A NEW APPROACH TO THE MEASURE OF CONCENTRATION:  
ABC (AREA, BARYCENTRE AND CONCENTRATION)

Gustavo De Santis

Dipartimento di Statistica “G. Parenti”, Viale Morgagni 59, 50134 Firenze, 
Italy E-mail: desantis@ds.unifi.it.

Abstract

Gini’s index of concentration may be viewed from a different, and simpler angle, by 
considering where the barycentre falls in an ordered, but not cumulated distribution of the 
possessed quantitative variable (on the y axis) among owners (on the x axis - the poorest 
on the left, the richest on the right). The abscissa of the barycentre (relative to its maximum 
and minimum) provides a measure of concentration that coincides with Gini’s G. Several 
empirical applications and a few theoretical considerations show that the ABC approach 
performs at least as well as - and sometimes better than - the traditional version of Gini’s 
index.

1. INTRODUCTION

Every quantitative, non negative variable y may be distributed more or less 
equitably among N potential “owners”: for instance assets among a group of 
individuals, number of years lived by a cohort of newborn in a life table, proportion 
of ancestors of noble origin in a sample of respondents, etc. Concentration is 
frequently measured as the relative distance of the actual case from two extreme, 
thoretical cases: the most equitable and the most inequitable one. The former 
occurs when everybody possesses the same (average) quantity; the latter is less 
obvious and may vary with the case at hand, but, as a first approximation, can be 
defined as the case when one owner owns everything, and all the others own 
nothing.

The most widely used indicator of concentration is Gini’s index G (section 1). 
This papers sets out to show that the concentration problem can also be viewed from 
a different, and to the best of my knowledge new, perspective, leading to the ABC 
measures (relative Area, Barycentre and Concentration) presented here.
Since $C=G$, as the appendix proves, the proposed measures do not introduce any substantial novelty in the analysis of concentration. But they may nonetheless prove useful, in that they present the issue in an original way, which may be easier to understand, especially for non-statisticians. This holds especially in those cases when $G$ works less well, because Gini’s index:

1. relies on the assumption that the variable $y$ can be transferred indefinitely, i.e. that the theoretical case of greatest possible concentration emerges when the richest individual (or class) owns everything, and all the others own nothing. Extensions to other cases are possible, but not straightforward;
2. requires the use of a different formula with individual or grouped data (frequency distribution), and specific cautions when $N$ is relatively low;
3. is cumbersome to apply with model distributions, continuous both in $x$ (owners) and $y$ (the quantitative variable that is owned);
4. cannot be decomposed easily, distinguishing in particular the part of $G$ that depends on the difference between and within subgroups.

2. GINI’S CONCENTRATION RATIO: A SYNTHETIC REMINDER

The purpose of Gini’s index is to provide a normalised measure of concentration, telling whether a given, quantitative and transferable, variable is more or less equitably distributed among the $N$ owners, or if a few of them possess most of it.

In order to do this, Gini proposed an approach that corresponds to the following steps. Firstly, let us put these $N$ owners on the horizontal axis “arrayed in ascending order (say, by height, as in the case of soldiers on parade) with equal distance between each two observations (soldiers)”, using Yitzhaki’s (1998, p. 22) words: the $i$-th owner possesses $y_i$. [With frequency distributions, the $j$th class will include $N_j=(\sum_i n_{ij})$ owners, each possessing $y_j$ on average, and $Y_j (=y_jN_j)$ will stand for the class total.] Secondly, let us consider not the number of owners, but their proportion to the total, with $P_i (=n_i/N)$ representing the simple, and

$$P_i = \left( \sum_{j=0}^{i} n_j / N \right)$$

the cumulated, proportion.

Similarly, as for the vertical axis, let us consider not directly the quantity owned, but the ratios $h_i = y_i / Y_{\text{max}}$ and $q_j = Y_j / Y_{\text{max}} (=y_jN_j/N_{\text{max}})$ where $Y_{\text{max}}$ is the

---

1 The original contribution can be found in Gini (1914). Most of the Italian textbooks on statistics discuss Gini’s index extensively: e.g., Leti (1983), Giusti (1983), Frosini (1990) and Piccolo (1998). As for the Lorenz curve, see Lorenz (1905) and Gini (1914).
theoretical maximum that an individual could possibly own, frequently assumed to be \( Y_{\text{max}} = \sum y_i \); in this theoretical case, the \( N^{\text{th}} \) owner\(^2\) (the last and richest), owns everything and all the others own nothing. \( h_i \) is, relative to a maximum, the amount possessed by the fraction \( p_i \), \( q_j \) the relative amount possessed cumulatively by the members of the \( j^{\text{th}} \) class\(^3\) and \( Q_i \) the cumulated, relative amount possessed by the first (that is, poorest) \( P_i \) individuals. By definition, \( P_0 = Q_0 = 0 \), and \( \sum p_i = P_N = \sum q_i = Q_N = 1 \). In other words, the graphical representation of the variables \( P \) (on the horizontal axis) and \( Q \) (on the vertical axis), lies within a normalised square, with side=1 (see, e.g. figures 1 and 2).

With individual data, Gini’s index can be calculated as

\[
G = \frac{\sum_{i=1}^{N-1} p_i - Q_i}{P_i} = \frac{\sum P - \sum Q}{\sum P} \tag{1}
\]

whereas, in the case of frequency distributions, with data grouped in \( J \) classes, it is generally preferable to use a slightly different formula

\[
\tilde{G} = 1 - \sum_{j=0}^{J-1} \left( P_j - P_j \right) \left( Q_j' + Q_j \right) = \sum_{j=0}^{J-1} PQ' - P'Q \left( \tilde{G} = \frac{N-1}{N} G \right) \tag{2}
\]

where \( P' \) and \( Q' \) stand for \( P_{i+1} \) and \( Q_{i+1} \), respectively, and where \( P_0 = Q_0 = 0 \), and \( P_1 = Q_1 = 1 \). Independently of the formula used, Gini’s index can be represented graphically as a ratio between the actual concentration area (light grey area in figure 1) and the maximum concentration area, corresponding, as a first approximation, to the area of the triangle \( OAB = 0.5 \). When the number of classes (or individuals) if finite, however, the area of maximum concentration is that of a slightly smaller triangle (\( OBC \) in figure 2), so that \( \tilde{G}_{\text{max}} = 0.5 - \frac{p_N}{2} \), where \( p_N = (1 - P_{N-1}) \) is the share of the last and richest group, or individual.

Gini’s index ranges between 0 (perfectly equitable distribution) and 1 (the richest owns everything and the others own nothing).

---

\(^2\) Or class of owners, with intensity distributions.

\(^3\) With individual data, \( q_i = h_j \)
3. A DIFFERENT GRAPHICAL REPRESENTATION OF THE SAME PROBLEM

Let us now consider a different graphical representation of the same problem. Nothing changes on the horizontal axis, where we still find proportions of owners, both as $p_i$ (simple) and $P_i$ (cumulated proportions). On the vertical axis, instead, we now put the $h_i$ (height) = $y_i / Y_{\text{max}}$, that is the simple proportions to the maximum of the quantitative variable $y$ possessed by the $i^{\text{th}}$ individual, or, with frequency distributions, by each member of the $i^{\text{th}}$ class. Graphically, this produces a series of rectangles, each with base $p_i$ and height $h_i$. Notice that, because of ranking, all the relevant points on the abscissa (beginning and end of the base of each rectangle) coincide with $P_i$, the cumulated proportions of ranked owners, and the distribution appears as an area located on the bottom right corner of the normalised square, with side = 1, as in figure 3, for instance. Notice, also, that, as before, our rectangles lie within a normalised square (side = 1).

Let us now consider three summary measures of this graphical representation. The first is $A$, the relative area of our histogram (i.e. the ratio between its area and the area of the normalised square that encompasses it).

$$A = \sum A_i = \sum p_i h_i = \frac{\sum p_i y_i}{Y_{\text{max}}} = \frac{\bar{y}}{Y_{\text{max}}} \quad (0 < A < 1) \quad (3)$$

where $A_i$ is the area of each rectangle, with base $p_i$ and height $h_i = y_i / Y_{\text{max}}$. Since

Fig. 1: Gini’s index as a measure of the relative area of Lorenz’s curve.

Fig. 2: Maximum concentration area for a finite and relatively small number of classes or individuals.

Source: Author’s simulations (cf. tables 1 and 2, column 2).
the average amount possessed is $\bar{y} = \frac{\sum p_i y_i}{\sum p_i} = \sum p_i y_i$, $A$ is also the ratio of the average $\bar{y}$ to the theoretical maximum $Y_{max}$. When $Y_{max} = \sum y_i = N\bar{y}$,

$$A = \frac{\bar{y}}{Y_{max}} = \frac{1}{N} \quad (0 < A < 1) \quad \text{(3')},$$

which means that $A \to 0$ if $N \to \infty$. In the case of table 1, instead, with $N=10$ owners and $\bar{y}=4$, $Y = \sum y_i = N\bar{y} = 40$, and $A=10\%$ (see the appendix for the details of the calculations of all ABC measures).

<table>
<thead>
<tr>
<th>Number of owners variable $y_i$ (e.g. income)</th>
<th>Actual</th>
<th>Max concentration</th>
<th>Perfect equity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) $(n_i)$</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total N</strong></td>
<td><strong>40</strong></td>
<td><strong>40</strong></td>
<td><strong>40</strong></td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

The second synthetic measure of interest to us is the abscissa of the barycentre $B$ of the distribution, indicated by the vertical arrow in the figures 3 to 6. Its formula is

$$B = \frac{\sum P_i^* A_i}{\sum A_i} \quad \text{(4)}$$

where $P_i^* = \frac{P_{i+1} + P_i}{2} = \frac{P' + P_i}{2}$ is the barycentre of each vertical bar, and $A_i$ its area.

$B$ is bounded, as figures 3-6 show. Since, by construction, potential owners appear in non-decreasing order of possession, with the richest grouped on the right side of the distribution, $B$ has a minimum in $B_{min} = 0.5$, which signals perfect equity (figure 4). $B$ has a maximum, too, which is $B_{max} = 1 - A/2$, or the barycentre of the rightmost rectangle, with height 1 ($=\sum p_i$), area $A$, and therefore also base $A$ (figure
5). If \( \Sigma y \) and \( N \to \infty \), \( A \to 0 \) and \( B_{\text{max}} \to 1.4 \)

The third and, for our purposes, most important measure is the concentration index \( C \), which measures how far on the right \( B \) falls, as compared to its theoretical limits (minimum and maximum)

\[
C = \frac{B - B_{\text{min}}}{B_{\text{max}} - B_{\text{min}}} = \frac{B - 0.5}{(1 - A/2) - 0.5} = \frac{2B - 1}{1 - A} \tag{5}
\]

Fig. 3: A graphical representation of the ABC approach.

Fig. 4: Horizontal distribution (perfect equity).

Fig. 5: Maximum concentration.

Fig. 6: From B to C (C=S1/S2).

Source: Author’s simulations (cf. table 1).

\(^4\) When \( \Sigma y \), \( B_{\text{max}} = 1-1/2N \).
Eq. (5) simplifies to \( \tilde{C} = 2B - 1 \) when \( B_{\text{max}} \to 1 \) and \( A \to 0 \). Besides, when \( Y_{\text{max}} = \sum y_i = N\bar{y} \), and therefore \( A = 1/N \),

\[
C = \frac{N}{N-1} \left( 2B - 1 \right) \left( \text{i.e.,} \tilde{C} = C \frac{N-1}{N} \right) \tag{5'}
\]

Graphically, what we evaluate is the actual distance of the barycentre \( B \) from its ideal starting point \( B_{\text{min}} = 0.5 \), or the length of the arrow \( S1 \) in figure 6, and we relate this to the maximum possible distance \( B_{\text{max}} - B_{\text{min}} \), corresponding to the length of arrow \( S2 \).

How does the ABC approach compare with the more traditional index \( G \)? The appendix shows that \( \tilde{C} = \tilde{G} \) (and, therefore, \( C = G \)). In this sense, \( C \) is useful only in that it sheds new light on how to look at the issue of concentration, and how to represent it graphically. The difference is that \( G \) focuses on the relative extension of the concentration area, using the non-negative elements \( (P_i - Q_i)/P_i \), while, in the ABC approach, the problem is: “how far on the right does the barycentre fall, and how does this compare with its theoretical maximum?”

However, ABC also improves over \( G \) in some respects, considered in the next sessions.

4. INDIVIDUAL DATA VS. FREQUENCY DISTRIBUTIONS

To find the correct value of Gini’s index, one must first consider the type of data at hand (individual data or frequency distributions? With \( N \to \infty \) ?), and then pick the right formula (1 or 2 ?), making the correct assumption on the maximum-concentration case (is it when everything is in the hand of the richest individual or the richest group ?).\(^5\)

For instance, compare tables 1 and A.1 (in the appendix): they describe exactly the same situation, but the use of equation (1) leads to different results: \( G_{(\text{Table 1})} = 0.5 \) (frequency distribution), while \( G_{(\text{table A.1})} = 0.44 \), which is the correct result (but individual data are not always available!). Using eq. (2), instead of eq. (1), does not solve the problem, because one gets \( \tilde{G}_{(\text{table 1})} = \tilde{G}_{(\text{table A.1})} = 0.4 \).

Contrary to Gini, ABC indexes are insensitive to how data are presented, either

\[^5\] For grouped data, one can also use the alternative formula \( G = \Delta/\Delta_{\text{max}} \) where \( \Delta \) is the simple mean difference (without repetitions) and \( \Delta_{\text{max}} \) its maximum, corresponding to twice the average when the richest person possesses everything, and all the others own nothing.
as a succession of individual data, or as frequency distributions, and can always be computed in the standard way. In this example, for instance, in both cases (frequency distribution and individual data) the relative area is \( A = 0.1 \), the barycentre is \( B = 0.7 \), and the resulting concentration is \( C = 0.44 \).

5. NO TOTALLY TRANSFERABLE VARIABLES

Not all variables are totally transferable. Gini’s index can be extended so as to take this constraint into account (see, e.g., Frosini, 1990), but in practice this proves difficult to handle. On the contrary, ABC works almost as easily when upper or lower bounds exist. Let us consider a few cases.

5.1 PROPORTIONS

Let us start with proportions, considering, for instance, the concentration of “fuzzy” poverty. A fuzzy poverty index is normally the synthesis of several indicators of relative deprivation, concentrating on income, assets, amenities in the house, etc.\(^6\) All the elementary components are normalised, so as to range between 0 (no sign of poverty) and 1 (definitely poor). The synthetic result of these elaborations, a weighted average of the elementary indexes, is itself an index ranging between 0 and 1 for each individual in the sample, where 0 denotes individuals with absolutely no sign of poverty whatsoever, and 1 marks individuals who are totally poor in all the dimensions considered.

Data appear as in table 2, for instance, and calculations are identical to those shown above, except that the theoretical maximum for each owner is now \( Y_{\text{max}} = 1 \) (or 100% poor), and not the sum of the observed values (\( \sum y_i = 185 \), bottom line of column 3). Correspondingly, the height of each rectangle, normally the ratio \( y_i / Y_{\text{max}} \), simplifies in this case to \( y_i \), so that columns 2 and 5 coincide.

Our relative area is \( A = \sum A_i = 0.185 \), because \( A \) is also the ratio \( y_i / Y_{\text{max}} \), or simply \( y_i \) in this case.\(^7\) The abscissa of the barycentre is \( B = \sum P_i A / \sum A = 0.153 / 0.185 = 0.827 \). The minimum for this barycentre is 0.5 (as always), while its maximum is \( B_{\text{max}} = 1 - A / 2 = 0.908 \). This leads to \( C = \frac{0.827 - 0.5}{0.908 - 0.5} = 0.803 \), which signals (in this purely hypothetical case) a high concentration of poverty, or, in other words, a relatively

\(^6\) Cerioli and Zani’s (1990) seminal contribution on this issue has been repeatedly refined. See, e.g., Betti et alii (2005).

\(^7\) Our index \( A \) coincides with the so-called “diffusion” of poverty.
sharp distinction between those who are better off and those who are worse off in economic terms.

### Tab. 2: Distribution of a fuzzy poverty index, and calculations of the ABC measures.

<table>
<thead>
<tr>
<th>Fuzzy measure of poverty</th>
<th>Owners from to</th>
<th>Owners</th>
<th>Average $y_i$</th>
<th>Total $Y_i$</th>
<th>Base $p_i$</th>
<th>Height $h_i=y_i/Y$</th>
<th>Barycentre $A_i$</th>
<th>Area $A_i*p_i h_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>from 0 to 0.4</td>
<td>500</td>
<td>0.00</td>
<td>0.00</td>
<td>0.5000</td>
<td>0.0000</td>
<td>0.2500</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.4 to 0.6</td>
<td>300</td>
<td>0.20</td>
<td>60.0</td>
<td>0.3000</td>
<td>0.2000</td>
<td>0.6500</td>
<td>0.0600</td>
<td>0.0390</td>
</tr>
<tr>
<td>0.6 to 0.7</td>
<td>40</td>
<td>0.65</td>
<td>26.0</td>
<td>0.0400</td>
<td>0.5000</td>
<td>0.8500</td>
<td>0.0425</td>
<td></td>
</tr>
<tr>
<td>0.7 to 0.8</td>
<td>30</td>
<td>0.75</td>
<td>22.5</td>
<td>0.0300</td>
<td>0.7500</td>
<td>0.9550</td>
<td>0.0225</td>
<td>0.0215</td>
</tr>
<tr>
<td>0.8 to 0.9</td>
<td>20</td>
<td>0.85</td>
<td>17.0</td>
<td>0.0200</td>
<td>0.8500</td>
<td>0.9800</td>
<td>0.0170</td>
<td>0.0167</td>
</tr>
<tr>
<td>0.9 to 1</td>
<td>10</td>
<td>0.95</td>
<td>9.5</td>
<td>0.0100</td>
<td>0.9500</td>
<td>0.9950</td>
<td>0.0095</td>
<td>0.0095</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>185</strong></td>
<td><strong>0.185</strong></td>
<td><strong>0.0153</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** $Y$ (highest possible fuzzy poverty measure) = 1. Source: Author’s simulations.

Note that it is possible to view the data of table 2 from the opposite perspective, i.e. to focus on non-poverty, and consider that 500 individuals have value 1 (i.e. are totally non-poor), 300 have value 0.8 (largely non poor), etc.

The results that one obtains in this case (calculations not shown here) are perfectly consistent with those found before. The relative area is $\hat{A} = .815$ (and, therefore, $A + \hat{A} = 1$); the barycentre is $\hat{B} = .468/.815 = .574$; and, finally, since the maximum for this barycentre is $(1 - \hat{A}/2) = .592$, the concentration index is $\hat{C} = .803 = C$.

This property holds more in general with the study of proportions: ABC measures do not depend on whether one calculates them on those who do or do not have a given characteristic.

### 5.2 LIFE TABLES

Intermediate cases exist between variables that can be considered entirely transferable (like income, for instance), and variables for which a given, well identified individual maximum exists (e.g. proportions, limited to unity). Life

---

8 This satisfies the condition that $BA + (1 - \hat{B}) \hat{A} = 0.5$, or, in other words, that the weighted mean of the barycentres of the two complementary distributions results in the overall barycentre of the normalised square, 0.5. The reason why one needs to include $(1 - \hat{B})$ instead of $\hat{B}$ in this formula is that, when both distributions are considered in the same figure (e.g. in figure 7), the two partial barycentres fall, by definition, one to the left and one to the right of 0.5.
tables are an example: they furnish, among other things, the number of years lived by each individual in a cohort of newborn, under a certain set of assumptions that need not concern us here.

The first and most interesting question to which a life table answers is: how long do these individuals live, on average? For instance, a hypothetical, typical Italian woman, subject all along her life to the mortality conditions prevailing in year 2004, would have lived about $e_0 = 83.7$ years, where $e_0$ stands for life expectancy at birth, or average duration of life (Istat, 2007).

What about the degree of concentration of this variable (number of years lived), and what about the hypothesized tendency toward “rectangularisation” in the shape of the survival curve, i.e. the idea that, as $e_0$ increases, concentration decreases and people tend to die more and more at the same (high) age? Gini’s index does not seem (to me) immediately applicable here because “it is equal to zero if all individuals die at the same age, and equal to 1 if all people die at age 0 and one individual dies at an infinitely old age.” (Shkolnikov, Andreev and Begun 2003, p. 311; emphasis added). For instance, the maximum degree of concentration in a life table like that of the Italian women in 2004 would emerge if 99.999 individuals died immediately after their birth, while one individual (Methuselah?) lived up to the age of more than 8.37 million years, thus preserving the total number of years lived by our hypothetical cohort of 100,000 individuals. This implies that Gini’s index will always tend to 0 when applied to actual life tables, as shown, for instance, in Shkolnikov, Andreev and Begun (2003).

A possible alternative is to assume the existence of a maximum life span $Y_{max}$. Admittedly, this is arbitrary, but not more than the traditional practice of having “one individual dying at an infinitely old age”. To this end, two approaches, among others, can be considered. One is to assume that $Y_{max}$ is a constant, corresponding to the maximum ever reliably recorded for a human being, i.e. 122 years, the age at death of Jeanne Calment, a French woman who died in 1997. The other is to assume that $Y$ is the age by which the proportion still surviving is reasonably small, like 2.5% or 1%, for instance. The assumption here is that, given external conditions (medicine, economy, pollution, etc.), the human beings described in the life tables were not expected to exceed the age of $Y_{max}$, and it is only by chance that a small fraction of them actually survived, for a short time, past that threshold.

Both logics can be applied to an illustrative set of life tables and the results in terms of concentration are shown in table 3, together with the traditional version of Gini. The elaborations suggest that the $C$ measure of concentration of the $ABC$ approach does depend on the assumed highest age span, and is, in this sense, partly arbitrary. However, independently of the assumptions on the maximum possible
age span, concentration consistently diminishes as $e_0$ increases, implying less inequality in the distribution of the number of years lived by each individual. In other words, when the average length of life increases, people tend to die more and more at similar ages, and rectangularisation in the survival curve occurs.

Tab. 3: Measures of concentration ($G$ and $C$) of the number of years lived in a few model life tables (Italy, women, several historical periods).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$G(Y_{max}=100.000e_0)$</td>
<td>37.6%</td>
<td>33.7%</td>
<td>27.8%</td>
<td>23.9%</td>
<td>11.6%</td>
<td>7.9%</td>
</tr>
<tr>
<td>$C'(Y_{max}=122)$</td>
<td>70.8%</td>
<td>62.6%</td>
<td>52.0%</td>
<td>45.6%</td>
<td>29.3%</td>
<td>25.5%</td>
</tr>
<tr>
<td>$C''(Y_{max}$ separates 2.5%$)$</td>
<td>84.2%</td>
<td>81.2%</td>
<td>74.5%</td>
<td>69.0%</td>
<td>53.3%</td>
<td>47.2%</td>
</tr>
<tr>
<td>$(values\ of\ Y_{max}\ in\ this\ case)$</td>
<td>(87.7)</td>
<td>(87.6)</td>
<td>(90.6)</td>
<td>(92.3)</td>
<td>(95.3)</td>
<td>(99.5)</td>
</tr>
</tbody>
</table>

Note. $Y_{max}$ is the highest theoretically possible age at death, under various assumptions. With Gini, this is the sum of the years lived in the community, or $N \cdot e_0$, where $e_0$ is the average length of life, and $N$ the number of individuals (here, $10^6$). With $C'$, $Y_{max}$ is fixed at 122, the highest age at death ever reliably recorded. With $C''$, the varying values of $Y_{max}$ (also reported in italics in the table) are calculated in such a way that only 2.5% of deaths occur after that age. Source: Own elaborations on the Human Mortality Database (http://www.mortality.org/, as of August 2007).

There is widespread agreement about this conclusion, and only approaches to measurement differ: some researchers use indicators like entropy (Keyfitz, 1977, pp. 64-66), or Anson’s “MLS & U” (Anson, 1992), but the majority resort to Gini (Wilmoth and Horiuchi 1999), etc. However, a closer look at table 3 suggests that Gini, which is relatively close to 0 even in situations of high mortality, may bias the picture, not only at any given point in time, but also in dynamic terms. Consider, for instance the evolution from $e_0=35.1$ (in 1880) to $e_0=81.8$ (in 1999): $G$ decreases from 37.6% to 7.9%, or 80% less.

With the alternative (bounded) indexes $C'$ and $C''$, relative declines appear more moderate (between -64% and -44%), and more in line with the general impression conveyed by a look at survival curves (not reported here): there is a substantial, but not such a dramatic, increase in equality in the age at death when life expectancy passes from very low to very high values. In short, the $C$ indexes discussed here (and especially $C''$, in my opinion) may constitute a valid alternative to the synthetic measures of inequality in survival chances that are customarily used.

---

9 When mortality is high, relatively more deaths occur in the first year of life, while the highest age at death is scarcely affected. This accentuates inequality in terms of individual length of life.
5.3 OTHER BIOMETRIC INDICATORS

Gini (1914) himself applied his own index to the data of table 4, to find that $G=5.9\%$.

Tab. 4: Heartbeats of 263 Native Americans around 1908.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>1</td>
<td>55</td>
<td>4</td>
<td>64</td>
<td>19</td>
<td>73</td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>56</td>
<td>19</td>
<td>65</td>
<td>3</td>
<td>74</td>
<td>3</td>
</tr>
<tr>
<td>48</td>
<td>3</td>
<td>57</td>
<td>7</td>
<td>66</td>
<td>32</td>
<td>75</td>
<td>1</td>
</tr>
<tr>
<td>49</td>
<td>1</td>
<td>58</td>
<td>24</td>
<td>67</td>
<td>5</td>
<td>76</td>
<td>2</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>59</td>
<td>7</td>
<td>68</td>
<td>18</td>
<td>78</td>
<td>3</td>
</tr>
<tr>
<td>51</td>
<td>3</td>
<td>60</td>
<td>23</td>
<td>69</td>
<td>1</td>
<td>80</td>
<td>1</td>
</tr>
<tr>
<td>52</td>
<td>5</td>
<td>61</td>
<td>2</td>
<td>70</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>2</td>
<td>62</td>
<td>19</td>
<td>71</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>12</td>
<td>63</td>
<td>11</td>
<td>72</td>
<td>12</td>
<td>Tot. 263</td>
<td></td>
</tr>
</tbody>
</table>

Source: Gini 1914.

However, the case of highest concentration that Gini is implicitly assuming is as follows:

a) 262 individuals have 0 heartbeats, and one (“the richest”) has 16343 (so that $B_{\text{max}}=0.9981$, as in figure 7a).

This assumptions seems to me to be scarcely tenable. More plausible alternatives, all preserving the total number of heartbeats to 16434, could be:

b) no one can have less than 40 heartbeats per minute. This means that, in the most extreme case, 262 individuals have 40 and 1 has 5863 heartbeats. Therefore $B_{\text{max}}=0.6775$ (figure 7b);

c) no one can have more than 85 heartbeats per minute. This means that, in the most extreme case, 70 individuals have 0 heartbeats; 192 have 85, and 1 has the remaining 23. In this case $B_{\text{max}}=0.6345$ (figure 7c);

d) both the above constraints apply. This means that, in the most extreme case, 133 individuals have 40 heartbeats; 129 have 85, and 1 has the remaining 58. Here, $B_{\text{max}}=0.5905$ (figure 7d).

Depending on the maximising assumption, the corresponding concentration measures $C$ range from 5.9% in case (a) (as Gini himself suggests) to 32.7% in case (d) - which I find more plausible. It is true that, as was the case for the study of life tables of section 4.2, this is based on some arbitrary choice on the acceptable minima and maxima for heartbeats, but these choices are implicit also in Gini (they are 0 and 16434, respectively), and, besides, it is relatively easy to verify (although
not shown here) that results do not vary too much if one tries alternative values for the lowest and highest acceptable number of heartbeats, as long as these thresholds vary within reasonable limits.

6. CONTINUOUS DISTRIBUTIONS

The ABC approach works equally well, and proves more tractable than Gini’s $G$, when the distribution of the quantity $y$ owned by each (properly ranked) owner $x$ is defined as a function $y=f(x)$ [$y \geq 0$], where $x$ ranges between 0 and 1, and both
y and x are continuous. In this case, the three basic measures of the ABC approach become

\[ A = \frac{\int_0^1 y \, dx}{Y_{\text{max}}}, \text{ with } x \in [0;1] \]  

(6)

\[ B = \frac{\int_0^1 xy \, dx}{\int_0^1 y \, dx}, \text{ again, with } x \in [0;1]. \]  

(7)

\[ C = \frac{B - 0.5}{B_{\text{max}} - 0.5} = \frac{B - 0.5}{(1 - A/2) - 0.5} = \frac{2B - 1}{1 - A} \]  

(8)

If the integers in (6) and (7) can be computed, the concentration index follows immediately. Figure 8, for instance, presents a few simple functions, where all the synthetic measures, for the sake of simplicity, have been calculated only with reference to the case of proportions (that is, with $Y_{\text{max}} = 1$).

![Fig. 8: Selected functions of y (owned variable) on x (owners, continuous variable).](image-url)
For example:

panel a) \( y = x/2 \). In this case, \( A = 1/4, B = 2/3, \) and \( C = 4/9 \);

panel b) \( y = \hat{x} \). In this case, \( A = 2/3; B = 3/5, \) and \( C = 3/5 \).

panel c) \( y = x^2 \). In this case, \( A = 1/3; B = 3/4, \) and \( C = 3/4; \) etc.

It is worth noticing that the two distributions in panels c) and d) produce the same \( ABC \) measures, despite their marked differences. This may serve as a reminder that distributions with the same \( ABC \) indexes (Area, Barycentre and Concentration) do not necessarily coincide, and may even differ sharply in some respects.

7. DECOMPOSITION OF \( ABC \) MEASURES

Finally, decomposition is difficult with Gini (cf. e.g., Bhattacharya and Mahalanobis, 1967; Dagum, 1997; Costa, 2004), but it proves easier to follow with \( ABC \), because the process can be reduced to a succession of movements of the barycentre as conditions change, one by one.

Figure 9 provides an example, with a quantitative variable \( y \) (e.g. income) that is distributed among possessors \( x \), classified in two subgroups. What we typically observe is the general distribution of figure 9 (panel 5), and we may want to try to understand what caused the barycentre to move from its ideal starting point of \( B_{\text{min}} = 0.5 \) (thick arrow of panel 1) to its actual position (\( B = .694 \) in this example; thick arrow of panel 5). In order to do this, we first need to calculate income averages: one for the whole sample (5.2, in this example) and one for each subgroup (which, in our case, turn out to be 4 and 6, respectively, for the “white” and the “grey” group). Then the argument goes as follows:

1. in the ideal case of perfect equity, income distribution would be as shown in panel 1, with everybody earning the average income of 5.2. The barycentre would be at \( B_0 = 0.5 \), by definition;

2. the second step is to imagine perfect equity within groups, ordered from the poorest to the richest (left to right). This would cause the barycentre to shift, from \( B_0 = 0.5 \) to \( B_1 \) (=.546, in this case - panel 2);

3. next, we consider each group, in turn, and we let the distribution within that group become what it actually is, noticing that the result does not depend on the order with which we proceed. The barycentre moves from \( B_1 \) to \( B_{2w} \) (.573, in this case - panel 3), to \( B_{2G} \) (.640 - panel 4), and so on, for as many groups as necessary;

---

10 *By definition, \( B = C \) when the distribution is of the type \( y = x^2 \): see appendix.*

11 *Subgroups can be more than two, as the appendix shows.*
Finally, we consider the further rightward movement of $B$ caused by the unconstrained ranking of individuals, which depends on the fact that subgroup distributions overlap, at least partially. We thus get to point $B_3$ (.694, in this case - panel 5).

**Fig. 9: Decomposition of $B$ into successive movements of the barycentre.**

Note: This is an illustrative, hypothetical case. The values on the horizontal axis are the barycentres of each rectangle. The vertical axis represents income, in this case in absolute values (not as a ratio to the maximum).
In summary, the general movement $\Delta B = B_3 - B_0$ (.194, in this case) can be broken down into specific, additive components:

1. between groups $= B_1 - B_0$ (0.046, in this case, or 24% of the total)
2. within groups $= B_2 - B_1$ (0.094, in this case, or 49% of the total), which can be further detailed into:
   - within group White $= B_{2W} - B_1$ (0.027, in this case - 14% of the total)
   - within group Grey $= B_{2G} - B_{2W}$ (0.067, in this case - 35% of the total)
3. overlapping $= B_3 - B_{2G}$ (0.054, in this case - 28% of the total)

The appendix proves that, for each component, the relative contribution to the movement of $B$ corresponds exactly to its relative contribution to the overall value of $C$, which can thus be additively decomposed into elementary parts.

Readers will note that not even the ABC approach succeeds in eliminating the overlapping element of this passage, which emerges also with other forms of decomposition, and which proves difficult to interpret. But this element is intrinsic in this type of measures: my contribution to concentration depends not only on how I fare within my group (within), and on how my group’s average fares with respect to the average of other groups (between), but also on how I fare with respect to each individual in each of the other groups.

**CONCLUSIONS**

In 1998, Yitzhaki found that there were “More than a dozen alternative ways of spelling Gini”. This paper shows that there is yet another way of spelling it, different from all the others, and, I contend, considerably simpler than them - possibly even simpler than the original, since it can be represented as a ratio of two segments, instead of a ratio of two areas. Nothing of what had been previously discovered about Gini’s index is disproved here: it is merely arrived at in a way which I find intuitively more immediate and mathematically simpler.


APPENDIX

A.1. DETAILS ON THE CALCULATION OF ABC

Table A.1 gives the details for calculating the ABC indexes on the data of table 1.

<table>
<thead>
<tr>
<th># of owners</th>
<th>Average income</th>
<th>Total income</th>
<th>Base</th>
<th>Height</th>
<th>Barycentre</th>
<th>Area</th>
<th>Weighted average</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_i</td>
<td>y_i</td>
<td>Y_i = n_i y_i</td>
<td>p_i</td>
<td>h_i = y_i / Y</td>
<td>P_i *</td>
<td>A_i = p_i h_i</td>
<td>A_i P_i *</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.100</td>
<td>0.000</td>
<td>0.050</td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.100</td>
<td>0.000</td>
<td>0.150</td>
<td>0.000</td>
<td>0.0000</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.100</td>
<td>0.050</td>
<td>0.250</td>
<td>0.005</td>
<td>0.00013</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.100</td>
<td>0.050</td>
<td>0.350</td>
<td>0.005</td>
<td>0.0018</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>0.100</td>
<td>0.100</td>
<td>0.450</td>
<td>0.010</td>
<td>0.0045</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>0.100</td>
<td>0.100</td>
<td>0.550</td>
<td>0.010</td>
<td>0.0055</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>6</td>
<td>0.100</td>
<td>0.150</td>
<td>0.650</td>
<td>0.015</td>
<td>0.0098</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>6</td>
<td>0.100</td>
<td>0.150</td>
<td>0.750</td>
<td>0.015</td>
<td>0.0113</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>8</td>
<td>0.100</td>
<td>0.200</td>
<td>0.850</td>
<td>0.020</td>
<td>0.0170</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>8</td>
<td>0.100</td>
<td>0.200</td>
<td>0.950</td>
<td>0.020</td>
<td>0.0190</td>
</tr>
</tbody>
</table>

Column (3) gives the total income within that group (not very useful, here, since we are considering individual data). Each rectangle in figure 3 has the same base \((p = 0.1)\), and a varying height \(h_i = y_i / Y_{max}\) where \(y_i\) is the average amount possessed by each individual in that class, and \(Y_{max}\) is the highest possible amount possessed by the richest individual (here \(Y_{max} = 40\), the total income in the sample). This gives the area \(A_i(p_i h_i)\) of each rectangle (column 7), each with its own barycentre \(P_i^* = \frac{P_i^* + P_i}{2}\) (column 6). The sum of all of the areas gives the total (relative) area \(A = \sum A_i\), which is 0.100 in this example. The weighted average of the 10 barycentres \(P_i^*\) gives the general barycentre \(B = \frac{\sum P_i^* A_i}{\sum A_i}\), equalling 0.7 in this case. Finally, the value of \(B\) must be interpreted in relative terms, as the relative distance from its minimum (0.5) and maximum (1-\(A/2\)). Since, in this case, the maximum is \(B_{max} = 0.95\), \(C = 0.444\).
A new approach to the measure of concentration: ABC…

219

A.2 PROOF THAT C COINCIDES WITH G (GINI’S INDEX)

Eq. (4) states that

\[ B = \frac{\sum P_i^* A_i}{\sum A_i} = \frac{\sum P_i^* p_i h_i}{\sum p_i h_i} = \frac{\sum P_i^* p_i y_i}{\sum p_i y_i} \] (A2.1)

where \( A_i \) represents the area of each rectangle, \( P_i^* \) its barycentre, \( p_i \) its base, and \( h_i \) and \( y_i \) its height (\( h_i = y_i / Y_{\text{max}} \) is in relative terms), or average amount possessed by each member of the \( i \)-th class. Normalizing \( B \) with respect to its minimum and maximum gives the concentration index \( C \)

\[ C = \frac{B - 0.5}{B_{\text{max}} - 0.5} \] (A2.2)

If we assume, for the sake of simplicity, that \( B_{\text{max}} = 1 \), (A2.2) transforms into

\[ \tilde{C} = 2B - 1 = 2 \frac{\sum P_i^* p_i y_i}{\sum p_i y_i} - 1 \] (A2.3)

Expressing all the variables in (A2.3) in terms of cumulative proportions \( P \) (for owners) and \( Q \) (for the owned variable), we get

\[ P_i^* = \frac{P_i^* + P_{i+1}}{2}; \quad p_i = P_i^* - P_{i+1}; \quad y_i = \frac{(Q_i' - Q_{i+1}) Y_{\text{max}}}{(P_i' - P_i) N} \] (A2.4)

where \( P'_i = P_{i+1}, \ Q'_i = Q_{i+1}, \ N \) represents the total number of observations, and \( Y_{\text{max}} = \sum y_i' \).

Using (A2.4), and keeping in mind that \( \sum (Q' - Q) = 1 \), equation (A2.3) may be re-written as

\[ \tilde{C} = 2 \left\{ \frac{1}{2} \frac{\sum (P' + P)(P' - P)(Q' - Q) Y_{\text{max}}}{(P' - P) N} - 1 \right\} = \frac{\sum (P' + P)(Q' - Q) Y_{\text{max}}}{\sum (P' - P)(Q' - Q) Y_{\text{max}}} - 1 = \frac{\sum (P' + P)(Q' - Q) Y_{\text{max}}}{\sum (P' - P)(Q' - Q) Y_{\text{max}}} - 1 = \frac{\sum (P' + P)(Q' - Q)}{\sum (P' - P)(Q' - Q) Y_{\text{max}}} - 1 \] (A2.5)

Besides

\[ ^{12} \text{I dropped subscripts for simplicity of notation, but remember that } N \text{ and } Y_{\text{max}} \text{ do not have subscripts.} \]
\[ \sum (P' + P)(Q' - Q) = (\sum P'Q' - \sum PQ) + \sum PQ' - \sum P'Q = 1 + \sum PQ' - \sum P'Q \]  
\hspace{1cm} \text{(A2.6)}

because the only differences between the series \((\sum P'Q')\) and \((\sum PQ)\) are the terms \(P_NQ_N (=1)\) in the first and \(P_0Q_0 (=0)\) in the second. In short, what we get is

\[ \tilde{C} = \sum PQ' - \sum P'Q = \tilde{G} \]  
\hspace{1cm} \text{(A2.7)}

and, therefore, also \(C = G\) (cf. eqs. 2 and 5').

### A.3 SUBPOPULATIONS, AND THE DECOMPOSITION OF ABC MEASURES

Sometimes, the population can be partitioned into \(Z\) subgroups (1, 2, ..., \(z\), ..., \(Z\)), and one may be interested in measuring the contribution of each subgroup to the overall value of \(B\) and \(C\). In this case, the first thing to do is to identify the individual theoretical maximum \(Y_{max}\) with respect to the entire distribution. Then calculations of the \(A\) and indexes \(B\) can proceed as usual, that is

\[ A = \sum_z \sum_i A_{zi} = \sum_z \sum_i p_{zi}h_{zi} = \frac{\sum_z \sum_i p_{zi}y_{zi}}{Y} \]  
\hspace{1cm} \text{(A3.1)}

\[ B = \frac{\sum_z \sum_i p_{zi}^*A_{zi}}{\sum_z \sum_i A_{zi}} \]  
\hspace{1cm} \text{(A3.2)}

The only difference is that the procedure can now be applied repeatedly, changing the values of \(y_{zi}\) at each round, preferably in the following order:

1. \(y_{zi} = \bar{y}\), which leads to \(B=0.5\), and serves as a starting point;
2. \(y_{zi} = \bar{y}_z\), and this, by difference with passage (1), measures the effect on \(B\) and \(C\) of an ideal situation in which there is no intra-group variability: all the individuals of group \(z\) possess the same \((z-)\)average amount of \(y\). Notice that, in doing this, we must arrange groups in an ascending order of \(\bar{y}_z\), with the poorer on the left;
3. \(y_{zi}\) takes its actual value, but individuals must be ordered only within subgroups, while the succession of groups is the same as in passage (2), from the poorest to the richest\(^{13}\). This passage serves to measure the within-component of \(G\), because it relaxes the assumption of perfect intra-group equity. It may be done
for all subgroups at once, or one by one (as in section 6, for instance), so as to measure the “within” component of each group. In all cases, results do not depend on the order with which one proceeds;

4. finally, we let each individual have his/her own amount of \( y \) and we rank individuals in the correct order, from the poorest to the richest. The difference with the final stage of passage (3) gives the effect on \( B \) and \( C \) of the overlapping of the various distributions.

This procedure may be made slightly speedier by simply calculating the value of \( B \) (instead of \( B \) and \( C \)) at each passage. Since \( C=(2B-1)/(1-A) \) and \( A \) is a constant for each distribution (by definition, since we use either actual or average values of \( y \)), \( C \) becomes a linear transformation of \( B \), so that whatever causes \( B \) to vary by \( x\% \) of its total variation will also cause \( C \) to vary by the same proportion \( x\% \).

**A.4 PROOF THAT B=C, IN MODEL DISTRIBUTIONS, WHEN Y=X^Z**

Remembering that \( C=(2B-1)/(1-A) \), the quantity \( B-C \) becomes

\[
B-C = \frac{1 - B - AB}{1 - A}
\]  

(A4.1)

which equals 0 for \( B+AB=1 \). Now, if \( y=x^z \), then \( A = \int y = \frac{1}{z+1} \) and \( B = \int xy = \frac{z+1}{z+2} \), because we calculate all integrals between 0 and 1. Therefore, we get

\[
B + AB = \frac{z+1}{z+2} + \frac{1}{z+1} \frac{z+1}{z+2} = 1
\]  

(A4.2)

which proves that \( B=C \) when \( y=x^z \).

**ACKNOWLEDGEMENTS**

Paper first presented at the *International Conference in Memory of Two Eminent Social Scientists: C. Gini and M. O. Lorenz*, Certosa di Pontignano (Siena), 23-26 May 2005. I thank professors Carla Rampichini and Benito Vittorio Frosini, for their encouragement and their thoughtful remarks on an earlier draft.

---

13 In some cases, this will lead to individuals of different groups not being properly ordered: i.e. a relatively rich individual belonging to a relatively poor group will appear to the left of some relatively poor individuals who belong to a relatively rich group, as in panel 4 of Figure 7.
BIBLIOGRAPHY


UN NUOVO APPROCCIO ALLA MISURA DELLA CONCENTRAZIONE:  
**ABC (AREA, BARICENTRO E CONCENTRAZIONE)**

**Riassunto**

L’indice di concentrazione di Gini può essere considerato sotto un profilo diverso, e più semplice, guardando al baricentro di una distribuzione ordinata, ma non cumulata, della variabile quantitativa posseduta (sull’asse y) tra i possessori (sull’asse x - i più poveri a sinistra, i più ricchi a destra). L’ascissa del baricentro (relativizzata ai valori teoricamente estremi, massimo e minimo) fornisce una misura della concentrazione che coincide con la G di Gini. Diverse applicazioni empiriche e alcune considerazioni teoriche mostrano che l’approccio ABC funziona: almeno come l’indice di Gini e, in alcuni casi, meglio.