresponds to formula (96) of Gini.

Then Gini calculated the mean income, which was named with the letter A, whose expression is the following

$$A = x_0 \frac{h-1}{h-2},$$

this led Gini to write formula (4.18) in the following way

$$\Delta_r = \frac{2A}{(2h-3)},$$

that expressed in section **41** of his book in the following way

$$\frac{\Delta_r}{A} = \frac{2}{(2h-3)}, \tag{4.19}$$

since $\alpha = h - 1$, then

$$\frac{\Delta_r}{A} = \frac{2}{(2\alpha - 1)}. \tag{4.20}$$

It was page 71, at the end of section **41** of the book, that Gini showed us some calculation about the behaviour of how the relative mean difference faced the changes in the parameter α of Pareto.

	$\alpha = 1{,}17$	$\alpha = 1{,}5$	$\alpha = 1{,}9$
$\frac{\Delta_r}{A}$	1,5	1,0	0,7

Formula (4.20), and the illustrated examples of the table show that the relative mean difference decreases (increases) when α increases (decreases) and the other way

[49] Gini supposed in the beginning that the maximum income is finite. But, in order to simplify the result, Gini supposed a maximum income equal to infinite. These calculations are similar to ours in section **4.3**.

round, which is the opposite to Pareto's[50] interpretation. Gini summed it up in the beginning of section **42** of his book by saying:

"*La quantità* $\alpha = h - 1$ *venne da lui assunta come indice di distribuzione dei redditi. Egli ammetteva che la disuguaglianza crescesse o diminuisse con* α. *Il risultato a cui noi siamo venuti, che tutti gli indici di variabilità crescono o diminuiscono col diminuire o col crescere di* α *mette fuori di dubbio che si debe dare ad* α *il significato opposto*"[51].

4.4 Relationship between the δ of Gini and the parameter α of Pareto

The relationship $\delta = \dfrac{\alpha}{\alpha - 1}$ which Gini tried to prove in 1910 was the purpose of sections **43** and **44** of the book written in 1912, where Gini carried out a rigorous investigation. Let's see Gini's procedure.

From Model I of Pareto (4.17), Gini calculated the following integrals

$$F_x = \int_x^\infty V\, t^{-h+1} dt = \frac{V\, x^{-h+1}}{(h-1)}, \qquad (4.21)$$

and

$$T_x = \int_x^\infty V t\, t^{-h+1} dt = \frac{V\, x^{-h+2}}{(h-2)}, \qquad (4.22)$$

where it must be demanded that $h > 2$.

From (4.22) he obtained that

$$\frac{1}{x} = \left[\frac{V\, x^{-h+2}}{(h-2)}\right]^{\frac{1}{(h-2)}},$$

and replacing in (4.21), he obtained

$$F_x = \frac{V}{(h-1)}\left[\frac{T_x (h-2)}{V}\right]^{\frac{(h-1)}{(h-2)}},$$

i.e.

$$F_x \propto T_x^{\frac{(h-1)}{(h-2)}}. \qquad (4.23)$$

Now if we compare (4.23) with formula (3.6) in 1909 that Gini used to define his concentration δ index, it can be deduced that $\delta = \dfrac{h-1}{h-2}$, and knowing that $h = \alpha + 1$, we

---

[50] Let's remember that here the inequality criterion of Pareto is different from the inequality criterion of Gini.

[51] Gini calculated several variability measures in the Model I of Pareto. He showed that if α increase (decreases), then the variability increases (decreases).

can obtain the relationship $\delta = \dfrac{\alpha}{\alpha - 1}$. Gini did the opposite proof, i.e., from (4.23) to obtain[52] (4.17).

With this proof, Gini obtained the dual Lorenz curve (4.23) from the Model I of Pareto and the opposite, the Model I of Pareto (4.17) from the dual Lorenz curve (4.23), i.e. the dual Lorenz curve that Gini used in his article in 1909 to define his concentration index δ proved dependent on the Model I of Pareto. It was in 1914 that Gini left his concentration measure δ to look for an inequality measure with a better validity[53].

5. The R Gini concentration ratio (1914)

On March 29th when Gini was teaching statistics in the University of Padova he presented his research paper "*Sulla misura della concentrazione e della variabilità dei carattere*". This paper was published that year in the *Atti del R. Istituto Veneto di Scienze, Lettere ed Arti,* volume LXXIII, part II, 1203-1258.

In this paper, Gini proposed his R concentration ratio, which is more general than the delta-index δ in 1909, R could be defined for all the cumulative distribution function of a non negative random variable with a finite expected value.

The paper is divided into **13** sections. We are interested in sections **1** and **2** where Gini defined his R concentration for ungrouped data; in section 3 where we can see two expressions of R that were applied to grouped data and class intervals; in section **6** we'll see how Gini proposed the double of the area between the equality diagonal and the Lorenz curve as a measure of concentration that can be applied in case we had to compare two Lorenz curves cut. Gini approximated this measure by this R concentration when the number of observations was large; in section **7** we'll study how Gini approximated R concentration from a Lorenz polygonal curve built with five points; in section **9** we'll see that R is the ratio between the mean difference without repetition and it's twice the mean. This result let Gini find a decomposition of the mean difference for grouped data in class interval. Finally, we will propose a proof of how Gini could connect the double of the concentration area with the relative mean difference.

5.1 The R Gini ratio: ungrouped data

In section **1**, Gini arranged the n individual from lower to greater incomes and defined the cumulative proportion of population, $p_i = \dfrac{i}{n}$, and the income, $q_i = \dfrac{A_i}{A_n}$, for the i-th individual, where $A_i = \sum_{k=1}^{i} x_k$, $i = 1,2,...,n$. Gini considered the points of the dual Lorenz curve in his 1909 article, but in this work Gini considered the points of the Lorenz curve. Gini summarized his δ index of 1909 and other indexes gathered in 1910. He finished this section saying:

---

[52] The proof to go from (4.23) to (4,17) was brought by [Mortara, G., 1911].

[53] In the beginning of this paper in 1914, Gini found an inequality measure of which he said: "*Di una misura della concetrazione independente dalla distribuzione del carattere*".

"*La presente nota ha lo scopo di proporre una misura della concentrazione, che sia indipendente dalla curva di distribuzione del carattere e permetta quindi di esseguire paragoni tra la concentrazione dei carattere più varii*"[54].

It was in section **2** that Gini proposed directly[55] his R concentration. Let's see his expression:

$$R = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i}, \quad (5.1)$$ [56]

which has values from $R = 0$, perfect equality, to $R = 1$, maximum concentration.

Gini didn't give any geometric interpretation to each part of formula (5.1), in the sense that for the numerator, the expression $\frac{\sum_{i=1}^{n-1}(p_i - q_i)}{n}$ is the <u>concentration area</u> between the empirical[57] Lorenz curve and the equality line [58]. For the denominator of (5.1), the expression $\frac{\sum_{i=1}^{n-1} p_i}{n}$ is the triangle area whose vertices are: $(0,0)$, $\left(\frac{n-1}{n},0\right)$ and $(1,1)$. This area is the value maximum of the concentration area.

5.2 Gini's concentration ratio from grouped data and class intervals.

A formula more operational than (5.1) is

$$R = \frac{2\sum_{i=1}^{n}(i-1)x_k}{(n-1)A_n} - 1, \quad (5.2)$$

which is only valid, as (5.1), for ungrouped data. In this formula (5.2), if we defined a new variable Z where the value $z_i = i - 1$ is the rank of each individual diminished in one unit, then we can immediately obtain the following formula

---

[54] The dependence of the inequality δ index of Model I of Pareto has led Gini to look for another concentration measure with more applications.

[55] Gini compared p_i and q_i with $R_i = \frac{p_i - q_i}{p_i}$, that is a valuation between the distances of the proportion of population and income.

[56] R is a mean pondered of R_i, with ponderations that are proportional to p_i values.

[57] We'll see in section **6** of this work how Gini built a curve with p_i values in the axis of abscissa, and the q_i values for the axis of ordinate. Instead of joining those two points by a segment and build an empirical Lorenz curve, he joined them by a "*una linea continua*" with continuous derivative in each point that is called concentration curve.

[58] Gini called it "*retta di equidistrubuzione*".

$$R = \frac{2\sum_{i=1}^{n} z_k x_k}{(n-1)A_n} - 1 = \frac{Cov(Z,X)}{\overline{z}\overline{x}}\text{[59]}. \qquad (5.3)$$

For the grouped data, $\{(x_k, n_k): k = 1,2,...,r\}$, Gini obtained the following formula:

$$R = \frac{\sum_{k=1}^{r} (N_{k-1} + N_k - 1)t_k}{(n-1)A_n} - 1 = \frac{\sum_{k=1}^{r} (N_{k-1} + N_k - n)t_k}{(n-1)A_n}\text{[60]}, \qquad (5.4)$$

where N_k are the cumulative absolute frequencies and $t_k = x_k n_k$, with $k = 1,2,...,r$, being $N_0 = 0$.

For grouped data in class intervals, $\{(L_{k-1}, L_k]n_k, t_k : k = 1,2,...,r\}$, being n_k and t_k the total of individuals and the total income in the interval $(L_{k-1}, L_k], k = 1,2,...,r$, we are going to express the variable statistics as X_{kl}, where for each k-th interval this variable has values for the incomes that are inside the interval. Similarly we consider the variable range Z_{kl}, which for the interval k-th has the ranges $N_{k-1} + l - 1$, for $l = 1,2,..,n_k$. With these definitions and using formula (5.2), we obtain the following Gini concentration ratio

$$R = \frac{2\sum_{k=1}^{r}\sum_{l=1}^{n_k} (N_{k-1} + l - 1)x_{kl}}{(n-1)A_n} - 1. \qquad (5.5)$$

If we consider now the variable $\delta_{kl} = x_{kl} - \overline{x}_k$, where $\overline{x}_k$ is the arithmetic mean of the values of X_{kl} in the k-th interval; and so the variable $\varepsilon_{kl} = z_{kl} - \overline{z}_k$, where $\overline{z}_k$ is the arithmetic mean of the values of Z_{kl} in the k-th interval, then we can write the formula (5.5) in the following way

$$R = \frac{\sum_{k=1}^{r} (N_{k-1} + N_k - 1)t_k}{(n-1)A_n} - 1 + \frac{2\sum_{k=1}^{r}\sum_{l=1}^{n_k} \delta_{kl}\varepsilon_{kl}}{(n-1)A_n}. \qquad (5.6)$$

From (5.6) we have obtained the expression of R

$$R = \frac{\sum_{k=1}^{r} (N_{k-1} + N_k - n)t_k}{(n-1)A_n} + \frac{2\sum_{k=1}^{r} n_k Cov(Z_k, X_k)}{(n-1)A_n}\text{[61]}, \qquad (5.7)$$

[59] You can see that we have expressed R as the covariance between the variables Z and X that we divide by the product of their arithmetic means. Although this formula is an immediate consequence, Gini didn't take into account in this work.

[60] This formula is the ratio between the covariance of the data $\{(x_k, \overline{z}_k; n_k): k = 1,2,...,r\}$ and the product $\overline{x}\overline{z}$, where $\overline{z}_k$ is the arithmetic mean of the ranges that correspond to the observation x_k, for all k.

where $Cov(Z_k, X_k)$ is the covariance between variable Z_k [62], that is the range of each value into the interval $(L_{k-1}, L_k]$, and variable X_k, which has the values of the incomes in the interval $(L_{k-1}, L_k]$. The first part of (5.7) is what Gini named R', that is his formula (15), which is useful as it only depends on the total individuals n_k, and on the total of the income t_k, for each class interval $(L_{k-1}, L_k]$, for $k = 1,2,...,r$. This term is called <u>concentration ratio "between",</u> R_E, that is the concentration ratio between the average value, $\bar{x}_k$, of each one of the intervals, with frequencies n_k, for $k = 1,2,...,r$. Nevertheless, the second part of (5.7) depends on the value of the income of each of the class intervals.

Another expression of (5.7) is

$$R = \frac{\sum_{k=1}^{r} (N_{k-1} + N_k - n)t_k}{(n-1)A_n} + \frac{\sum_{k=1}^{r} n_k b_k (n_k^2 - 1)}{6(n-1)A_n}, \qquad (5.8)$$

where b_k is the slope of the regression line $X_k = a_k + b_k Z_k$, into each class interval $k = 1,2,...,r$. When the value of the individuals' incomes into the intervals is unknown, it's not possible to use formula (5.8). In this case, Gini supposed that the values of variable X into each interval followed an arithmetic progression where the last value was the upper class limit of interval. This hypothesis is equivalent to supposing that $b_k = \frac{a_k}{n_k}$. With this supposition Gini approximated expression (5.8) by the following formula

$$R'' = R_E + \frac{\sum_{k=1}^{r} (n_k^2 - 1)a_k}{6(n-1)A_n}. \qquad (5.9)$$

[63]

Where $a_k = L_{k-1} - L_k$. This formula (5.9) is the (17) of Gini[64].

If we suppose that the values of variable X_{kl} are known into each of the class intervals we'll be able to calculate from (5.3) a formula for the truncate concentration ratio R_k in the interval $(L_{k-1}, L_k]$, that is

---

[61] The second part of this formula is not gathered by Gini, but its deduction is immediate from the second part of formula (5.6) of Gini.

[62] For example, variable Z_k takes the value j-1 for the jkh income of the class interval $(L_{k-1}, L_k]$.

[63] This approximation can have values greater than the unity. Its validity depends on the good adjustments of the data and the size of n_k into each class intervals of the straight lines $X_k = a_k + b_k Z_k$.

[64] When the totals of the income are unknown into the intervals, Gini will take the class marks, $\hat{x}_k = \frac{L_k + L_{k-1}}{2}$, being $\hat{x}_k n_k$ an estimation of the totals.

$$R_k = \frac{Cov(Z_k, X_k)}{\bar{z}_k \bar{x}_k}, \tag{5.10}$$

where $\bar{z}_k$ y $\bar{x}_k$ are the means in $(L_{k-1}, L_k]$ of Z_k and X_k, respectively. From (5.10) we can write (5.7) as

$$R = R_E + \sum_{k=1}^{r} \left(\frac{t_k}{A_n}\right)\left(\frac{n_k - 1}{n-1}\right) R_k{}^{65}, \tag{5.11}$$

where the second term of (5.11) is the concentration ratio "intra", R_D. Formula (5.11) is equivalent to formula (24) of Gini.

Formula (5.11) is a decomposition of the Gini concentration ratio R in the concentration ratio R "between", R_E, and the concentration ratio "into", R_D.

Comparing formulas (5.9) and (5.11), it can be showed that if the concentration ratios R_k are equal to $\frac{1}{6}$, and the size of n_k is big enough, then (5, 11) approximates to (5.9). We can see that the approximation of Gini, formula (5.9), overestimates R when the concentration ratios R_k are lower than $\frac{1}{6}$; and that if the concentration ratios R_k are greater than $\frac{1}{6}$, then formula (5.9) underestimates R.

5.3 The double of concentration and the Gini concentration ratio

In section **6** of the paper, Gini proposed the double of concentration of the Lorenz curve as a new measure of concentration that would connect it with its concentration ratio R.

It started saying that apart from his ratio concentration R there was a graphical method that some authors like [Lorenz, 1905], [Chatelain, 1907, 1910, 1911] and [Séailles, 1910] proposed to judge the greater or lower inequality of the income distribution. In the footnote Gini added [G. P. Watkins, 1905, 1908 y 1909], [W. M. Person, 1909] and the book of [W. J. King, 1912].

Page 1229, Gini represented the points (p_i, q_i), for $i = 1,2,...,n$, in a cartesian coordinate system and traced a continuous curve with a continuous derivative on the

[65] This decomposition will be valid when the individuals are arranged by decreasing incomes, which means that the first individuals N_1 belong to the first group, the next individuals N_2 belong the second group etc. When the groups are arranged by a variable different from the income, for example by geographic areas, formula (5.11) comes with a term named interaction, that is the difference between the concentration ratio "into" (with the ranks that the individuals have in the population) and the concentration ratio "inside" (with the ranks that the individuals have in their group). This can be seen in: [Sastry and Kelkar, 1994].
When the variable total income is the sum of the different sources of incomes, for example the work incomes, capital incomes etc.., and you want to measure how the concentration of the total income is affected by an increase of the concentration of one of the sources, formula (5.3) is very useful to connect the concentration of the total income with the concentration of different incomes. This can be seen in [Lerman and Yitzhaki, 1985].

aforementioned points. Gini also traced a segment called "retta di equidistribuzione" joining the points $(0,0)$ and $(1,1)$. The curve resultant was the one known as the Lorenz curve (see graphic-I of Gini)

Page 1230, graphic-I bis, Gini added to the Lorenz curve other curves: (1) the points (q_i, p_i), for $i = 1,2,...,n$, generate the curve that Lorenz gathered in his paper in 1905; (2) the points $(1 - p_i, 1 - q_i)$, for $i = 1,2,...,n$, generate what today is called the dual Lorenz curve that was gathered by Chatelain (1907, 1910, 1911) and (3) the points $(1 - p_i, q_i)$, for $i = 1,2,...,n$, generate a curve gathered by Séailles (1910). Gini gathered in his paper another kind of concentration curves.

Next, Gini went back to the Lorenz curve represented in graphic-I and declared that

"*La curva di concentrazione è tanto meno accentuata quanto meno disuguale è la distribuzione della ricchezza, fino a diventare, nel caso di perfetta uguaglianza di distribuzione, una retta (retta di equidistribuzione*".

"*Gli autori sopra nominati trassero partito di questa propietà della curva di concentrazione per eseguire confronti sulla distribuzione della ricchezza*".

"*Disegnando sullo stesso diagramma più relative a tempi o luoghi diversi, essi erano in grado di giudicare in quale tempo o in quale luego la ricchezza risultava più concentrata".*

For Gini, with this graphical method, the concentration curve A will be less inequality than other B if A has a smaller bend ("*meno accentuate")* than B.

Page 217 of the paper of Lorenz, he declared that

"*With unequal distributions, the curves will always begin and end in the same points as with an equal distribution, but they will be bent in the middle; and the rule of interpretation will be, as the bow is bent, concentration increases*".

Lorenz considered the concentration ratio as the elastic stick of an arch saying that: a concentration curve A will be more concentrated than B if the stick increases its bend. Lorenz set this bend in the middle ("in the middle") of the concentration curve.

On the other hand, Gini declared that this graphical method had two drawbacks already recognized by Lorenz and King:

1. It didn't contribute to a precise measure of the concentration.
2. Not even let, in some cases, to judge about the greater or lower concentration. So when the concentration curves cut themselves (Lorenz gathered in his paper two concentration curves that cut[66] themselves), this graphical method didn't allow to judge the value of the concentration.

[66] This example of two empirical Lorenz curves, (A) and (B), which are crossed, is interesting because curve (B) is obtained by curve (A) when the individuals that are in the middle of the incomes of (A)

To find a solution to these drawbacks, Gini proposed, for the first time, the ratio between the area of concentration (area between the concentration curve and the equality line) and the triangle area (with the vertices: $(0,0)$, $(1,1)$ and $(1,0)$) as a concentration measure. That is to say that Gini proposed the double of the concentration area as a concentration[67] measure.

How did Gini connect this new inequality measure with his concentration ratio R?

Gini made an approximation of the area below the concentration curve with rectangles whose bases were $\frac{1}{n}$ long and heights were below the concentration curve. Equally Gini did the same with the triangle (with vertices: $(0,0)$, $(1,1)$ and $(1,0)$).

Gini declared that:

1. The area of n rectangles, $\sum_{k=1}^{n-1} \frac{q_k}{n}$, tended to the area below the concentration curve when n was large.

2. The area of n rectangles, $\sum_{k=1}^{n-1} \frac{p_k}{n}$, tended to 0,5 (area of the triangle with vertices $(0,0)$, $(1,1)$ y $(1,0)$) when n was large.
3. The difference, $\sum_{k=1}^{n-1} \frac{(p_k - q_k)}{n}$, between $\sum_{k=1}^{n-1} \frac{p_k}{n}$ and $\sum_{k=1}^{n-1} \frac{q_k}{n}$, was the area of n rectangles that tended to the concentration area when n was large.

Next, the concentration ratio

$$R = \frac{\sum_{i=1}^{n-1} \frac{(p_i - q_i)}{n}}{\sum_{i=1}^{n-1} \frac{p_i}{n}},$$

tended more and more to the double of the concentration area when n is large.

With these arguments, Gini connected his concentration ratio R with the double of the area of concentration.

transfer incomes to the individuals that are poorer and the ones that are richer. Then you can see that progressive and regressive transfer happen at the same time. The values of the Gini concentration are: $R_A = 0{,}133$ and $R_B = 0{,}16$. This means that the ratio of Gini increases the concentration in (B) respecting to (A). But if you calculate the Bonferroni index, the values are: $B_A = 0{,}199$ and $B_B = 0{,}176$. So, for this last index, the concentration in (B) decreases respecting the concentration of (A). The difference between these indexes is that the index of Gini ponderates in same way the progressive and regressive transferences (formula (4,7)), while the Bonferroni index gives more weights to the progressive transferences than to the regressive ones (Imedio, 2007, p. 109).

[67] Gini has forgotten his inequality criterion that was proposed in his article in 1909, as when the Lorenz curves crossed, the dual Lorenz curves do it too and then his general inequality criteria can't be applied.

5.4 An approximation of Gini ratio from five points of the concentration curve.

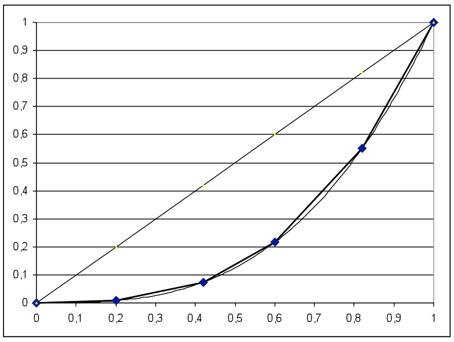
In the end of section **6** Gini said that,

"*…per descrivere la curva di concentrazione, non è necessario di conoscere tutti i valori di p_i e q_i. Basta la conoscenza di 4 o 5 di questi valori per che la curva possa essere descritta con sufficiente approssimazione*.".

In the beginning of section **7**, Gini said that,

"*Questa osservazioni suggeriscono un altro procedimento per determinare praticamente il valore de* R."

Gini proposed a graphical method to obtain an approximation of the concentration area and consequently the value of R.



It is in section **8** that Gini took five points of the concentration curve to approximate the value of R. Let's see the procedure.

From the concentration curve made by the points $\{(p_i, q_i): i = 1,2,...,n\}$, Gini chose five points $\{(p_{i_k}, q_{i_k}): k = 1,2,...,5\}$ to join them by segments and so generate a polygonal concentration curve. Next Gini calculated the area over this polygonal curve with trapeziums, for example, the fourth trapezium had the greater base, p_{i_4}, the lower base, p_{i_3}, and the height, $q_{i_4} - q_{i_3}$.

Then the area is equal to

$$\sum_{k=1}^{5}\left(\frac{p_{i_{k-1}} + p_{i_k}}{2}\right)\left(q_{i_k} - q_{i_{k-1}}\right),$$

where $p_{i_k} = \frac{N_{i_k}}{n}$ and $q_{i_k} - q_{i_{k-1}} = \frac{t_{i_k}}{A_n}$, being $N_{i_0} = 0$. Replacing in the last formula we obtain

$$\sum_{k=1}^{5}\frac{\left(N_{i_{k-1}} - N_{i_k}\right)t_{i_k}}{2nA_n}.$$

Now, the double of the concentration area of the polygonal curve is

$$2\left(\sum_{k=1}^{5}\frac{\left(N_{i_{k-1}} - N_{i_k}\right)t_{i_k}}{2nA_n} - \frac{1}{2}\right) = \sum_{k=1}^{5}\frac{\left(N_{i_{k-1}} - N_{i_k} - n\right)t_{i_k}}{nA_n} = \frac{n-1}{n}R_E,$$

that tends to the concentration ratio R_E, when n is large.

As the double of the concentration area of the curve of n points (when n is large), tends to the value R of concentration, then we can approximate R with the value of R_E. This is the essential reasoning of Gini[68].

5.5 A proof that the Gini concentration ratio is equal to the relative mean difference

In section **9**, Gini showed the relation between his concentration ratio R and the mean difference without repetition. After revising sections 11-13 of his book in 1912, Gini declared the following:

"*Dimostreremo ora che il rapporto di concentrazione coincide col rapporto della differenza media al valors massimo che questa può assumere, o in altre parole, col rapporto della differenza media al doppio della media aritmetica del carattere.* ".

The demonstration that Gini presented in pages 1237-1238 was confusing, as Gini "forced" the demonstration in order to make the following formula true:

$$R = \frac{2\sum_{k=1}^{n}(k-1)x_k}{A_n(n-1)} - 1 = \frac{1}{A_n(n-1)}\sum_{k=1}^{k^*}(n+1-2k)(x_{n+1-k} - x_k) = \frac{\Delta}{2\frac{A_n}{n}}.$$

The identity above is easily obtained from formulas (4.6) and (5.3).

From formula (4.16) and the relation between the R of Gini and his mean difference without repetition Δ, we obtain again formula (5.3); and if we define the concentration ratio of Gini from the mean difference with repetition, it's easy to obtain the following formula

$$R_r = \frac{\Delta_r}{2\bar{x}} = \frac{2Cov(X,F)}{\bar{x}}{}^{69},$$

where $F_h = \frac{h}{n}$, for $h = 1,2,...,n$.

Also Gini used the relation between R and the mean difference without repetition Δ, to obtain two new expressions of formula (5.11).

The first one:

$$\Delta_E = \Delta - \frac{\sum_{k=1}^{r} n_k(n_k - 1)\Delta_k}{n(n-1)},$$

[68] Gini will raise this problem in the paper "*Intorno alle curve di concentrazione*" that was presented in the 20th International Congress of Statistic that took placed in Madrid (September, 1931)

[69] This expression of the concentration ratio of Gini is the discrete version of the continuous case.

corresponds to formula (23) of the Gini paper, section **9**, being Δ_E the mean difference without repetition of the mean values of each class interval and taking the total of individuals of each interval[70] as the absolute frequencies.

The symbol Δ is the mean difference without repetition of all the data, and Δ_k is the mean difference without repetition of the truncate data in the interval k-th[71].

The second one:

$$R' = R - \frac{\sum_{k=1}^{r} n_k (n_k - 1)\Delta_k}{2(n-1)A_n}^{72}, \tag{5.12}$$

where $R' = R_E$. This formula (5.12) has been used in [Aghevli and Mehran, 1981] as an optimization criterion to group the data of the income in intervals.

5.6 An explanation of how Gini connected the double concentration area with the relative mean difference

It can be wondered how Gini connected his variability measure and his mean difference with the concentration area.

Let's see an explanation

In his book in 1912, Gini deduced the following formula,

$$\frac{\Delta_r}{2\mu} = \frac{1}{2\alpha - 1}.$$

Also in the same book, Gini showed the following relation:

$$\delta = \frac{\alpha}{\alpha - 1},$$

that connected the above formula with δ, that is to say

$$\frac{\Delta_r}{2\mu} = \frac{\delta - 1}{\delta + 1}.$$

as the dual Lorenz curve associated to Model I of Pareto was,

$$\tilde{q} = \tilde{p}^{\frac{1}{\delta}}, \quad \tilde{q} = 1 - q, \quad \tilde{p} = 1 - p.$$

Finally, we obtain

---

70 It's the mean difference without repetition "between".
71 It's the mean difference without repetition "intra".
72 This formula corresponds to formula (24) of Gini.

$$2(ConcentrationArea) = 2\left(\int_0^1 \widetilde{p}^{\frac{1}{\delta}} d\widetilde{p} - \frac{1}{2}\right) = \frac{\delta - 1}{\delta + 1} = \frac{\Delta_r}{2\mu}.$$

Some of the results showed in this paper have been obtained from Gini's book (1912) and we have gathered them in sections 5.3 and 5.4. In section **12** of the 1914 paper Gini gathered, for Model I of Pareto, the formula $R = \frac{1}{2h-3} = \frac{1}{2\alpha - 1}$, where $h = \alpha + 1$. All these events reinforce that the calculations gathered in this paper led Gini to connect the double of the concentration area with its relative mean difference.

References

Aghevli, N. B.; and Mehran, F. (1981): *Optimal Grouping of Income Distribution Data*. J. American Statistical Association, vol. LXXVI, n. 373, p.22-26.

Barbut, M. (2007): *La mesure des inégalités. Ambigüités et paradoxes*. Librairie Droz, Genève-Paris.

Basulto, J.; Busto, J.; y Sánchez, R. (2009): *El concepto de desigualdad en Vifredo Pareto (1848-1923).* V Congreso Internacional de Historia de la Estadística y de la Probabilidad de España, Santiago de Compostela.

Berrebi, Z. M. and J. Silber (1987). *Dispersión, Asymmetry and the Gini Index of Inequality*. Internacional Economics Review, vol. 28, No. 2, June, p.331-338.

Bortkiewicz, L. v. (1931): *Die Disparitätsmasse der Einkommensstatistik.* XII Session de L'Institut International de Statistique. Tokio, p.189-298.

Chatelain, E. (1907): *Les sucession declarées en 1905*. Revue politique et parlamentaire, Paris.

Chatelain, E. (1910): *La trace de la curbe des sucessions en France*. Journal de la Societé de Statistique de Paris. Paris, p. 362 y siguientes.

Chatelain, E. (1911): *La fortune française d'aprés les sucessions en 1909*. La Democratie. Paris, *20* Janvier.

David, H.A. (1968): *Gini's mean difference rediscovered*. Biometrika, p. 573-574.

David, H.A. (1998): *Early sample measures of variability*. Statistical Science, vol. 13, nº 4, p.368-377.

Gini, C. (1909): *Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza.* Giornale degli Economisti, anno XX (serie II), n. 1, p. 27-83.

Gini, C. (1910): *Índice di Concentrazione e di Dipendenza. Biblioteca dell'Economista, serie V, vol. XX*, Utet, Torino.

Gini, C. (1912): Variabilità e Mutabilità: contributo allo Studio delle distribuzioni e delle relazioni statistiche. Facoltá di Giurisprudenza della R. Universitá dei Cagliari, anno III, parte 2ª.

Gini, C. (1914): *Di una misura della concentrazione indipendente dalla distribuzione del carattere*. Atti del R. Istituto Veneto di Scienze, Lettere ed Arti, tomo LXXIII, parte II, pags. p.1203-1258.

Gini, C. (1931): *Intorno alle curve di concentrazione.* Bullitin del ISI, tomo XXVI, segunda entrega, p.423-484.

Imedio Olmedo, L. (2007): *Algunas consideraciones sobre el índice de Bonferroni*. Estadística Española, vol. 49, Num. 164, p. 103-135.

King, W. J. (1912): *The elements of statistical methods*. New York, The Macmillan Company.

Lerman and Yitzhaki, S. (1985): *Income inequality effects by income sources: approach and applications to the United Status*, The Review of Economic and Statistics, vol. 67, No. 1, p.151-156.

Lorenz, M. O. (1905): *Methods of Measuring the Concentration of Wealth*. American Statistical Association. Vol. IX, n. 70, June, p. 209-219.

Pareto, V. (1895): *La legge della domanda*. Giornale degli Economisti. Janvier, p.59-68.

Pareto, V. (1896a): *La courbe de la répartition de la richesse*. Université de Lausanne. Faculté de Droit à l'occasion de l'Exposition nacional suisse, Genève, Lausanne, CH. Viret-Geton Impr., p.373-387.

Pareto, V. (1896b): *La courbe des revenus.* Le monde économique, 25 juillet, p.127-137.

Pareto, V. (1897): *Cours d'économie politique.* Tomo II, Rouge, Lausanne.

Pareto, V. (1909): *Manuel d'Economie politique*. Traducción del libro *Manuel di economia politica con una introduzione alla sienza social* (1906). Società Editrice Libraria.

Pareto, V. (1965): *Écrits sur la courbe de la répartition de la richesse.* Genève. Librairie Droz, Edición de 1967.

Person, W.M. (1909): *The variability in the distribution of wealth and income*. The Quarterly Journal of Economics, vol. XXIII, N. 3.

Pollastri, A. (1990): *A Comparison of the tradicional estimators of parameter* α *the Pareto distribution*. Sudies in Contemporary Economics, Income and Wealth Distribution, Inequality and Poverty, Camilo Dagun y Michele Zenga (Eds). Springer-Verlag.

Mortara, G. (1911): *Note di economia inductiva (sulla distribuzione dei redditt)*. Giornale degli Economisti e Rivista di Statistica, Serie terza, anno XXII, vol. XLII, p. 455-471.

Sastry V.S. and Ujwala R. K. (1994): *Note on the descomposition of Gini inequality*, The Review of Economic and Statistics, vol. 76, No. 3, p.584-586.

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

Sen, A. (1973): *On economic inequality*. Expanded edition with a substantial annexe by J. E. Foster and A. Sen (1997). Claredon Press. Oxford.

Watkings, G. P. (1905). *Comment on the method of measuring concentration of wealth.* Publications of American Statistical Association. *n. 72,* December.

Watkings, G. P. (1908): *An interpretation of certain statistical evidence of concentration of wealth.* Publications of American Statistical Association. *n. 82,* March.

Watkings, G. P. (1909): *The measuring of concentrations of wealth*. Publications of American Statistical Association. *vol*. XXIV*,* n. 1*, November*