J Econ Inequal (2012) 10:421–443
DOI 10.1007/s10888-011-9188-x

The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini

Lidia Ceriani · Paolo Verme

Received: 20 January 2011 / Accepted: 23 March 2011 / Published online: 10 June 2011

Abstract The scope of this paper is to celebrate the 100th anniversary of the Gini index by providing the original formulae. Corrado Gini introduced his index for the first time in a 1912 book published in Italian under the name of "Variabilità e Mutabilità" (Variability and Mutability). This article provides selected extracts of Part I of the book dedicated to measures of variability. We find that Gini proposed no less than 13 formulations of his index, none of which is known today to the large public. We also find that Gini anticipated some of the developments that derived from the study of his index.

Keywords Inequality · Income distribution · Corrado Gini · Gini index

1 Introduction

One hundred years ago (1912), Corrado Gini (1884–1965) published the book "Variability and Mutability" (*"Variabilità e Mutabilità"*, [2]) where he presented for the first time the index that today is known as the "Gini Index". The original book published in Italian has never been translated into English and the Gini index became known in the anglophone world probably because Gini engaged in a brief exchange with Hugh Dalton in the Economic Journal in 1920–1921 [1, 4]. Gini also discussed

L. Ceriani (✉)
Università Bocconi, Milan, Italy
e-mail: lidia.ceriani@unibocconi.it

P. Verme
The World Bank, Washington, DC, USA
e-mail: pverme@worldbank.org

P. Verme
Università di Torino, Turin, Italy

his index and its relation with the Lorenz curve in a 1914 article [3] that was later translated into English [5]. However, despite the numerous formulations of the Gini index that have appeared in the literature over the years, the original formulations are largely unknown.

As suggested by the title, Variability and Mutability is divided into two parts, the first discussing indices of variability—including various versions of the Gini index—and the second discussing indices of mutability. Gini explains the difference between the two classes of indices as the first (variability) being devoted to the measurement of quantitative phenomena and the second (mutability) being devoted to the measurement of qualitative phenomena.

This article focuses exclusively on the first part of the book where Gini introduces and discusses indices of variability. In this part of the book, Gini defines his index as *"the mean difference from all observed quantities"*, shows the crucial differences between his index and popular measures of variability until then (the simple or probabilistic mean deviation from the median; and the simple, the square or the probabilistic mean deviation from the arithmetic mean), provides different formulations to easily compute his index with various types of data and indicates the fields where each formulation might be more appropriate to use. The text is enriched with many curious and practical applications that make the reading pleasant and surprisingly modern.

Gini himself was a rather extraordinary man of culture with degrees in law, mathematics and biology, interests spanning across the social and natural sciences, excellent knowledge of various languages and a very extensive publication record. One biography[1] claims that Gini authored more than 800 publications, 56 of which can be found today on JSTOR. These include articles in almost all major international journals in economics, statistics, political science and sociology, an achievement that very few social scientists have been able to match since and that Harvard University rewarded with a honorary doctoral degree of science in 1936. Gini published "Variability and Mutability" at the age of 28, two years after he obtained the Chair of Statistics from the University of Cagliari.

The scope of this paper is to celebrate the 100th anniversary of the Gini index, to provide the original formulations of the index and to illustrate how Gini anticipated some of the developments that derived from the study of his index. Many of the insights present in the book have been largely overlooked probably because the book was never translated into English and did not reach an international audience. We attempt here to recover some of the highlights from the original publication.

A few notes on the text of this article. The original text of the part on "Variability" counts 8 sections, 53 paragraphs and 94 pages rich in examples and mathematical demonstrations. The paper provides a mix of summaries of the various sections and literal translations, where we thought that the original text should be preserved. We also needed to be selective with the mathematical expressions and we opted to focus on the parts where the various forms of the Gini index appear. All mathematical expressions are reported as in the original text. For clarity and for easy reference to the original book, the numbers of the original paragraphs have been preserved. The section titles include translations of the original titles while the section numbering

[1] http://www.umass.edu/wsp/statistics/tales/gini.html

and the numbering of equations is our own. Table 1 provides a summary of all different formulations of the Gini index found in the original text.

2 Different aspects in the study of the variability of characters (*Diversi aspetti nello studio della variabilità dei caratteri*)

11 The analysis of variability, Gini explains, is concerned with two different categories of characters (objects). The first category is made of characters that have the same intensity during observation but that, due to measurement errors, can result in different measures of the same intensity. Natural sciences are typically concerned with these characters. For example, the measurement of a mountain's height can result in slightly different values with repeated observations because measurement instruments can hardly be perfect. The second category is made instead of characters

Table 1 Summary of the different formulations of the Gini index

Form	Δ	Δ_R	Equations
I	$\frac{2}{n(n-1)}\sum_{i=1}^{\frac{n+1}{2}}(n+1-2i)(a_{n-i+1}-a_i)$	$\frac{2}{n^2}\sum_{i=1}^{\frac{n+1}{2}}(n+1-2i)(a_{n-i+1}-a_i)$	(1), (2)
II	$\frac{2}{n(n-1)}\sum_{i=1}^{\frac{n+1}{2}} d_{i,n-i+1}(a_{n-i+1}-a_i)$	$\frac{2}{n^2}\sum_{i=1}^{\frac{n+1}{2}} d_{i,n-i+1}(a_{n-i+1}-a_i)$	(4), (5)
	$\frac{1}{n(n-1)}\sum_{i=1}^{n} d_{i,n-i+1}\lvert a_i-a_{n-i+1}\rvert$	$\frac{1}{n^2}\sum_{i=1}^{n} d_{i,n-i+1}\lvert a_i-a_{n-i+1}\rvert$	(6), (7)
III	$\frac{4}{n(n-1)}\sum_{i=1}^{n} d_{i,M}\lvert a_i-M\rvert$	$\frac{4}{n^2}\sum_{i=1}^{n} d_{i,M}\lvert a_i-M\rvert$	(13), (14)
IV	$\frac{4}{n(n-1)}\sum_{k=1}^{s} d_{kM}\, f_k\lvert x_k-M\rvert$	$\frac{4}{n^2}\sum_{k=1}^{s} d_{kM}\, f_k\lvert x_k-M\rvert$	(15), (16)
V	$\frac{n}{n-1}\frac{\sum_{i=1}^{n} d_{i,n-i+1}\lvert a_i-a_{n-i+1}\rvert}{2\sum_{i=1}^{n} d_{i,n-i+1}}$	$d\frac{\sum_{i=1}^{n} d_{i,n-i+1}\lvert a_i-a_{n-i+1}\rvert}{2\sum_{i=1}^{n} d_{i,n-i+1}}$	(19), (20)
VI	$\frac{n}{n-1}\frac{\sum_{i=1}^{n} d_{i,M}\lvert a_i-M\rvert}{\sum_{i=1}^{n} d_{i,M}}$	$\frac{\sum_{i=1}^{n} d_{i,M}\lvert a_i-M\rvert}{\sum_{i=1}^{n} d_{i,M}}$	(22), (23)
VII	$\frac{2}{n(n-1)}\sum_{h=1}^{n}(T_h-F_h a_h)$	$\frac{2}{n^2}\sum_{h=1}^{n}(T_h-F_h a_h)$	(37), (36)
VIII	$\sqrt[m]{\frac{1}{n(n-1)}\sum_{k=1}^{n}\sum_{i=1}^{n}\lvert a_i-a_k\rvert^m}$	$\sqrt[m]{\frac{1}{n^2}\sum_{k=1}^{n}\sum_{i=1}^{n}\lvert a_i-a_k\rvert^m}$	(38), (39)
IX	${}^2\Delta_R=\sqrt{2}({}^2S_A)$	${}^2\Delta=\sqrt{\frac{2n}{n-1}}({}^2S_A)$	(43), (42)
X	$\frac{H(n+1)}{3}$	$\frac{H(n^2-1)}{3n}$	(51), (52)
XI	$\frac{4}{n(n-1)}\frac{(n-1)H^{\frac{n+3}{2}}-(n+1)H^{\frac{n+1}{2}}+H^{\frac{3}{2}}+H^{\frac{1}{2}}}{(H-1)^2}$	$\frac{4}{n^2}\frac{(n-1)H^{\frac{n+3}{2}}-(n+1)H^{\frac{n+1}{2}}+H^{\frac{3}{2}}+H^{\frac{1}{2}}}{(H-1)^2}$	(57), (58)
	$\frac{4}{n(n-1)}\frac{(n-1)H^{\frac{n+3}{2}}-(n+1)H^{\frac{n+1}{2}}+2H}{(H-1)^2}$	$\frac{4}{n^2}\frac{(n-1)H^{\frac{n+3}{2}}-(n+1)H^{\frac{n+1}{2}}+2H}{(H-1)^2}$	(61), (62)
XII	$\frac{n(2s-1)!}{(n-1)2^{2s-1}(s-1)!(s-1)!}$	$\frac{(2s-1)!}{2^{2s-1}(s-1)!(s-1)!}$	(70), (69)
XIII	$\frac{2}{2h-3}\frac{n}{n-1}A$	$\frac{2}{2h-3}A$	(78), (77)

that can be measured with precision. Social sciences sometimes measure phenomena of such kinds. For example, population and income can be measured with finite and precise quantities if proper data are available. Clearly, the questions to address with these two different classes of characters are different. The question to address with the first class of characters is *"How much the different measures differ from the real value"*. The problem to address with the second class is, *"How much the different objects differ from each other"*.

12 Astronomers were the first to develop methods to study variability within the first class of characters. Since measurement errors have by definition the same probability of having positive or negative values, the arithmetic mean of the different measures can be interpreted as the most probable effective value of a character that cannot be measured with precision. In this case, an index that computes the difference between quantities and their arithmetic mean is an appropriate index of variability and so are indices such as the quadratic mean deviation, the simple mean deviation and the probabilistic deviation from the arithmetic mean.

13 However, one should ask the question—Gini argues—of whether the same class of measures should be used to answer the second question. Since the questions to answer with exact and inexact measures are different, the indices that measure the variability of such measures should also be different. It is therefore necessary to find an appropriate way to measure the intensity of the difference between the magnitudes of precisely known quantities.

3 The mean difference between several quantities (*La differenza media tra più quantità)*

14 The aim of this section, Gini writes, is to find a formula that can explain the arithmetic mean of differences between n quantities: $a_1, a_2, \ldots, a_{n-2}, a_{n-1}, a_n$, where $n \in \mathbb{N}$. Quantities are assumed to be in ascending order in such a way that $a_{i-1} \leq a_i$ for each $i = 1, 2, \ldots, n$. Starting from these definitions, Gini provides a page of calculus that leads to the first formulation of his index:

$$\Delta = \frac{2}{n(n-1)} \sum_{i=1}^{\frac{n+1}{2}} (n + 1 - 2i)(a_{n-i+1} - a_i) \tag{1}$$

We will call Δ the **mean difference between the** n **quantities** (*"Chiameremo* Δ *differenza media tra le n quantità".*)

15 The sum of the $n(n-1)$ differences between each quantity and all the others is the same as the sum of the n^2 differences between each quantity and all quantities (i.e. we consider also the difference between each quantity and itself, which is zero). Therefore, the average of this n^2 differences is:

$$\Delta_R = \frac{2}{n^2} \sum_{i=1}^{\frac{n+1}{2}} (n + 1 - 2i)(a_{n-i+1} - a_i) \tag{2}$$

We will call Δ_R the **mean difference with repetition between the n quantities** (*Chiamiamo Δ_R differenza media con ripetizione tra le n quantità)*. It follows that

$$\Delta_R = \frac{n-1}{n}\Delta. \tag{3}$$

According to Gini, both measures have a particular utility. The mean difference between n quantities is most appropriate for measuring differences between exact measures in social sciences such as income while the mean difference with repetition can be used to show the relation between the mean difference and the average difference from the median or from the mean and has different applications in statistics (an issue that Gini explores further, paragraphs 20, 21, 23, and 24).

16 Next, Gini introduces the notion of degree (*grado*), what we would generally refer to as *rank* in the modern welfare jargon. Defining *rank* as the positional distance between two consecutive quantities, then the *rank-distance* between two quantities a_i a_t is $d_{i,t}$, or the number of ranks between these quantities. Also, when $i + t = n + 1$, i.e. $t = n - i + 1$, we refer to a_i and a_t as *symmetric quantities*. Now, if we define $d_{i,n-i+1} = n + 1 - 2i$ the *the rank-distance between two symmetric quantities* a_i and a_{n-i+1}, Eq. 1 and 2 can be rewritten as follows:

$$\Delta = \frac{2}{n(n-1)} \sum_{i=1}^{\frac{n+1}{2}} d_{i,n-i+1}(a_{n-i+1} - a_i) \tag{4}$$

$$\Delta_R = \frac{2}{n^2} \sum_{i=1}^{\frac{n+1}{2}} d_{i,n-i+1}(a_{n-i+1} - a_i) \tag{5}$$

or also as:

$$\Delta = \frac{1}{n(n-1)} \sum_{i=1}^{n} d_{i,n-i+1}|a_i - a_{n-i+1}| \tag{6}$$

$$\Delta_R = \frac{1}{n^2} \sum_{i=1}^{n} d_{i,n-i+1}|a_i - a_{n-i+1}| \tag{7}$$

17 At this point, Gini provides an empirical example of his indices using a sample of meat prices in the Parisian meat market between 1867 and 1910. The example shows how, by arranging the prices series into two symmetric series (split by the median price), one can easily compute manually the mean difference and the mean difference with repetition.[2]

18 The example described above also introduces this paragraph where Gini presents his indices in the form of distances from the median. Call M the median

[2]In 1912 statisticians were understandably concerned with the degree of computability of indices in addition to the characteristics of such indices.

value between n ordered quantities. In case n is even, the distance between M and $a_{\frac{n}{2}}$ or $a_{\frac{n}{2}+1}$ is assumed to be half-rank. Then, the absolute difference between two symmetric quantities is equal to the sum of absolute differences between the median value and such quantities:

$$|a_{n-i+1} - a_i| = |a_i - M| + |a_{n-i+1} - M| \tag{8}$$

Therefore:

$$\sum_{i=1}^{n} |a_{n-i+1} - a_i| = 2 \sum_{i=1}^{n} |a_i - M| \tag{9}$$

Moreover, the rank-distance between two symmetric quantities is equal to two times the rank-distance between one of those quantities and the median value:

$$d_{i,n-i+1} = 2d_{i,M} = 2d_{n-i+1,M} \tag{10}$$

Therefore:

$$\sum_{i=1}^{n} d_{i,n-i+1} = 2 \sum_{i=1}^{n} d_{i,M} \tag{11}$$

and:

$$\sum_{i=1}^{n} d_{i,n-i+1} |a_{n-i+1} - a_i| = 4 \sum_{i=1}^{n} |a_i - M| \tag{12}$$

Hence, we can rewrite Eq. 6 as:

$$\Delta = \frac{4}{n(n-1)} \sum_{i=1}^{n} d_{i,M} |a_i - M| \tag{13}$$

and Eq. 7 as:

$$\Delta_R = \frac{4}{n^2} \sum_{i=1}^{n} d_{i,M} |a_i - M| \tag{14}$$

19 Next, Gini tackles the question of **repeated values** in a series. In such cases, the author argues that Eqs. 13 and 14 are the best choice to compute the mean difference and the mean difference with repetition. To illustrate this point, Gini changes notation as follows. Let s be the different values assumed by the n quantities; let x_k $(k = 1, 2, \ldots, s)$ be the $k-$th value and let f_k be the number of quantities assuming value x_k. The rank-distance between value x_k and the median M, $d_{k,M}$, is the the mean of rank-distances between M and the f_k quantities assuming value x_k. We can therefore rewrite Eqs. 13 and 14 as:

$$\Delta = \frac{4}{n(n-1)} \sum_{k=1}^{s} d_{kM} f_k |x_k - M| \tag{15}$$

$$\Delta_R = \frac{4}{n^2} \sum_{k=1}^{s} d_{kM} f_k |x_k - M| \tag{16}$$

An empirical example follows these formulations and illustrates the computation of such indices. This time Gini uses anthropometric measures of conscripted soldiers born in Italy between 1859 and 1863.

20 Consider now the case when i assumes all values between 1 and n (the case with **no repeated values)**. For n even, $d_{i,n-i+1}$ assumes twice all values of the odd numbers smaller than n. For n odd, $d_{i,n-i+1}$ assumes twice all values of the even numbers smaller than n. The sum of all odd numbers smaller than an even n is equal to $\frac{n^2}{4}$, and the sum of all odd numbers smaller than an odd n is equal to $\frac{n^2-1}{4}$ which, for n big enough, can be approximated back to $\frac{n^2}{4}$. Therefore, for n big enough (even or odd), we obtain the following:

$$2\frac{n^2}{4} = \sum_{i=1}^{n} d_{i,n-i+1} \tag{17}$$

that is equal to writing

$$n^2 = 2\sum_{i=1}^{n} d_{i,n-i+1}. \tag{18}$$

And therefore, we can rewrite Eqs. 6 and 7 as:

$$\Delta = \frac{n}{n-1}\frac{\sum_{i=1}^{n} d_{i,n-i+1}|a_i - a_{n-i+1}|}{2\sum_{i=1}^{n} d_{i,n-i+1}} \tag{19}$$

$$\Delta_R = \frac{\sum_{i=1}^{n} d_{i,n-i+1}|a_i - a_{n-i+1}|}{2\sum_{i=1}^{n} d_{i,n-i+1}} \tag{20}$$

Equation 20 shows that the mean difference with repetition between n quantities equals one half the weighted average of differences between symmetric quantities, where weights are proportional to the rank-distance between each pair of quantities.

21 From Eqs. 18 and 10 we obtain:

$$n^2 = 4\sum_{i=1}^{n} d_{i,M} \tag{21}$$

which allows us to reformulate Eqs. 13 and 14 as follows:

$$\Delta = \frac{n}{n-1}\frac{\sum_{i=1}^{n} d_{i,M}|a_i - M|}{\sum_{i=1}^{n} d_{i,M}} \tag{22}$$

$$\Delta_R = \frac{\sum_{i=1}^{n} d_{i,M}|a_i - M|}{\sum_{i=1}^{n} d_{i,M}} \tag{23}$$

Equations 22 and 23 are exact for n even and approximated to less than $\frac{1}{n^2}$ for n odd. Equation 23 shows that the mean difference with repetition between n quantities equals one half the weighted average of differences from the median value, where weights are proportional to the rank-distance between each quantity and the median.

This is an important result because it underlines that the mean difference between n quantities differs from the mere average difference from the median, since it gives more weight to larger differences. This is true also if we compare the average difference from the median with the quadratic average difference from the median. In the case of the quadratic average difference, the weight is proportional to the magnitude of the difference and, in case of the mean difference, the weight is proportional to the rank-distance between the quantity and the median.

22 It is known, Gini writes, that the simple mean deviation from the median is minimum as compared to the simple mean deviation from any other quantity.[3]

To determine the magnitude of the difference between the simple mean deviation from the median and the simple mean deviation from any other quantity, let H be any other quantity different from the median. By definition, therefore, the number of the quantities in the rank above H is not equal to the number of the quantities in the rank below H, but it is different from the number of the quantities in the rank below H by some K (which can be either positive or negative).

Consider first the case where n is odd and K is even. The simple mean deviation from the median M is:

$$ {}^1S_M = \frac{1}{n}\left(\sum_{i=\frac{n+1}{2}}^{n} a_i - \sum_{i=1}^{\frac{n+1}{n}} a_i\right) $$

while the simple mean deviation from H, for $H > M$ is:

$$ {}^1S_H = \frac{1}{n}\left[\sum_{i=\frac{n+1+K}{2}}^{n} a_i - \left(\frac{n+1-K}{2}H\right) + \left(\frac{n+1+K}{2}H\right) - \sum_{i=1}^{\frac{n+1+K}{2}} a_i\right] $$

which is equal to:

$$ {}^1S_H = \frac{1}{n}\left(\sum_{i=\frac{n+1}{2}}^{n} a_i - \sum_{i=\frac{n+1}{2}}^{\frac{n+1+K}{2}-1} a_i - \sum_{i=1}^{\frac{n+1}{2}} a_i - \sum_{i=\frac{n+1}{2}+1}^{\frac{n+1+K}{2}-1} a_i + KH\right) $$

from which we obtain:

$$ {}^1S_H = {}^1S_M + \frac{1}{n}\left(HK - 2\sum_{\frac{n+1}{2}}^{\frac{n+1}{2}+\frac{K-2}{2}} a_i + M - H\right) $$

Analogously, it is possible to demonstrate that, given $H < M$:

$$ {}^1S_H = {}^1S_M + \frac{1}{n}\left(2\sum_{\frac{n+1}{2}}^{\frac{n+1}{2}+\frac{K-2}{2}} a_i - HK + H - M\right). $$

[3] Laplace [6], *Théorie analytique des probabilités*, Paris, Courcier, II Supplément, pp. 42–43.

The last two formulas can be written as

$$ {}^1S_H = {}^1S_M + \frac{2}{n} \sum_{\frac{n+1}{2}}^{\frac{n+1}{2} \pm \frac{K-2}{2}} |a_i - H| - \frac{1}{n} |H - M| \tag{24} $$

where the sign is equal to + if $H > M$ and − if $H < M$. In the same way, for n odd and K odd:

$$ {}^1S_H = {}^1S_M + \frac{2}{n} \sum_{\frac{n+1}{2}}^{\frac{n+1}{2} \pm \frac{K-1}{2}} |a_i - H| - \frac{1}{n} |H - M| \tag{25} $$

For n even and K odd

$$ {}^1S_H = {}^1S_M + \frac{2}{n} \sum_{\frac{n+1}{2} \pm \frac{1}{2}}^{\frac{n+1}{2} \pm \frac{K-2}{2}} |a_i - H| - \frac{1}{n} |H - M| \tag{26} $$

and, finally, for n even and K even

$$ {}^1S_H = {}^1S_M + \frac{2}{n} \sum_{\frac{n+1}{2} \pm \frac{1}{2}}^{\frac{n+1}{2} \pm \frac{K-2}{2}} |a_i - H| - \frac{1}{n} |H - M| \tag{27} $$

Therefore, the simple average difference between all quantities and any quantity in a series differs from the simple mean deviation between all quantities and the median value for an amount that is not in relation with the rank difference of individual quantities from the median. Hence, a simple mean deviation between any quantities cannot be a precise measure of the mean difference between quantities.

23 It is also known, Gini continues, that the quadratic mean deviation from the arithmetic mean is minimum with respect to the quadratic mean deviation from any other quantity. Let H be any quantity, let A be the arithmetic mean, and let ${}^2S_H$ and ${}^2S_A$ be the respective quadratic mean deviation. We obtain:

$$ {}^2S_H^2 = {}^2S_A^2 + (A - H)^2 \tag{28} $$

The quadratic mean deviation from any quantity is related to the quadratic mean deviation from the median by the following:

$$ {}^2S_H^2 = {}^2S_M^2 + (A - H)^2 - (A - M)^2 \tag{29} $$

or:

$$ {}^2S_H^2 = {}^2S_M^2 + (H - M)(H + M - 2A) \tag{30} $$

The quadratic mean deviation from any quantity is different from the quadratic mean deviation from the median and this, in turn, is different from the simple mean deviation from the median by a quantity which is not in relation with the rank difference of the single quantities from the median. Therefore, Gini concludes, the quadratic mean deviation from the arithmetic mean cannot result, in general, in a precise measure of the average difference between quantities.

24 There is another easy way to express the mean difference and the mean difference with repetition. The mean difference of all n quantities from the $h-$th one is:

$$ {}^1S_h = \frac{1}{n}\left\{\sum_{k=h}^{n}(a_k - a_h) + \sum_{k=1}^{h-1}(a_h - a_k)\right\} \tag{31} $$

which can also be expressed as:

$$ {}^1S_h = \frac{1}{n}\left\{2\sum_{k=h}^{n} a_k - \sum_{k=1}^{n} a_k - na_h + 2(h-1)a_h\right\} \tag{32} $$

Let

$$ F_h = n - h + 1; \quad T_h = \sum_{k=h}^{n} a_k; \quad T_1 = \sum_{k=1}^{n} a_k \tag{33} $$

Therefore

$$ {}^1S_h = \frac{1}{n}\left(2T_h - T_1 + na_h - 2F_h a_h\right) \tag{34} $$

Let now h assume all values between 1 and n. The mean of the $h = 1, 2, \ldots, n$ values ${}^1S_h$ is the mean difference with repetition of the n quantities. It is immediate to see that

$$ \Delta_R = \frac{1}{n^2}\left(2\sum_{h=1}^{n} T_h - nT_1 + nT_1 - 2\sum_{h=1}^{n} F_h a_h\right) \tag{35} $$

From which we obtain the mean difference with repetition

$$ \Delta_R = \frac{2}{n^2}\sum_{h=1}^{n}(T_h - F_h a_h) \tag{36} $$

and the mean difference without repetition

$$ \Delta = \frac{2}{n(n-1)}\sum_{h=1}^{n}(T_h - F_h a_h)\,. \tag{37} $$

For characters distributed along certain curves (here Gini does not specify what curves), it is convenient to use the formulations in Eqs. 36 and 37 to calculate the mean difference values.

25 We can think of infinite mean differences between n quantities, which correspond to the infinite values that m can assume in the following Eq. 38:

$$ {}^m\Delta = \sqrt[m]{\frac{1}{n(n-1)}\sum_{k=1}^{n}\sum_{i=1}^{n}|a_i - a_k|^m} \tag{38} $$

and, analogously, we can think of infinite mean differences with repetition :

$$ {}^{m}\Delta_R = \sqrt[m]{\frac{1}{n^2}\sum_{k=1}^{n}\sum_{i=1}^{n}|a_i - a_k|^m} \tag{39} $$

Up to now we have considered the simple mean deviation and the simple mean deviation with repetition with $m = 1$. We now consider the quadratic mean deviation and the quadratic mean deviation with repetition that we obtain with $m = 2$.

These values of ${}^{2}\Delta$ and ${}^{2}\Delta_R$ are in a simple relation with the quadratic mean deviation from the arithmetic mean ${}^{2}S_A$. Let ${}^{2}S_h$ be a quadratic mean deviation of the n quantities from the $h-$th quantity a_h. This can be expressed as a function of the quadratic mean deviation from the arithmetic mean ${}^{2}S_A$:

$$ {}^{2}S_h^2 = {}^{2}S_A^2 + (A - a_h)^2 \tag{40} $$

Let h take all values from 1 to n. The mean of the n values of ${}^{2}S_h^2$ will correspond to the mean of the squares of the n^2 differences between each quantity and all n quantities of the series. Therefore:

$$ {}^{2}\Delta_R^2 = \frac{1}{n}\sum_{h=1}^{n}{}^{2}S_h^2 = {}^{2}S_A^2 + \frac{1}{2}\sum_{h=1}^{n}(A - a_n)^2 = 2({}^{2}S_A^2) \tag{41} $$

from which

$$ {}^{2}\Delta_R = \sqrt{2}({}^{2}S_A) \tag{42} $$

$$ {}^{2}\Delta = \sqrt{\frac{2n}{n-1}}({}^{2}S_A) \tag{43} $$

4 On indices of variability of characters in the case of partial series *(Degli indici di variabilità dei caratteri nel caso di seriazioni parziali)*

26 In this section, Gini discusses the issue of the partial observation of values of a series, due to a random extraction of m observed values out of a complete series of n values. The question is whether and how much the square mean difference from the arithmetic mean, the mean difference and the mean difference with repetition computed on the partial observations differ from the ones computed on the complete series. Moreover, Gini stresses the difference between random extraction with or without repetition. Measurement errors of physical quantities are of the first kind while statistical observations of biological, anthropological, demographic or economic phenomena are of the second type.

27–28 In the following two paragraphs, Gini demonstrates the relation between the square mean difference from the arithmetic mean computed on partial observations

($^2S^2_{A_p}$) and the square mean difference from the arithmetic mean computed on the complete series ($^2S^2_{A_g}$). In the case of partial observations generated by random extraction with repetition:

$$^2S^2_{A_p} = \frac{m-1}{m}\,^2S^2_{A_g} \tag{44}$$

In the case of partial observation generated by random extraction without repetition:

$$^2S^2_{A_p} = \frac{m-1}{m}\frac{n}{n-1}\,^2S^2_{A_g} \tag{45}$$

29 Next, Gini demontrates the relation between the mean difference computed on partial observations (Δ_p) and the mean difference computed on the complete series (Δ_g). In the case of partial observations generated by random extraction without repetition:

$$\Delta_p = \Delta_g \tag{46}$$

and for the mean difference with repetition:

$$\Delta_{Rp} = \frac{m-1}{m}\frac{n}{n-1}\Delta_{Rg} \tag{47}$$

In the case of partial observations generated by random extraction with repetition, we obtain for the mean difference:

$$\Delta_p = \frac{n-1}{n}\Delta_g \tag{48}$$

and for the mean difference with repetition:

$$\Delta_{Rp} = \frac{m-1}{m}\Delta_{Rg} \tag{49}$$

Therefore, Gini shows that to pass from the partial to the complete series, the same coefficient applies to the square mean difference from the arithmetic mean and to the mean difference with repetition (but not to the mean difference).

30 According to Gini, the purpose of the analysis of variability is twofold. On the one hand, one can be interested in analyzing how the characters have changed within the time and space boundaries to which the observed values belong. On the other hand, we may be interested in checking the varying trend of the character. In the first case, Gini talks of *concrete variability* (*variabilità concreta*), in the second case of *limit variability* (*variabilità limite*). To obtain the *limit variability* value of the variability indices computed on the complete series, one needs to use formulas 45, 46 and 47 replacing m with n and n with ∞.

5 On indices of variability of characters in the case of parallel series *(Degli indici di variabilità dei caratteri nel caso di seriazioni parallele)*

31 Let ${}_A f_k$ and ${}_B f_k$ be the number of times characters A and B assume the intensity x_k, $(k = 1, 2, \ldots, s)$. The two series are *equal* if ${}_A f_k = {}_B f_k, \forall x_k$. Two series are instead

parallel if ${}_A f_k = H_B f_k$, where H is a constant. Two parallel series are in fact parallel curves when represented in logarithmic scale.

32 Two parallel series have the same simple mean deviation, quadratic mean deviation and mean difference with repetition. However, the less numerous series has a larger mean difference.

33 Gini shows next the desirability of this last property. In two parallel series, he says, the different values of x_k have the same ratio (the series 1, 2, 2 has the same ratio between 1s and 2s as the series 1, 1, 2, 2, 2, 2). But in the most numerous series, there are more observations assuming the same value. Therefore, in this last case, the variability should be lower.

6 On indices of variability of characters in some types of series *(Degli indici di variabilità dei caratteri in alcuni tipi di seriazioni)*

34 In the following paragraphs, Gini determines the value of the simple mean deviation, the quadratic mean deviation and the mean difference for a set of specific series. The scope of this section is to point out whether a constant relation among these measures occurs, or, on the contrary, the relation varies with the type of series. And, for the same type of series, whether the relation among these measures is constant or varies with the number of observations or the intensity of the character's variability.

35 The first type of series considered by Gini is the one for which the intensity of the differences between each character and the median grows arithmetically with the rank-distance between each character and the median. In other words, $\|a_i - M\| = Hd_{i,M}$ where $H \gtreqless 1$ is a constant. Therefore, the terms of the series form an arithmetic progression with common difference H. The following formulas (50)–(54) summarize the results.

$$ {}^2S_M = H\sqrt{\frac{n^2-1}{12}} \tag{50} $$

$$ \Delta = \frac{H(n+1)}{3} \tag{51} $$

$$ \Delta_R = \frac{H(n^2-1)}{3n} \tag{52} $$

$$ {}^2\Delta = H\sqrt{\frac{n(n+1)}{6(n-1)}} \tag{53} $$

$$ {}^2\Delta_R = H\sqrt{\frac{n^2-1}{6}} \tag{54} $$

36 A series made of n consecutive natural numbers is a particular case of the previous series where $H = 1$.

37 The second type of series considered by Gini is the one where the intensity of the differences between each character and the median grows geometrically with the rank-distance between each character and the median. In other words:

$$\|a_i - M\| = \begin{cases} 0 & \text{if } d_{i,M} = 0 \\ H^{d_{i,M}} & \text{otherwise} \end{cases}$$

The following formulas summarize the results. For n even:

$${}^1S_M = \frac{2}{n} \frac{H^{\frac{n+1}{2}} - H^{\frac{1}{2}}}{H - 1} \tag{55}$$

$${}^2S_M = \sqrt{\frac{2}{n} \frac{H^{n+1} - H}{H^2 - 1}} \tag{56}$$

$$\Delta = \frac{4}{n(n-1)} \frac{(n-1)H^{\frac{n+3}{2}} - (n+1)H^{\frac{n+1}{2}} + H^{\frac{3}{2}} + H^{\frac{1}{2}}}{(H-1)^2} \tag{57}$$

$$\Delta_R = \frac{4}{n^2} \frac{(n-1)H^{\frac{n+3}{2}} - (n+1)H^{\frac{n+1}{2}} + H^{\frac{3}{2}} + H^{\frac{1}{2}}}{(H-1)^2} \tag{58}$$

and for n odd:

$${}^1S_M = \frac{2}{n} \frac{H^{\frac{n+1}{2}} - H}{H - 1} \tag{59}$$

$${}^2S_M = \sqrt{\frac{2}{n} \frac{H^{n+1} - H^2}{H^2 - 1}} \tag{60}$$

$$\Delta = \frac{4}{n(n-1)} \frac{(n-1)H^{\frac{n+3}{2}} - (n+1)H^{\frac{n+1}{2}} + 2H}{(H-1)^2} \tag{61}$$

$$\Delta_R = \frac{4}{n^2} \frac{(n-1)H^{\frac{n+3}{2}} - (n+1)H^{\frac{n+1}{2}} + 2H}{(H-1)^2} \tag{62}$$

38 In the previous examples, the n observations in the series took all different values. Gini next considers examples of series where observations may take the same values. Let $x_0, x_1, \ldots, x_s$ be the $s+1$ distinct values assumed by the n terms of the series. Let $x_{k+1} = x_k + 1$, and f_k be the frequency of the value x_k, for $k > s$. First Gini assumes the following functional form:

$$f_k = \frac{s!}{k!(s-k)!} \tag{63}$$

which corresponds to the frequency curve of the Newton binomium $(p+q)^s$, where $p = q$, and, moreover, it has a limit-representation (for s big) in the Gaussian curve,

or curve of random errors. The results are as follows (all results apply for each value of s, except Eq. 66 which is valid for s big enough, and Eq. 70 which is valid for $s \to \infty$):

$$^2S_{A=M} = \frac{\sqrt{s}}{2} \tag{64}$$

$$^1S_{A=M} = \sqrt{\frac{s}{2\pi}} \tag{65}$$

$$^2\Delta_R = \sqrt{\frac{s}{2}} \tag{66}$$

$$^2\Delta = \sqrt{\frac{sn}{2(n-1)}} \tag{67}$$

$$\Delta_R = \frac{(2s-1)!}{2^{2s-1}(s-1)!(s-1)!} \tag{68}$$

$$\Delta = \frac{n(2s-1)!}{(n-1)2^{2s-1}(s-1)!(s-1)!} \tag{69}$$

$$\Delta_R = \Delta = \sqrt{\frac{s}{\pi}} \tag{70}$$

From Eqs. 65, 66, 67, 68, 70, given the hypothesis that quantities are distributed according to a Gaussian curve, Gini finds the following relations between variability indices:

$$\Delta = \sqrt{2}\,^1S_A \tag{71}$$

$$\Delta = \frac{2}{\sqrt{\pi}}\,^2S_A \tag{72}$$

$$^2\Delta = \sqrt{2}\,^2S_A \tag{73}$$

$$^2\Delta = \frac{1}{\sqrt{\pi}}\,^1S_A \tag{74}$$

39 Gini considers next the example of a continuous series of x_k terms, where x_1 is the smallest value and x_s is the biggest value. The number of times the character assumes a value between x_k and $x_k + dx_k$ is described by:

$$f_k = Vx_k^{-h}dx_k \tag{75}$$

Now, let $F_k = \int_{x_k}^{x_s} Vx_\lambda^{-h}dx_\lambda$, x_s be much bigger than x_k, and $h-1$ be greater and not too close to 0, then

$$F_k = \frac{1}{h-1}Vx_k^{-h+1} \tag{76}$$

After some algebraic manipulations, and under the hypothesis that (i) x_1 is much smaller than x_s and (ii) $h-2$ is greater than and not too close to zero,

Gini determines the values for the mean difference and the mean difference with repetition:

$$\Delta_R = \frac{2}{2h-3}A \tag{77}$$

$$\Delta = \frac{2}{2h-3}\frac{n}{n-1}A \tag{78}$$

Moreover, Gini determines the value of the simple mean deviation, which, under the hypothesis that (i) M is negligible with respect to x_s and (ii) $h-1$ is greater than and not too close to zero, takes the form

$${}^1S_M = \frac{h-1}{h-2}(1-2^{\frac{1}{1-h}})M = \left(2^{\frac{1}{h-1}}-1\right)A \tag{79}$$

and, under the hypothesis that (i) A is negligible with respect to x_s and (ii) $h-2$ is greater than and not too close to zero, the simple mean deviation takes the form

$${}^1S_M = \left(2^{\frac{1}{h-1}}-1\right)A. \tag{80}$$

Under the hypothesis that (i) A is negligible with respect to x_s and (ii) $h-2$ is greater than and not too close to zero, Gini defines the simple deviation from the arithmetic mean as

$${}^1S_A = 2(h-2)^{h-2}(h-1)^{1-h}A \tag{81}$$

and the square mean deviation from the arithmetic mean and the square mean difference with repetition as

$${}^2S_A = \frac{1}{\sqrt{(h-1)(h-3)}}A \tag{82}$$

$${}^2\Delta_R = \sqrt{\frac{2}{(h-1)(h-3)}}A \tag{83}$$

40 Gini suggests to use Eqs. 78–81 for computing the variability of all fiscally assessed incomes because he noted that the various hypotheses underlying these equations were verified empirically and because the distribution of these incomes (above a given value) is well approximated by the curve described by Eq. 75 (a finding attributed by Gini to Pareto). Gini gives some examples using data on income levels in Austria (1900), Amsterdam (1906–1907) and in the Reign of Saxony (1892) to back up these arguments.

41 Next, Gini notes that all the variability indexes discussed are negatively correlated with h and that the expected difference between incomes of two random individuals varies across different times and countries from 70% to 150% of the mean

income. These are extremely high values considering that in the case of maximum inequality, when just one individual owns all country's income, the mean difference is 200% of mean income.

42 In 1895, Pareto had observed that the distribution of incomes can be approximated by the curve described in Eq. 75. He defined $\alpha = h - 1$ as *income distribution index*, assuming that inequality would change proportionally with α. Moreover, since values of α empirically determined had not changed much along time and across countries at the time Pareto performed his analysis, he concluded that the distribution of incomes is constant and independent of any social, political or economic influence. Gini agrees with the definition of α given by Pareto, but he disagrees with the second conclusion: the distribution of incomes, he affirms, is very dissimilar in time and space.

43 Gini proposes a new formula to approximate the distribution of fiscally assessed incomes, different from Eq. 76, which, he says, fits empirically better the true data

$$F_k = \frac{1}{K} T_k^{\delta} \tag{84}$$

where δ and K are constants, and $T_k = \int_{x_k}^{x_s} V x_\lambda^{1-h} dx_\lambda$. Gini defines δ as the *concentration index* of global incomes, since Eq. 84 can be rephrased as:

$$\frac{F_k}{F_1} = \left(\frac{T_k}{T_1}\right)^{\delta} \tag{85}$$

and therefore δ is the constant to which to elevate the fraction of highest incomes in order to obtain the fraction of individuals who owns these incomes.

44 In this paragraph, Gini examines the relation between formula (84) and formula (76). Under given circumstances, which are seldom satisfied, a constant relation exists between the two formulas. Gini demonstrates the desirability of using Eq. 84 over Eq. 76 in a number of circumstances, and provides practical examples using data on rents, wages, and wealth.

45 Given a series of n quantities with arithmetic mean A, inequality is maximum when one quantity in the series is equal to $T = An$, and the other $n - 1$ quantities are equal to 0. In this case:

$$^1S_A = 2\frac{n-1}{n} A \tag{86}$$

$$^1S_M = A \tag{87}$$

$$^2S_A = \sqrt{n-1}\, A \tag{88}$$

$$^2S_M = \sqrt{n} A \tag{89}$$

$$\Delta = 2A \tag{90}$$

$$^2\Delta = \sqrt{2n} A \tag{91}$$

46 Inequality is minimum if all quantities in the series are equal. Therefore: $^1S_A = {}^2S_A = {}^1S_M = {}^2S_M = \Delta = {}^2\Delta = 0$.

47 Comparing the mean difference and the simple or quadratic deviations in the different types of series considered so far, sometimes a constant relation applies.

In the case of minimum inequality (paragraph 46) it is:

$$\Delta = {}^1S_M = {}^1S_A = {}^2S_M = {}^2S_A.$$

In the case of the curve of random errors (paragraph 38) the relations are exactly:

$$\Delta = 1.414\ {}^1S_M = 1.414\ {}^1S_A$$

$$\Delta = 1.128\ {}^2S_M = 1.128\ {}^2S_A.$$

In the case of arithmetic progression (paragraph 35), and for n big enough:

$$\Delta = 1.333\ {}^1S_M = 1.333\ {}^1S_A$$

$$\Delta = 1.306\ {}^2S_M = 1.306\ {}^2S_A.$$

In the case of hyperbolic curve (paragraph 39), these relations vary as the h power of x_k, changes. For n big enough:

$$\Delta = {}^1S_M \frac{2}{(2h-3)(2^{\frac{1}{h-1}}-1)}$$

$$\Delta = {}^1S_A \frac{(h-1)^{h-1}}{(2h-3)(h-2)^{h-2}}$$

$$\Delta = {}^2S_A \frac{2\sqrt{(h-1)(h-3)}}{2h-3}$$

As h grows, the ratios $\frac{\Delta}{{}^1S_M}$, $\frac{\Delta}{{}^1S_A}$ and $\frac{\Delta}{{}^2S_A}$ become smaller.

In the case of maximum inequality (paragraph 45), the following relation is constant:

$$\Delta = 2\,{}^1S_M$$

For n big enough, the following constant relation applies between the mean difference and the simple mean deviation form the arithmetic mean:

$$\Delta = \frac{n}{n-1}\,{}^1S_A.$$

Instead, relations between the mean difference and the quadratic mean deviation from the mean and the median value grow as n becomes larger. Thus

$$\Delta = \frac{2}{\sqrt{n-1}}\,{}^2S_A$$

$$\Delta = \frac{2}{\sqrt{n}}\,{}^2S_M.$$

Moreover, in the case of deviations growing geometrically with their rank-distances (paragraph 37), the relations between variability indexes are functions of H and n.

Gini concludes that the relations between the mean difference, the simple mean deviation and the quadratic mean deviation from the median and the simple and

quadratic mean deviation from the arithmetic mean vary with the type of series, and—within the same type of series—they may be functions of constant values.

48 Gini gives some examples of discordancies of results obtained with different variability indexes using data extracted from the Parisian meat market and the cephalic indexes of Italian conscripted soldiers.

7 On the most convenient index to measure the variability of characters *(Dell'indice piú conveniente per misurare la variabilità dei caratteri)*

49 Given that a constant relation exists between different variability indices when observations are infinitely numerous, the choice of a specific index to compare variability for different series of infinite observations is irrelevant. But, Gini remarks, the relation between indices varies when the number of observations is limited, which, in practice, is the most frequent case. Therefore, Gini continues, the choice of the index becomes relevant. In particular, it depends on whether the researcher is interested in evaluating the limit-variability (of the character's infinitely extended series of observations) or the concrete variability (based on time- and space-limited observations). Physicists and astronomers are interested in the exact measure they can infer from n different observations: they are therefore interested in evaluating the limit-variability. In other disciplines the scope of the analysis on the same set of observations can be different. If the aim is to infer the limit-variability, the best index is the quadratic mean deviation, which has the higher probability to return the result obtained on an infinitely large number of observations. Instead, if the aim is to measure the concrete-variability, the aim of the research itself should lead towards the choice of the best index.

50 As pointed out already in paragraph 11 the scope of the analysis is different if (i) the series is made of true different measures or (ii) the series is made of different observations affected by measurement error of a true, unknown, measure. In the second case, the aim of the variability analysis is to determine how much the observed measures are different from the true measure. Hence, the arithmetic mean of deviations (or the simple mean deviation) is the most appropriate tool, since it represents the expected value of the difference between the true measure and one of the observed measure, randomly chosen. In the first case, instead, the aim of the analysis is to determine how much the true observed measures are different from each other. The most appropriate tool in this case is the arithmetic mean between all possible differences that can be found between the quantities (i.e. the mean difference), since it represents the expected value of the difference between two randomly chosen quantities. In both cases, the underlined assumption is that deviations and differences have an importance which is proportional to their magnitude. Even if, Gini says, this assumption is often true, it is possible that deviations and differences have an importance which is more or less than proportional to their magnitude. Therefore, Gini continues, it may be appropriate to use a weighted average of deviations or differences.

51 In practice, to compute such indices, the researcher needs to know either the true value (ii) or the value of all possible observations (i). But this information is not

always known. When studying the variability of approximate measures of the same quantity, the analysis aims at finding a way to determine the expected difference between the surveyed measures and the true unknown measure. Assuming that the expected value of the positive measurement errors is equal to the expected value of the negative measurement errors, the arithmetic mean of the surveyed measures is the expected value of the true measure. Therefore, the analysis of variability can be performed by replacing the true unknown measure with the arithmetic mean of the surveyed ones. This procedure is not free from consequences on the simple mean deviation, and Gini remarks that a correction coefficient had not been computed to translate the value of the simple mean deviations from the arithmetic mean to the expected value of the simple mean deviation from the true value. Such a correction had instead been computed for the quadratic mean deviation. Let n be the number of observed quantities, the expected value of the quadratic mean deviation of the n quantities from the true value is equal to the quadratic mean deviation from their arithmetic mean times $\frac{n}{n-1}$. Gini underlines that both indices (simple mean deviation and quadratic mean deviation from the arithmetic mean) should be used together since neither of the two can be considered better than the other and both have shortcomings. The simple mean deviation from the arithmetic mean does not consider the replacement of the true value with the arithmetic mean while the square mean deviation does not impose a weight to errors.

52 Next, Gini considers the other alternative (i), where each surveyed measure corresponds to an existing true quantity. Correspondence does not mean coincidence, since measurement errors are always possible. Often, measurement errors are negligible in comparison to differences occurring between quantities. In this case, the mean difference between surveyed quantities is the best index of the characters' variability. However, sometimes measurement errors are not negligible, and the mean difference between surveyed quantities does not coincide with the mean difference between true measures. Note that errors can be accidental (assuming that the error curve is symmetric) or systematic. In the first case, Gini demonstrates that the mean difference between surveyed quantities is larger than the mean difference between true quantities. Moreover, accidental measurement errors have a negligible influence on the simple mean deviation, but a large influence on the mean difference between quantities. Gini suggests to use both indices when a series shows significant measurement errors or when one needs to compare two series of which one has greater measurement errors than the other.

In the second case, when measurement errors are systematic, there is not a unique rule. Gini gives two examples. If errors are such that all quantities are overstated or understated by a constant K, then simple mean deviation, quadratic mean deviation and mean difference are equivalent for the surveyed and the true quantities. If errors are such that all quantities are overstated or understated proportionally to their magnitude, then simple mean deviation, quadratic mean deviation and mean difference are equivalent if applied on the logarithm of the surveyed quantities or on the logarithm of the true quantities.

53 Some series are in between series where the quantities correspond to true measures and series where quantities are approximative expressions of the same quantity. These are series where quantities not only correspond to true values,

but they can also be interpreted as aberrant expressions diverging from the typical measure. Gini gives the example of heights in a population of a specific race. Each measure corresponds to one individual true height, but this height is also the individual departure from the typical height of the race. Again, it is the scope of the analysis which leads to the choice of the most appropriate index of variability. Another example helps understanding the issue. The variability of working ability across individuals in some population can be interpreted as one of the determinants of the economic inequality between individuals. If this is the case, the researcher is interested in evaluating the individual differences in working ability and the mean difference is the most appropriate index to serve that purpose. On the other hand, the variability of working ability of different individuals is a measure of the homogeneity of types in the population. In this case, Gini suggests the use of the arithmetic mean of deviations from the typical working ability, which can be found in the most frequent, normal or modal value.

54 Sometimes it is not possible to know all possible values of a series, but quantities are grouped in categories, which can be small or large. If categories are small, (as in the case of heights which are approximated at centimeters), no particular issue arises, since each quantity can be considered as the upper limit, or the lower limit, or the partial sum of limits of the category. But quantities are often grouped in large categories which may or may not be of the same dimension. In this case, computing the mean difference, the square mean difference and the square mean deviation becomes impossible. It is often still possible to compute the simple mean deviation from the arithmetic mean and from the median, when the arithmetic mean, the median, the sum and the number of all quantities above and below the arithmetic mean and the median are known. Even if the number of quantities within each category is unknown, it is still possible to make a direct approximation of the simple mean deviation from the mean and the median or the square mean deviation from the mean, provided that the lower limit of the lower category and the upper limit of the upper category are known, and that the hypothesis of uniform distribution of quantities within categories is reasonable.

55 In this section, Gini considers also the case where data are grouped into large classes, where only the arithmetic mean of each class is known. In this case, measures of variability cannot be estimated directly but, if quantities follow approximately the distribution described in formula (76), they can be estimated as described in paragraphs 49 and 50.

8 On the comparisons between variability of characters with different average values *(Del confronto tra la variabilità dei caratteri che presentano valori medi differenti)*

56–61 Gini spends a few paragraphs trying to answer the question of whether, in addition to the absolute value of the differences, also the magnitude of the characters' values should play a role in the analysis of variability. He concludes that a unique solution does not exist; instead, different solutions may apply to different cases according to the nature and the purpose of each research. If the averages of the characters' values are numerically different because of the use of a different unit of

measurement for quantities of the same nature (e.g. income expressed in different currencies), index of absolute differences cannot be comparable, and the different unit of measure must be homogenized. The true problem arises when we aim to compare quantities of different nature, for which a unique unit of measurement cannot be found. It is not always satisfactory, in fact, to simply use the ratio between the absolute index of variability and the average value of the different characters. Therefore, neither the absolute approach nor the relative approach seem to be applicable to all situations.

9 Indices of oscillation and evolution *(Indici di oscillazione e di evoluzione)*

62 Whenever a character assumes different intensities successively rather than simoultaneously, the analysis of variability may take into account also the order of appearance. Gini makes use of an example where we observe the annual price of an item for n consecutive years. The aim of the study can be different, he says. First, it may be necessary to determine the expected error in the estimate of the price in one year starting from the information on the price in some other year. In this case, the most appropriate tool is the arithmetic mean of the $n(n-1)$ differences between all annual prices. Second, the object of analysis may be to find the expected error in determining the price in one year given the price in the previous year. The most appropriate index would be the arithmetic mean of $n-1$ differences between consecutive prices. Last, it may be interesting to know if and how much the price in some year is larger or smaller than the expected value in the subsequent year. In this case, the index to be used is the sum of all $n-1$ possible differences between consecutive terms, divided by $n-1$, i.e. the difference between the first and the last term, divided by $n-1$. The arithmetic mean of consecutive successive values measures the oscillation of these values over time. Gini defines this index *oscillation index* and he labels it O. The algebraic difference between the first and the last term divided by $n-1$ measures the increment or decrement of the observed character over the time-span considered. Gini defines this index *evolution index* and he labels it E. If all differences between consecutive values are the same, then $O = |E|$, otherwise $O > E$. Gini remarks that, in the same way as with the variability indices, also for the oscillation index it may be more appropriate to use a weighted average instead of the arithmetic mean. Moreover, the same argument about relative and absolute indices used for variability also applies to evolution and oscillation indices.

63 Normally, the oscillation index (O) of a series of n elements is different from the mean difference (Δ) computed on the same series. When the oscillation index is bigger, it means that differences between consecutive quantities are larger, on average, than the differences between all quantities, and therefore the character has a tendency to diverge. When the oscillation index is smaller than the mean difference, it means that differences between consecutive quantities are smaller, on average, than the differences between all quantities, and therefore the character has a tendency to converge. Gini defines as $R = \frac{\Delta}{O}$ the ratio which explains the relations between some character's consecutive intensities. He gives an example using data about prices of different kinds of meat in Paris for the years 1867–1910.

References

1. Dalton, H.: The measurement of the inequality of incomes. Econ. J. **30**, 348–461 (1920)
2. Gini, C.: Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche. C. Cuppini, Bologna (1912)
3. Gini, C.: Sulla misura della concentrazione e della variabilità dei caratteri. Atti R. Ist. Veneto Sci. Lett. Arti **LXXIII**(II), 1203–1248 (1914)
4. Gini, C.: Measurement of inequality of incomes. Econ. J. **31**(121), 124–126 (1921)
5. Gini, C.: On the measurement of concentration and variability of characters. METRON - Int. J. Stat. **LXIII**(1), 3–38 (2005)
6. Laplace: Théorie Analytique des Probabilités. II Supplément. Courcier, Paris (1820)