Measuring Inequality: The Origins of the Lorenz Curve and the Gini Coefficient

Michael Schneider, La Trobe University, 2004

Devised almost a century ago, both the Lorenz curve and the Gini coefficient are still widely used today as a measure of the degree of inequality exhibited by any set of figures. This paper traces the origins of these concepts, describing how Max Otto Lorenz and Corrado Gini each arrived at a new measure of inequality, and exploring their motives in devoting themselves to achieving this objective.

As Porter (1994, p.46) put it, '[s]tatistics, pre-eminent among the quantitative tools for investigating society, is powerless unless it can make new entities'. The Lorenz curve and Gini coefficient are examples of new entities, their purpose being to measure the degree of dispersion of a set of figures.

In the 1830s, almost a century before Lorenz and Gini, the Belgian astronomer Adolphe Quetelet, one of a few 'ambitious quantifiers' (Porter, 2001, p.17), moving from astronomy to society, used the 'bulls-eye of a target as a simile for his *l'homme moyen*' (Klein, 1997, p.3), the average man. The focus of his attention was the bulls-eye, representing the mean of each of the physical characteristics of a diverse set of men, the mean being the sum of the observations relating to any particular characteristic divided by the number of observations. At this stage Quetelet's 'notion of dispersion was still intimately tied to the notion of error' (Klein, 1997, p.163).

In 1844, however, Quetelet made a quantum leap, transforming 'the theory of measuring unknown physical quantities, with a definite probable error, into the theory of measuring ideal or abstract properties of a population' (Hacking, 1990, p.108). From the time of Quetelet one line of progress in statistical theory took the form of dropping the assumption that all but one of a set of figures involve error, and trying instead to 'make' new entities relating to the set of figures as a whole.

Quetelet himself 'found bell-shaped distributions in social statistics and named the

Correspondence may be addressed to Michael Schneider, Department of Economics and Finance, La Trobe University, Bundoora, 3086, Australia; email: m.schneider@latrobe.edu.au. An early version of this paper was presented at the 2000 conference of the History of Economic Thought Society of Australia, held at the University of Sydney. Helpful comments by participants in that conference, and by my colleague Suzanne Sommer, are gratefully acknowledged. Any remaining errors or omissions are my responsibility.

vision the “law of accidental causes” … [and] is credited with the first published images of normal and skewed probability distributions’ (Klein, 1997, p.164). In a diagram with observed figures on the horizontal axis and the number of occurrences of each figure on the vertical axis, he depicted normal distribution by a symmetrical bell-shaped curve, with the mean at its mid-point, and skewed distribution by an asymmetric bell-shaped curve (reproduced in Figures 7.1a and 7.1b in Klein, 1997, p.164). Some fifty years later Karl Pearson was the first to articulate another new entity important to the study of dispersion, namely the concept he named the standard deviation, now defined as the square root of (the sum of the squares of deviations from the mean divided by the number of figures), which determines whether the bell in the dispersion diagram is tall and narrow or short and wide.[1] However, neither mean nor standard deviation was to play any part in either Lorenz’s or Gini’s measure of yet another characteristic of dispersion, namely degree of inequality.

Measuring inequality is no easy matter. Suppose that a quantity x is distributed between a number n; for ease of exposition we shall assume that n is a set of individuals, though the argument applies to any set. Distribution is equal if every individual is coupled with the quantity x/n. All other distributions are unequal. If n is two, the degree of inequality is measured by the extent to which the fraction assigned to either individual, and therefore to the other, departs from one half. If n is greater than two, however, measuring inequality is not so straightforward. Compare, for example, the distribution 2, 5, 5 with the distribution 3, 3, 6; it is not immediately obvious which is the more unequal, or indeed what it would mean to say that one is more unequal than the other.

Lorenz attempted to deal with this question in an article entitled ‘Methods of Measuring the Concentration of Wealth’, published in 1905 when Lorenz (who was born in Burlington, Iowa, in 1876) was a Ph.D. candidate and Instructor in Economics at the University of Wisconsin during the Richard T. Ely era.

The measurement of inequality, or ‘concentration and diffusion’, was one of the questions taken up in Ely’s *Studies in the Evolution of Industrial Society* (1903). In the Preface to this book Ely wrote that it ‘is the author’s duty and pleasure to express his appreciation for the varied assistance given him by his colleague, Mr. Max O. Lorenz, Assistant in Economics. The untiring and very efficient efforts of Mr. Lorenz

[1] See Klein, 1997, p.177, and Stigler, 1986, p.328.

have lightened his labors and added to the value of this book' (Ely, 1903, p.x). Lorenz's hand is undoubtedly to be detected in Chapter VI of Part II of Ely's book, entitled 'The Concentration and Diffusion of Wealth'.[2] Here Ely asked '[i]s it sufficient to say that when wealth is equally distributed there is no concentration, that when it is all in the hands of one person, there is the greatest possible concentration, and that between these two extremes there is every possible gradation?' (Ely, 1903, p.256). He replied in the negative, justifying his answer with an example comparing two distributions of a total income of one hundred dollars among ten persons, the first reading 19, 17, 15, 13, 11, 9, 7, 5, 3, 1, the second 16, 16, 16, 16, 16, 4, 4, 4, 4, 4, and pointing out that while the richer half has a greater share of the total income in the second case, inequality between individuals in both the richer half and the poorer half has been eliminated. In Ely's words, 'there may be movements both of concentration and diffusion going on simultaneously and one has to be balanced against the other' (Ely, 1903, p.257), the implicit question being that of how the 'balancing' is to be done.

No attempt was made to answer this last question. The remainder of Ely's chapter is instead devoted to the following criticisms of conclusions drawn by others about the 'concentration' of income. First, the income distribution statistics for the Grand Duchy of Baden, Prussia, Saxony, England and the United States that had been used to draw conclusions about concentration were all open to one objection or another. Second, it was invalid to conclude, as some had done, that an increase over time in the number of millionaires demonstrates increasing concentration, as their incomes may have increased no faster than mean *per capita* incomes. Third, Leroy-Beaulieu was wrong to conclude that since there were only 2570 persons in Paris in 1896 with an income of over 100,000 francs a year the distribution of income was widely diffused, as such a conclusion cannot be drawn without knowledge of the proportion of total income held by these persons (and, we may add, without information about the number of persons there were in Paris in that year, knowledge which Leroy-Beaulieu however perhaps took for granted).[3]

[2] Note that in accordance with the common practice of the day, Ely (and Lorenz) meant by 'wealth' the flow of income.
[3] Ely cited *Essai sur la répartition des richesses et sur la tendance à une moindre inégalité des conditions*, 4th ed., Paris, 1896, p.564.

Lorenz's subsequent article did not concern itself with the question of reliability of statistics, nor indeed with 'the significance of a very unequal distribution of wealth' (Lorenz, 1905, p.209), but did take up the measurement problem of 'balancing' movements of concentration and diffusion. He generalized the second and third 'Ely' criticisms listed above by stating that a necessary condition for measuring changes in inequality is that we 'take account simultaneously of changes in wealth and changes in population' (Lorenz, 1905, p.213). He added that this is not however a sufficient condition, citing by way of evidence a conclusion drawn by George K. Holmes in his 1892-3 article entitled 'Measures of Distribution'.[4]

In this pioneering work Holmes proposed that inequality in income distribution be measured by the difference between the income received by the median person and the median amount of income.[5] He defined the former as the income received by the class (defined by equal income) closest to median plus that fraction which when applied to the class closest to median and added to the total number of poorer persons results in a total exactly equal to that of the total number of richer persons plus the remaining portion of the class closest to median, and defined the latter as the income of persons belonging to the class whose income is closest to median plus that fraction which when applied to the total income of this class generates an amount which when added to the total incomes of poorer classes results in a total exactly equal to that of richer classes plus the remaining amount of the total income of the 'median' class.

An example of Holmes' complex measure is the following. Suppose that one person receives \$1, two persons receive \$2, and one person receives \$3, the total incomes received by the \$1, \$2 and \$3 classes consequently being \$1, \$4 and \$3 respectively. Then according to the Holmes definition the dollar income received by the median person would be 2 plus one half (the latter being the fraction of the median class which equates the number in the 'poorer' class with that in the 'richer' class), or

---

[4] Holmes subsequently estimated the distribution of wealth in the United States in 1890. See Williamson and Lindert (1981), p.57.

[5] The term 'median', which denotes that member of a set whose value is exceeded by one half of the members of the set and exceeds the other half, was used by Cournot as early as 1843. The median value has the property that it minimises the sum of the absolute differences from the other members of the set. Keynes (1911, p.223) stated that this property 'was known to Fechner (who first introduced the median into use) …', but it was in fact first proved geometrically by Boscovich in 1760, and analytically by Laplace in 1799, while Fechner's proof dates only from 1874 (see Hald, 1998, pp.378 and 604).

$2.50, and the median dollar amount of income would be 2 plus three quarters (the 'poorer class' dollar income of 1 plus three quarters times 4 equalling the 'richer class' dollar income of 3 plus one quarter times 4), or $2.75. The difference between these two medians being 25 cents, Holmes' measure implies that an alternative distribution of income be judged more or less unequal than this one according to whether the difference between its medians is more or less than 25 cents. Take, for example, the alternative case in which $8 is distributed equally between four persons. Holmes' 'income received by the median person' in dollars would be 2 plus one half, or $2.50, and his 'median amount of income' in dollars would also be 2 plus one half, or $2.50, the difference between the two medians in this case of equal distribution being zero.

This measure was however objected to by Lorenz, on the ground that the difference between the two medians depends not only on the distribution of income but also on its total level; double the income of everyone, for example, and the difference between the two medians is doubled, in spite of the fact that the distribution is the same as before.

Having rejected Holmes' measure, Lorenz returned to the point at which Ely gave up, namely measuring concentration according to the shares received by the richer and the poorer halves of a community, though he recognised that 'such a measure does not tell the whole story' (Lorenz, 1905, p.215), because it does not cover the degree of concentration within each half. This turned out to be the starting point for the discovery of what has become known as the Lorenz curve. For Lorenz then noted that Arthur L. Bowley in his *Elements of Statistics* (1901) had suggested dispersion be measured in terms of 'quartiles', and that drawing on this Thomas Sewall Adams (Adams and Sumner, 1905, p.538) proposed specifically that 'assuming the members of a community arranged in order according to their wealth, we find the first and second quartiles, and divide their difference by their sum' [Lorenz, 1905, p.215; $(Q_2 - Q_1)/(Q_2 + Q_1)$ in Adams and Sumner], and use this as a measure of inequality.[6] Thinking in terms of changes in concentration over time, Lorenz commented that '[t]his quotient will vary from 0 to 1, and the nearer 1, the greater the concentration. This is the best numerical measure that has yet been suggested, although it may also

[6] 'Quartiles' were both invented and named by Francis Galton in 1875 (see Hald, 1998, pp.602-4).

hide some of the changes that are going on' (Lorenz, 1905, pp.215-6).

An attempt not only to reveal, so to speak, the detailed 'changes that are going on', but also to compare 'concentration' in one society with that in others, had already been made by Vilfredo Pareto, who for countries and cities where figures were available compared income levels (x) with the numbers of persons above each income level (N). Pareto (1897, p.305) recorded, for example, that in 1893-4 in Great Britain 400,648 persons had an income exceeding £150, 234,185 an income exceeding £200, and so on.[7] He reached the unexpected conclusion that for all the societies examined, the relationship between the two variables could be represented by the equation $N = A/x^{\alpha}$, where A and α are constants.

This equation, which has since become known as Pareto's Law, implies a linear logarithmic relationship between N and x, namely $\log N = \log A - \alpha \log x$. Pareto, like Lorenz and Gini, was working during a time characterised by what Klein (1997, p.17) has dubbed 'the golden stage (sic) for graphs … [namely] empirical investigation in the late nineteenth and early twentieth centuries'. It is thus not surprising that Pareto depicted this linear logarithmic relationship graphically. Pareto's graph is reproduced in Figure 1, where log N is plotted on the vertical axis (designated AC by Pareto) and log x on the horizontal axis (designated AB by Pareto), mn representing income distribution in England and pq representing income distribution in Ireland, both of these taking the form of a straight line with a slope of $-\alpha$.

Pareto's proposed measure of distribution was α; while Pareto concluded that '… a diminution in the slope α indicates a lesser inequality of incomes'[8] (Pareto, 1897, p.312), the surrounding text makes it clear that he believed inequality in income distribution will be less the steeper the line.[9] A nearly equal distribution of income, for example, would be represented in this diagram by a very steep line showing the relatively small number of persons remaining in the 'relatively rich' group as (the log

---

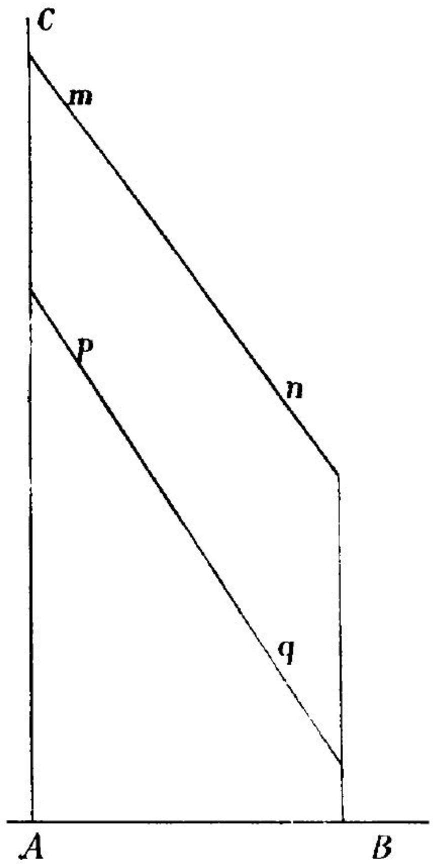
[7] Note, however, that Pareto's first exposition of his ideas on distribution is to be found in an article entitled '*La legge della domanda*', published in 1895 in *Giornale degli Economisti*.

[8] '*… une diminution de l'inclinaison α, indique une moindre inégalité des revenus*'.

[9] It is possible that Pareto may have been thinking of the vertical axis (rather than the horizontal) as the point from which to measure the coefficient alpha; this would result in a 'run' of 1 and a 'rise' of α, rather than vice-versa. In this case concentration would indeed be measured by the steepness of the curve.

of) each income level successively higher than that represented by the point A in figure 1 is considered.

Figure 1



Source: Pareto (1897), Fig. 47, p.305.

Pareto (1897, p.305; see Figure 1 again) found that for Great Britain and Ireland the two lines were almost exactly parallel, with a slope of about -1.5. He further found that the value of $-\alpha$ for all societies where data were available was close to 1.5, varying only between 1.13 and 1.89, and that in those societies where data were available for more than one year the value of α varied very little over time (Pareto, 1897, p.312).

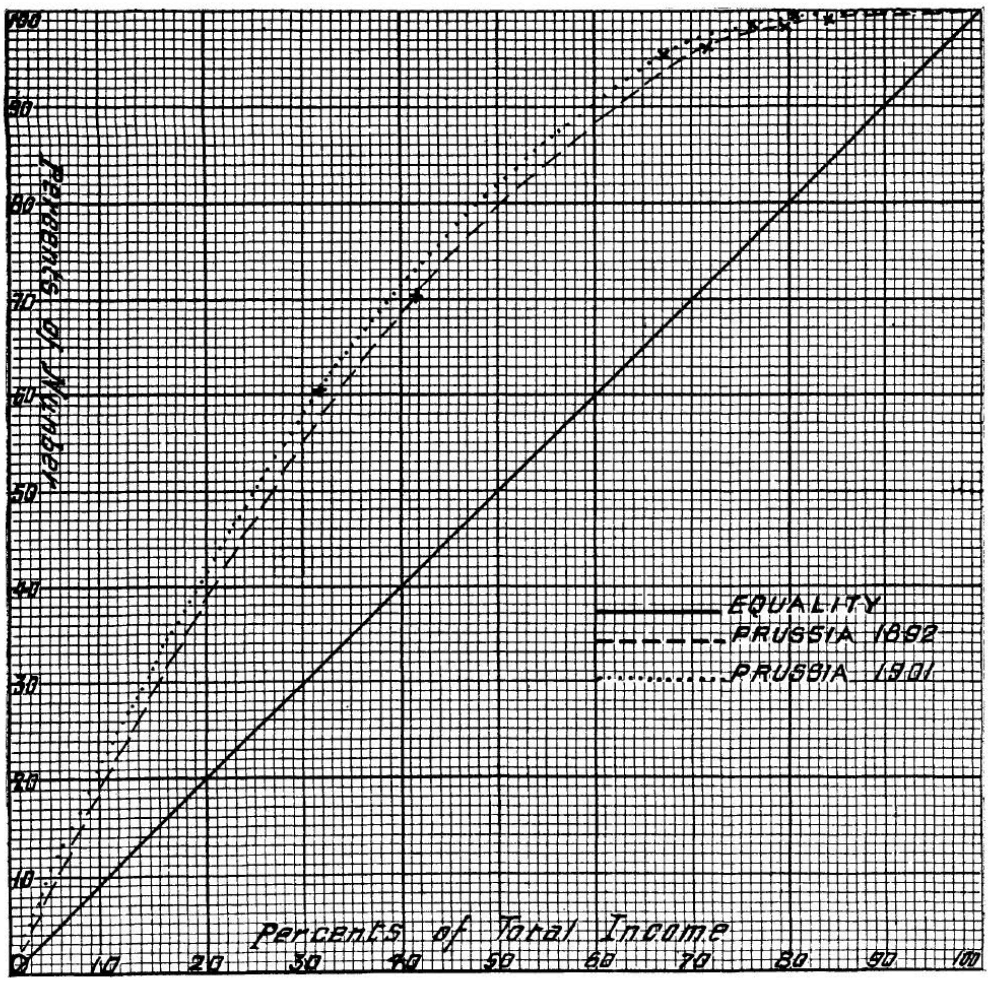
Pareto's measure of inequality was however found to be deficient by Lorenz, on the ground that when an unchanging degree of concentration is combined with growth in per capita income the value of α does not remain constant but increases (for example, for N to remain the same when x doubles, α must increase). Further, it is insensitive to changes in the incomes of those whose incomes exceed the highest income level considered; it is unaffected, for example, by any increase in the income of the person with the highest level of income.

A better procedure, Lorenz argued, would be to '[m]easure along the horizontal axis cumulated per cents. of the population from the poorest to the richest, and along the vertical axis logarithms of the *cumulated amounts* of wealth held by the successive per cents. of the population from poorest to richest' (Lorenz, 1905, p.216). Suggesting however that logarithmic curves are 'more or less treacherous', because one is apt to think of them as absolute amounts, Lorenz finally arrived at the 'Lorenz curve' by reversing the axes (without however mentioning the change, let alone explaining it, thereby imitating one of Marshall's habits) and using the vertical axis now to measure the percent of total income held by successive percents of the population, ranging from the poorest to the richest. In the case of this graph a perfectly equal distribution would generate a forty-five degree line, and in other cases, where income initially rises more slowly than population and finally rises more quickly, distribution is more equal the closer the curve is to the forty-five degree line. This enabled Lorenz to show that assuming the statistics to be correct, the distribution of income in Prussia in 1901 was more unequal than it had been in 1892 (see figure 2).

Lorenz concluded his article by recognising that that his curve would not always yield such unambiguous results. This was perhaps an after-thought, as the drawer of the diagram illustrating the ambiguous case was clearly a different person from the drawer of the previous diagram in the article (compare figure 2 with figure 3).[10] In order to demonstrate the ambiguous case Lorenz took as an example distribution of $100 among a group of ten persons in a first period as 6, 7, 8, 9, 10, 12, 12, 12, 12, 12, and in a second period as 8, 8, 8, 8, 8, 8, 8, 14, 14, 16. As Lorenz showed, this yields two curves that intersect at fifty percent of the population, illustrating over time 'a tendency toward an equal distribution in the lower half, but a contrary tendency in the upper half' (Lorenz, 1905, p.219). How to determine which of these two cases represents greater inequality of distribution was a problem that remained for Gini to solve.

[10] Both terms and handwriting in the second diagram differ from those in the first.

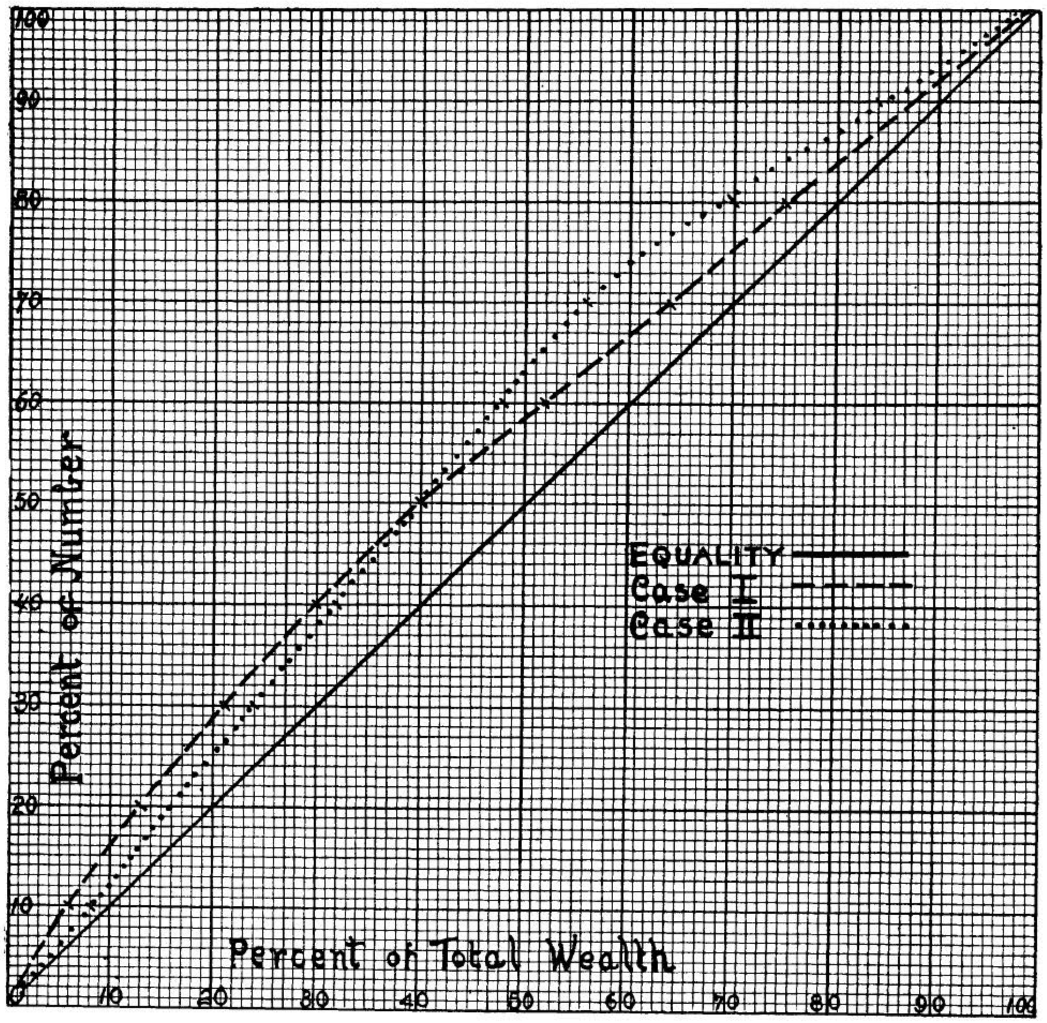
Figure 2



Source: Lorenz (1905), p.218.

Why did Lorenz devote himself to devising a new measure of inequality? Although Lorenz published his 'Lorenz curve' article during his time as a doctoral candidate, his Ph.D. topic, namely 'Outline of the Economic Theory of Railroad Rates', did not require a measure of inequality, and consequently was not the source of the Lorenz curve. Rather, Ely's handsome acknowledgment of Lorenz's assistance with *Studies in the Evolution of Industrial Society*, this book's inclusion of a chapter on income distribution with its pregnant reference to the possibility that 'there may be movements both of concentration and diffusion going on simultaneously', and Lorenz's criticisms of then existing measures of inequality all suggest that having been drawn into the field of income distribution by Ely, Lorenz set himself the task of devising a measure of inequality more satisfactory than those already existing. As noted above, Lorenz's article did not concern itself 'with the significance of a very unequal distribution of wealth'; while many of those who have used the Lorenz curve have been motivated by a concern for distributive justice, this does not appear to have played any part in its devising.[11]

Figure 3



Source: Lorenz (1905), p.219.

We now turn to Gini's attempt to devise a new measure of inequality. Born in 1884 at Motta di Livenza, near Treviso in northern Italy, Corrado Gini studied law,

[11] Note however that '[i]n politics he [Lorenz] was a Democrat' (White, 1940, p.490).

statistics, economics, mathematics and biology at the University of Bologna, and 'from this base his subsequent scientific work developed in two principal directions: the social sciences and statistics' (Castellano, 1965, p.3). In 1908 he published his doctoral thesis under the title of *Il sesso dal punto di vista statistico* (*Gender from the Statistical Point of View*).[12] A statistical study of the sex ratio at birth, in this thesis, 'beginning with an exposition of past theories, he [Gini] proceeds through existing statistical information, new hypotheses suggested by this material, and verifiable consequences to be drawn from these hypotheses to a final check of theory against the statistical data' (Castellano, 1965, p.4). The Gini coefficient likewise arose out of Gini's examination of existing statistical information.

What we now know as the Gini coefficient was the outcome of three pieces of work by Gini entitled '*Indici di concentrazione e di dependenza*' ('Indices of concentration and dependence'), published in 1910, '*Variabilità e mutabilità*' ('Dispersion and qualitative variates' - the writer's translation follows Regazzini, 1997, p.292), written after Gini had become a full professor at the University of Cagliari in 1910, and published by that institution in 1912, and '*Sulla misura della concentrazione e della variabilità dei caratteri*' ('On the measure of concentration and of the dispersion of observations'), published in 1914.

In his 1910 article Gini sought a measure of inequality. His 'coefficient of inequality' was arrived at in the following way. Supposing an increasing series a_1, a_2, … a_n, Gini first derived what he called the *intensità* (literally 'intensity') of the last m members of the series by dividing their total by their number. Dividing this in turn by the *intensità* of the series as a whole, he obtained what he called an 'index of concentration'. Alternatively, dividing the *intensità* of the last m members of the series by the *intensità* of the remaining (n-m) members, he obtained his coefficient of inequality.[13]

---

[12] Presented at the University of Bologna, Gini's thesis was awarded the Vittorio Emanuele Prize for social sciences (see Dagum, 1987a, p.529).

[13] Gini's index of concentration and coefficient of inequality have the following properties, not explicitly noted by Gini. In the case of complete equality their value is 1. In the case where all members of the series except the last are zero ('extreme inequality'), the value of the former is the value of the m^{th} member of the series divided by the sum of the series divided by n, namely n, while the value of the latter is infinity. In all other cases the value of the former will lie between n and 1, and that of the latter between infinity and 1; and the value of both will decline monotonically as m increases.

Gini illustrated his coefficient of inequality by reference to income distribution figures for the large cities of the Kingdom of Saxony in the years 1888, 1892, 1904 and 1908. He noted without comment that whether the coefficient of inequality increases or decreases between any pair of years in many cases depends on what value of m is used to measure it. This is an important deficiency of both the index of concentration and the coefficient of inequality as measures of dispersion. As an illustration, take the two series mentioned at the beginning of this paper, namely 2, 5, 5 and 3, 3, 6. For a value of m of 1, Gini's indices of concentration are 5/4 and 3/2 respectively; for a value of m of 2, they are 5/4 and 9/8 respectively. And for a value of m of 1, Gini's coefficients of inequality are 10/7 and 2 respectively; for a value of m of 2, they are 5/2 and 3/2 respectively.

Nonetheless, since Gini first suggested his index of concentration it has been widely used, particularly where imperfect data preclude the use of the Gini coefficient; for example, income inequality is commonly measured by the percentage of national income received by the top (say) ten percent of the population. Further, this measure turned out to be an important building-block in the discovery of the 'Gini coefficient'.[14] It would have anticipated the Gini coefficient more closely if Gini had divided the *intensità* of the first, rather than the last, m members of the series by that of the series as a whole, as he would then have arrived at an 'index of concentration' with values between zero and one, a characteristic shared by the Gini coefficient.

In his 1912 slim volume Gini, while acknowledging the usefulness of employing deviations from the mean in astronomy, for example, asked whether 'for the study of the dispersion of demographic, anthropological, biological and economic observations the question should not be "By how much do diverse actual magnitudes differ between each other?", rather than "By how much do diverse outlying quantities differ from their arithmetic mean?" (Gini, 1939, p.206; translation by the writer).[15]

Suggesting that in some cases the answer would be in the affirmative, Gini proceeded to derive the relevant measure as follows. For a set of n observations, designated by a_1, a_2, … a_n according to increasing size, the sum of the differences between each observation and each other will be:

---

[14] In this article Gini also demonstrated that income inequality is not an increasing function of Pareto's α, as he interpreted Pareto as believing, but a decreasing function.

$(a_n-a_1) + (a_n-a_2) + \ldots (a_n-a_{n-1}) +$

$(a_{n-1} - a_1) + \ldots (a_{n-1}-a_{n-2}) +$

$\ldots +$

(a_2-a_1)

Gini proceeded to derive what he called 'the arithmetic mean of the differences between n quantities' (indicated by the symbol Δ) by dividing this sum by the number of differences, namely ½n(n-1). Gini reduced his Δ to the following formula:

$$\Delta = [2/n(n-1)] \sum_{i=1}^{(n+1)/2} (n+1-2i)(a_{n-i+1}-a_i)$$

By definition, however:

$$\Delta = \sum_{j=1}^{n} \sum_{i=1}^{n} |a_j-a_i|/½n(n-1)$$

This formulation makes it easier to deduce two important properties of Δ, the second of which was explicitly noted by Gini in his 1912 article (see Gini, 1939, p.273); when $a_j=a_i$ for all values of i, reflecting equality between all members of the series, Δ falls to its minimum value of 0, and when all members of the series except a_n are zero, Δ rises to a maximum value of $(n-1)a_n/½n(n-1)$, that is $2a_n/n$, which is twice the mean value of the series. In the case of both of the series 2, 5, 5 and 3, 3, 6, Gini's Δ is 2.

The actual Gini coefficient did not appear until the publication of Gini's 1914 article. Drawing on the concept of an 'index of concentration' set out in his 1910 article, Gini defined q_i as the sum of the first i members of an ascending series divided by the sum of the n members of the series, that is:

$$q_i = \frac{\sum_{j=1}^{i} a_j}{\sum_{j=1}^{n} a_j}$$

[15] Bowley (1937, p.114) was later to say of the idea implied by the former that 'the conception is simple and logical'.

Subsequently defining p_i as i/n, Gini concluded that inequality will increase according to the extent that p_i exceeds q_i; to take a simple example, if i is the median member the degree of inequality is measured by the extent to which the sum of the i (lowest) members of the series falls short of one half. Building on this, Gini derived what he called the 'ratio of concentration' ('*rapporto di concentrazione*'), R, namely (note that p and q must be equal for the n^{th} member of the series, which is therefore in a sense irrelevant):

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

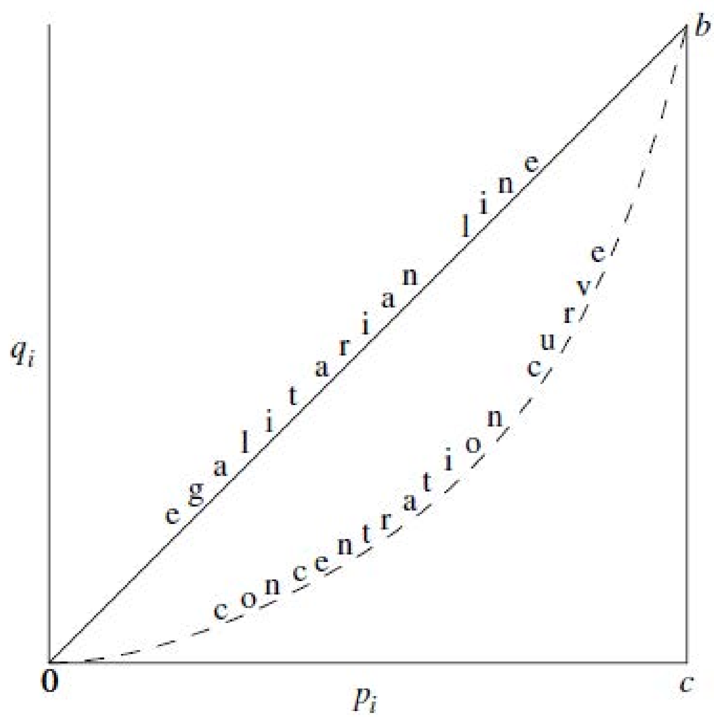
This is what we now know as the Gini coefficient, which in a rough way of speaking can be said to represent an average of Gini's indices of concentration.

As Gini demonstrated (see Gini, 1939, pp.395-7), the Gini coefficient can also be derived by dividing Gini's Δ by twice the mean of the n members of a series. As we have already seen, twice the mean of the n members of a series is the maximum value of Gini's Δ. The Gini coefficient thus has the convenient property of varying only between zero and one; in the case of perfect equality q_i equals p_i, causing the Gini coefficient to equal zero, and in the case of all members of the series bar one having a value of zero q_i equals zero, causing the Gini coefficient to be one. An intermediate case is to be found in the two series 2, 5, 5 and 3, 3, 6, where the Gini coefficient for both turns out to be 0.25; it is thus not surprising that there is no obvious answer as to which of these two series represents the greater degree of inequality.

Why did Gini devote so much of his time between 1908 and 1914 to devising a new measure of inequality? As an academic he was of course seeking to advance his reputation by publishing articles containing new ideas in the discipline he chose initially, namely the methodology of statistics. Gini's specific objective, in the words of Castellano (1965, p.13), was to produce 'a chiselling work of analysis, mainly employing specialized techniques towards limited and particular aims, yet governed and oriented by an implicit intuitive over-all vision'. His predominant aim during this period was the derivation of a measure of inequality not subject to the limitation that it was dominated by the concept of the mean. It should not be overlooked, however, that ‘[p]ractical problems were always the stimulus for Gini's methodological work’ (Castellano, 1965, p.4), the practical problem faced by Gini during this period being to devise an improved measure of inequality in the distribution of income.

The inter-relationship between the Gini coefficient and the Lorenz curve was noted by Gini in section 6 of his 1914 article, when he wrote that ‘in connection with that which we propose as an appropriate measure of concentration, we also come to the perfecting of a graphical method that some authors, namely Lorenz, Chatelain, and Séailles, have already proposed as a guide to the greater or less inequality of the distribution of wealth’ (Gini, 1939, p.386, translation by the writer); Gini’s representation of the Lorenz curve is reproduced in figure 4.[16]

Figure 4



Source: Gini (1914)
On the measurement of concentration and variability of characters
METRON, International Journal of Statistics, vol.LXIII, n.1, 2005.

[16] While Dagum (1987b, p.530) claims that some refer to ‘the Lorenz-Gini curve since it was independently introduced by both authors’, none of Gini’s published writings before his 1914 article contains anything resembling such a curve, and the cited quotation from that article seems to imply that Gini came across the curve in the course of reading works by Lorenz, Chatelain and Séailles.

Gini's 'perfection' of the Lorenz curve included a demonstration that the 'ratio of concentration' is measured by the area between the equal distribution line (*retta di equidistribuzione*) and the concentration curve (*curva di concentrazione*) divided by the area of the triangle of which it forms part[17], these two areas representing the numerator and the denominator respectively of the ratio of concentration. Gini thus found a solution to the problem of intersecting Lorenz curves left unresolved by Lorenz. For Lorenz's series 6, 7, 8, 9, 10, 12, 12, 12, 12, 12 the Gini coefficient is 2/15, while for his series 8, 8, 8, 8, 8, 8, 8, 14, 14, 16 it is 4/25. However, since the published work of Gini suggests that he may well have discovered the famous coefficient in the course of 'perfecting' the Lorenz curve, it is arguable that the Gini coefficient should be known as the Gini-Lorenz coefficient, even though Lorenz was not aware of all the properties of the new entity he had 'made'.

REFERENCES

Adams, T.S. and Sumner, H.L. (1905) *Labor Problems*. New York: Macmillan.

Anon (1969) 'Lorenz,Max Otto', entry in *The National Cyclopædia of American Biography* (White, J.T., ed.). 5:490.

Anon (1973) 'Lorenz, Max Otto', entry in *Who Was Who in America?*. Chicago: Marquis Who's Who, V:440.

Bowley, A.L. (1937) *Elements of Statistics*, 6th edn. New York: Staples Press Limited.

Dagum, C. (1987a) 'Gini, Corrado (1884-1965)', entry in *The New Palgrave: A Dictionary of Economics*, vol.2. London and Basingstoke: Macmillan, p.529.

Dagum, C. (1987b) 'Gini ratio', entry in *The New Palgrave: A Dictionary of Economics*, vol.2. London and Basingstoke: Macmillan, pp.529-32.

Ely, R.T. (1903) *Studies in the Evolution of Industrial Society*. New York: Macmillan.

Gini, C. (1908) *Il sesso dal punto di vista statistico*. Milan: Sandrom.

Gini, C. (1910) '*Indici di concentrazione e di dependenza*', in *Biblioteca dell' Economista*. Torino: Unione Tipografico – Editrice Torinese, XX (1a & 2a): 453-469. Reprinted in Gini (1939) with additional footnotes, pp.1-188.

Gini, C. (1912) '*Variabilità e mutabilità*', *Studi economico-giuridici pubblicati per cura della Facoltà di Giurisprudenza della R.Università di Cagliari* III (2a). Reprinted in Gini (1939) with additional footnotes, pp.189-358.

Gini, C. (1914) '*Sulla misura della concentrazione e della variabilità dei caratteri*', *Atti del R.Istituto Veneto di Scienze, Lettere ed Arti* LXXIII (II): 1203-48. Reprinted in Gini (1939) with additional footnotes, pp.359-408).

Gini, C. (1939) *Memorie di Metodologica Statistica*, vol.I, *Variabilità e concentrazione*. Milano: Dott.A.Giuffrè – Editore.

Hacking, I. (1990) *The Taming of Chance*. Cambridge: Cambridge University Press.

[17] An alternative way of expressing the ratio of concentration is twice the area between the equal distribution line and the concentration curve divided by the area of the completed square.

REFERENCES

Adams, T.S. and Sumner, H.L. (1905) *Labor Problems*. New York: Macmillan.

Anon (1969) 'Lorenz,Max Otto', entry in *The National Cyclopædia of American Biography* (White, J.T., ed.). 5:490.

Anon (1973) 'Lorenz, Max Otto', entry in *Who Was Who in America?*. Chicago: Marquis Who's Who, V:440.

Bowley, A.L. (1937) *Elements of Statistics*, 6th edn. New York: Staples Press Limited.

Dagum, C. (1987a) 'Gini, Corrado (1884-1965)', entry in *The New Palgrave: A Dictionary of Economics*, vol.2. London and Basingstoke: Macmillan, p.529.

Dagum, C. (1987b) 'Gini ratio', entry in *The New Palgrave: A Dictionary of Economics*, vol.2. London and Basingstoke: Macmillan, pp.529-32.

Ely, R.T. (1903) *Studies in the Evolution of Industrial Society*. New York: Macmillan.

Gini, C. (1908) *Il sesso dal punto di vista statistico*. Milan: Sandrom.

Gini, C. (1910) '*Indici di concentrazione e di dependenza*', in *Biblioteca dell' Economista*. Torino: Unione Tipografico – Editrice Torinese, XX (1a & 2a): 453-469. Reprinted in Gini (1939) with additional footnotes, pp.1-188.

Gini, C. (1912) '*Variabilità e mutabilità*', *Studi economico-giuridici pubblicati per cura della Facoltà di Giurisprudenza della R.Università di Cagliari* III (2a). Reprinted in Gini (1939) with additional footnotes, pp.189-358.

Gini, C. (1914) '*Sulla misura della concentrazione e della variabilità dei caratteri*', *Atti del R.Istituto Veneto di Scienze, Lettere ed Arti* LXXIII (II): 1203-48. Reprinted in Gini (1939) with additional footnotes, pp.359-408).

Gini, C. (1939) *Memorie di Metodologica Statistica*, vol.I, *Variabilità e concentrazione*. Milano: Dott.A.Giuffrè – Editore.

Hacking, I. (1990) *The Taming of Chance*. Cambridge: Cambridge University Press.

Hald, A. (1998) *A History of Mathematical Statistics from 1750 to 1930*. New York: John Wiley & Sons.

Holmes, G.K. (1892-3) 'Measures of Distribution', *Journal of the American Statistical Association* III: 141-57.

Keynes, J.M. (1911) *A Treatise on Probability*. London: Macmillan.

Klein, J.L. (1997) *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938*. Cambridge: Cambridge University Press.

Leroy-Beaulieu, P.P. (1896), *Essai sur la répartition des richesses et sur la tendance à une moindre inégalité des conditions,* 4th edn. Paris: Guillaumin.

Lorenz, M.O. (1905) 'Methods of measuring the concentration of wealth', *Journal of the American Statistical Association* (new series) 70: 209-17.

Pareto, V. (1895) '*La legge della domanda*', *Giornale degli Economisti* 7: 59-68.

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

Porter, T.M. (1996) 'Making Things Quantitative', in *Accounting and Science: Natural inquiry and commercial reason*. Cambridge: Cambridge University Press.

Porter, T.M. (2001) 'Economics and the History of Measurement', in *The Age of Economic Measurement* (Klein, J.L. and M.S.Morgan, eds). Durham: Duke University Press.

Regazzini, E. (1997) 'Gini, Corrado', entry in *Leading Personalities in Statistical Sciences: from the 17th century to the present* (Johnson, N.L. and Kotz, S., eds). New York: Wiley, pp.291-6.

Stigler, S.M. (1986) *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, Massachusetts: Harvard University Press.

Williamson, J.G. and P.H.Lindert (1981) ‘Long-term trends in American wealth inequality’, in J.D.Smith (ed.), *Modeling the Distribution and Intergenerational Transmission of Wealth. Chicago: University of Chicago Press.*