Statistica & Applicazioni
Vol. III, n. 2, 2005
pp. 5-24

On the Decomposition of the Gini's Mean Difference and Concentration Ratio*

Paolo Radaelli

Michele Zenga[§]

Summary: *In this paper we propose a decomposition for the Gini's mean difference of a real variate obtained as the sum of c variates observed on a finite population. The decomposition is mainly based on two different sortings of the values of each variate: one is the natural sorting of the values, the other one is obtained by ranking the values according to the corresponding totals. The obtained result is then applied to get a similar decomposition for the Gini's concentration ratio. This latter decomposition is then compared with other decompositions proposed in literature. Finally the decompositions for the Gini's mean difference and concentration ratio are extended to the more general case of a linear combination of variates.*

Keywords: *mean difference, concentration ratio, decomposition, uniform ranking.*

1. Introduction

In this paper we show, in an elementary way, that the mean difference of the sum Y of the variates $X_1,\ldots,X_j,\ldots,X_c$, can be obtained as the difference of the sum of the mean differences of each variate X_j with a non-negative quantity that measures the departure of the data matrix from the

* Presented at the *International Conference in Memory of Two Eminent Social Scientists: C. Gini and M. O. Lorenz. Their impact in the XX-th century development of probability, statistics and economics* – 23-26 May 2005 – Siena.

[§] Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali – Università degli Studi di Milano-Bicocca – P.zza dell'Ateneo Nuovo, 1, 20126 Milano (e-mail: paolo.radaelli@unimib.it michele.zenga@unimib.it).

The present work reflects the common thinking of the two authors, even if, more specifically, M. Zenga wrote sections 1-3 and P. Radaelli wrote the remaining sections.

uniform ranking (*cograduation*) matrix. The mean difference $\Delta(Y)$ of Y is equal to the sum $\sum_j \Delta(X_j)$ of the mean differences of each variate X_j only when there is a "uniform ranking" among the variates.

By utilizing the decomposition of $\Delta(Y)$ we have decomposed, in an analogous way, the Gini's concentration ratio of a sum of non-negative variates.

The result obtained in this paper shows that the Gini's mean difference, like the standard deviation, may be a useful measure variability when a variate Y can be expressed as a sum of c variates; this is the case, for example, of the decomposition of the income by different sources.

Finally the decomposition is extended to a variate Y obtained as a linear combination of the variates X_j ; $j=1,\dots,c$.

2. Definitions and Notation

Let $X_1,\dots,X_j,\dots,X_c$ be c variates observable on each of the N units of a finite population. In each of the N rows of matrix (1) the values of the c variates are reported:

$$\begin{bmatrix} x'_{11} & \dots & x'_{1j} & \dots & x'_{1c} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x'_{i1} & \dots & x'_{ij} & \dots & x'_{ic} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x'_{N1} & \dots & x'_{Nj} & \dots & x'_{Nc} \end{bmatrix}. \tag{1}$$

From now on, when not differently indicated, the indexes are supposed to assume the values: $i,l=1,\dots,N;\ j=1,\dots,c$.

With $Y=\sum X_j$ we denote the sum of the c variates. The N values of Y arranged in increasing order of magnitude are:

$$y_{(1)} \le \dots \le y_{(i)} \le \dots \le y_{(N)}. \tag{2}$$

The matrix (3) is obtained from matrix (1) permuting the rows according to the N increasing values $y_{(i)}$:

$$\begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1c} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ic} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nj} & \cdots & x_{Nc} \end{bmatrix}. \tag{3}$$

In other words, in matrix (3) we have:

$$x_{i1} + \ldots + x_{ij} + \ldots + x_{ic} = y_{(i)} \quad i = 1, 2, \ldots, N. \tag{4}$$

In (3), the c values on a row belong to one of the N units of the population. Furthermore, the increasing order of the sums $y_{(i)}$ does not imply that the same sorting is fulfilled for each of the c variates, in other words it is not generally true that:

$$x_{1j} \leq \ldots \leq x_{ij} \leq \ldots \leq x_{Nj} \quad j = 1, 2, \ldots, c.$$

The matrix (5) is obtained from matrix (3), arranging in increasing order the values of each column:

$$\begin{bmatrix} x_{(11)} & \cdots & x_{(1j)} & \cdots & x_{(1c)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{(i1)} & \cdots & x_{(ij)} & \cdots & x_{(ic)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{(N1)} & \cdots & x_{(Nj)} & \cdots & x_{(Nc)} \end{bmatrix}. \tag{5}$$

In other words:

$$x_{(1j)} \leq \ldots \leq x_{(ij)} \leq \ldots \leq x_{(Nj)} \quad j = 1, 2, \ldots, c. \tag{6}$$

Matrix (5) can be defined *uniform ranking* (*cograduation*) *matrix* since the values of each variate (a column of the matrix) are ranked in increasing order.
Note that in the matrix (5) the values on a single row do not necessarily refer to the same unit of the population.
Adding up each row of (5), we obtain the theoretical values:

$$y^*_{(i)} = x_{(i1)} + x_{(i2)} + \ldots + x_{(ij)} + \ldots + x_{(ic)} \qquad i = 1, 2, \ldots, N. \tag{7}$$

Gini's mean difference (without repetition) of the variate X that takes the values $x_1, \ldots, x_i, \ldots, x_N$ on the N units of a finite population is given by:

$$\Delta(X) = \frac{1}{N(N-1)} \sum_i \sum_l |x_i - x_l|. \tag{8}$$

It is well known (Gini, 1914) that the statistics:

$$S(X) = \sum_i \sum_l |x_i - x_l| \tag{9}$$

is given by:

$$S(X) = 2\sum_i x_{(i)} (2i - N - 1). \tag{10}$$

3. Decomposition of the Mean Difference of a Sum

The mean difference of the variate $Y = \sum X_j$ is given by:

$$\Delta(Y) = \frac{S(Y)}{N(N-1)} \tag{11}$$

where, according to (10):

$$S(Y) = \sum_i \sum_l |y_i - y_l| = 2\sum_i y_{(i)} (2i - N - 1). \tag{12}$$

Substituting (4) in (12) we have:

$$\begin{aligned} S(Y) &= 2\sum_i \left[\sum_j x_{ij}\right](2i - N - 1) \\ &= \sum_j 2\sum_i x_{ij} (2i - N - 1). \end{aligned} \tag{13}$$

We can rewrite (13) as follows:

$$S(Y)=\sum_j 2\sum_i\left(x_{(ij)}-x_{(ij)}+x_{ij}\right)(2i-N-1)$$
$$=\sum_j 2\sum_i x_{(ij)}(2i-N-1)-2\sum_j\sum_i\left(x_{(ij)}-x_{ij}\right)(2i-N-1). \tag{14}$$

From (10):

$$S\left(X_j\right)=2\sum_i x_{(ij)}(2i-N-1),$$

thus:

$$S(Y)=\sum_j S\left(X_j\right)-2\sum_j\sum_i\left(x_{(ij)}-x_{ij}\right)(2i-N-1). \tag{15}$$

The sum $\sum_j\sum_i\left(x_{(ij)}-x_{ij}\right)(2i-N-1)$ is equal to:

$$2\sum_j\sum_i\left(x_{(ij)}-x_{ij}\right)i-(N+1)\sum_j\sum_i\left(x_{(ij)}-x_{ij}\right)=2\sum_j\sum_i\left(x_{(ij)}-x_{ij}\right)i$$

since $\sum_i\left(x_{(ij)}-x_{ij}\right)=0 \quad j=1,\dots,c.$
Substituting in (15), we get:

$$S(Y)=\sum_j S\left(X_j\right)-4\sum_j\sum_i\left(x_{(ij)}-x_{ij}\right)\, i. \tag{16}$$

By theorem 368 of Hardy et al. (1952, p. 261), the sum:

$$\sum_i i\; x_{ij}$$

is greatest when the values $x_{ij},\quad i=1,\dots,N$ are arranged in increasing order, that is:

$$\sum_i i\; x_{ij}\le\sum_i i\; x_{(ij)}.$$

Therefore, for each $j=1,\dots,c$, we have:

$$\sum_i\left(x_{(ij)}-x_{ij}\right)\, i\ge 0 \tag{17}$$

with equality only if $x_{ij}=x_{(ij)},\quad i=1,\dots,N$.

It derives that:

$$\sum_j \sum_i \left(x_{(ij)} - x_{ij} \right) i \geq 0 \tag{18}$$

with equality only if $x_{ij} = x_{(ij)}, \quad \forall i, j$; that is when the matrix (3) is equal to the uniform ranking matrix (5).
From (16) and (18) we get:

$$S(Y) = S(X_1 + \ldots + X_c) \leq \sum_j S(X_j). \tag{19}$$

The non-negative term $4\sum_j \sum_i \left(x_{(ij)} - x_{ij} \right) i$ can be interpreted as a measure of departure of the data matrix (3) from the uniform ranking (cograduation) matrix (5).
Dividing (16) by $N(N-1)$ we obtain the *subtractive decomposition* for the Gini's mean difference of Y :

$$\Delta(Y) = \sum_j \Delta(X_j) - \frac{4}{N(N-1)} \sum_j \sum_i \left(x_{(ij)} - x_{ij} \right) i. \tag{20}$$

Obviously:

$$\Delta(Y) = \Delta(X_1 + \ldots + X_c) \leq \sum_j \Delta(X_j) \tag{21}$$

with equality only in the case of uniform ranking of the *c* variates.

Zenga (2003) derived from decomposition (20) the following normalized distributive compensation index:

$$C = 1 - \frac{\Delta(Y)}{\sum_j \Delta(X_j)} = 1 - \frac{S(Y)}{\sum_j S(X_j)}$$

whose range is [0; 1] in particular:

- $C = 0$ if there is no compensation among the *c* variates i.e. there is uniform ranking (cograduation);
- $C = 1$ if there is maximum compensation among the *c* variates i.e. *Y* is constantly equal to the arithmetic mean of *Y*.

The compensation index *C* was also investigated by Maffenini (2003) who decomposes *C* in order to evaluate the contribution of each variate to the

overall compensation; furthermore the author studies the index behaviour in the case of independence among the variates and applies the methodology to italian families incomes.
Moreover Borroni and Zenga (2003) propose a test of concordance based on the distributive compensation ratio *C* and compare it with other classical rank correlation methods such as the Spearman's rho and the Kendall's tau. Some developments on the power of this test are the object of the investigation by Borroni and Cazzaro (2005).

4. Decomposition of Gini's Concentration Ratio of a Sum

In this section we assume that the *c* variates X_j are non-negative and that their mean values are positive:

$$M_1(X_j) = \frac{1}{N}\sum_i x_{ij} > 0 \quad j = 1,\ldots,c. \tag{22}$$

Obviously:

$$M_1(Y) = \sum_j M_1(X_j) > 0. \tag{23}$$

Gini's concentration ratio of a non-negative variate X with mean value $M_1\left(X\right) > 0$, is defined by:

$$R(X) = \frac{\Delta(X)}{2M_1(X)}. \tag{24}$$

Dividing both terms in (20) by $2M_1\left(Y\right)$, we obtain:

$$
\begin{aligned}
R(Y) &= \frac{\Delta(Y)}{2M_1(Y)} \\
&= \frac{\Delta(X_1)+\ldots+\Delta(X_c)}{2M_1(Y)} - \frac{4}{2M_1(Y)N(N-1)}\sum_j \sum_i \left(x_{(ij)} - x_{ij}\right) i \\
&= \frac{\frac{\Delta(X_1)}{2M_1(X_1)}2M_1(X_1)+\ldots+\frac{\Delta(X_c)}{2M_1(X_c)}2M_1(X_c)}{2M_1(Y)} - \frac{2\sum_j \sum_i \left(x_{(ij)} - x_{ij}\right) i}{M_1(Y)N(N-1)} \\
&= \frac{R(X_1)M_1(X_1)+\ldots+R(X_c)M_1(X_c)}{M_1(X_1)+\ldots+M_1(X_c)} - \frac{2\sum_j \sum_i \left(x_{(ij)} - x_{ij}\right) i}{M_1(Y)N(N-1)}.
\end{aligned}
$$

The share of the variate X_j on the sum Y is given by:

$$
\omega_j = \frac{\sum_i x_{ij}}{\sum_j \sum_i x_{ij}} = \frac{M_1(X_j)}{M_1(Y)}; \tag{25}
$$

obviously $\sum \omega_j = 1$.

Thus the decomposition of $R(Y)$ can be rewritten as:

$$
R(Y) = \sum_j R(X_j)\ \omega_j - \frac{2\sum_j \sum_i \left(x_{(ij)} - x_{ij}\right) i}{M_1(Y)N(N-1)}. \tag{26}
$$

Equation (26) is an easy decomposition that allows Gini's concentration ratio of a sum to be obtained as the difference of the weighted arithmetic mean of the concentration ratios of each variate X_j with a non-negative quantity that measures the departure of the data matrix (3) from the uniform ranking (cograduation) matrix (5).

5. Comparison with others Gini's Concentration Ratio Decompositions

5.1 *The Decomposition Proposed by Rao*

Rao (1969) proposed two decompositions of Gini's concentration ratio: the first one is by sub-populations, the second one is by components of income[1]. In this section we compare the latter with the decomposition (26).

[1] The subject is also discussed in Kakwani (1980).

Let x_{ij} be the income of the *i*-th family ($i = 1, \ldots, N$) due to the *j*-th component X_j ($j = 1, \ldots, c$). The total income of the *i*-th family is given by: $y_i = \sum_j x_{ij}$.

The whole income composition of the *N* families can be reported in a matrix similar to (1).

Suppose to permute the rows of this matrix so that the families are arranged in increasing order according to their total incomes $y_{(i)}$ (matrix (3)).

The objective of this approach is to explain the inequality of total incomes by the inequalities observable for each of the *c* income sources. In order to reach this result Rao considers, for each income component, two different sortings of the *N* families:

i) in the first one the values of each income component are sorted in increasing order of magnitude; in other words for the *j*-th component we have:

$$x_{(1j)} \le \ldots \le x_{(ij)} \le \ldots \le x_{(Nj)}.$$

The result is a uniform ranking (cograduation) matrix (5);

ii) in the second one the income components are sorted in increasing order according to their total incomes (matrix (3)).

The families are then partitioned, for both sortings, in *k* subsets so that each set includes *N/k* families. In order to make the comparison with decomposition (26) easier we set *k* = *N* so that each subset includes only one family.

Let:

$$\overset{*}{q}'_{ij} = \sum_{t=1}^{i} x_{tj} \qquad i = 1, \ldots, N$$

indicate the cumulative sums of the *j*-th component incomes in matrix (3) and let:

$$q'_{ij} = \sum_{t=1}^{i} x_{(tj)} \qquad i = 1, \ldots, N$$

be the same cumulative sums in matrix (5).

The value of Gini's concentration ratio for the *j*-th income component (*j*-th column of matrix (5)) $R(X_j) = \Delta(X_j) / 2M_1(X_j)$ can be obtained by the following formula related to the Lorenz curve (Gini (1914)):

$$R(X_j) = \frac{\sum_{i=1}^{N-1} \left[i\ M_1(X_j) - q'_{ij} \right]}{\sum_{i=1}^{N-1} i\ M_1(X_j)} = \frac{\sum_{i=1}^{N-1} \left(p_i - q_{ij} \right)}{\sum_{i=1}^{N-1} p_i} \tag{27}$$

where $p_i = i/N$ and $q_{ij} = q'_{ij} / \left[N\ M_1(X_j) \right]$.

Expression (27) is applied by Rao also on the cumulative sums $\overset{*}{q}'_{ij}$ obtaining the statistics:

$$\overset{*}{R}(X_j) = \frac{\sum_{i=1}^{N-1} \left[i\ M_1(X_j) - \overset{*}{q}'_{ij} \right]}{\sum_{i=1}^{N-1} i\ M_1(X_j)} = \frac{\sum_{i=1}^{N-1} \left(p_i - \overset{*}{q}_{ij} \right)}{\sum_{i=1}^{N-1} p_i} \tag{28}$$

where $\overset{*}{q}_{ij} = \overset{*}{q}'_{ij} / \left[N\ M_1(X_j) \right]$.

Note that $\overset{*}{R}(X_j)$ is not Gini's concentration ratio since the *N* elements x_{ij} in a column of matrix (3) may not necessarily be ranked by increasing order[2] as are the elements in a column of matrix (5).
Furthermore Rao shows that:

$$-R(X_j) \le \overset{*}{R}(X_j) \le R(X_j) \tag{29}$$

where the lower and the upper bounds are reached respectively if in the *j*-th column of matrix (3) the *N* families *j*-th incomes are in descending or ascending order.
Rao shows that the concentration ratio $R(Y)$, computed on the total incomes, can be obtained as the difference of the weighted arithmetic mean of components concentration ratios, with weights given by the shares (25) of each component on the total income, with a non-negative quantity that Rao

[2] It is not necessarily true that: $p_i \ge \overset{*}{q}_{ij} \quad i = 1, \ldots, N-1.$

defines "*an overall measure of the extent to which component-inequalities offset each other*". Using the notation above, Rao's decomposition is:

$$R(Y)=\sum_j R\left(X_j\right)\,\omega_j-\sum_j R\left(X_j\right)\,\omega_j\left[1-\frac{\overset{*}{R}(X_j)}{R\left(X_j\right)}\right]. \qquad (30)$$

Comparing decompositions (26) and (30) we observe that, for both, we have to subtract a non-negative term from the weighted arithmetic mean of concentration ratios computed on each income component. These terms are respectively:

$$\frac{2\sum_j\sum_i\left(x_{(ij)}-x_{ij}\right)\,i}{M_1(Y)N(N-1)} \qquad (31)$$

and:

$$\sum_j R\left(X_j\right)\,\omega_j\left[1-\frac{\overset{*}{R}(X_j)}{R\left(X_j\right)}\right]. \qquad (32)$$

Obviously (31) is equal to (32) but they are different in the way they have been obtained and in their interpretation. The term depends, in both cases, on the different families sorting for individual components with respect to the one obtained for the total income. The interpretation is clear in (31) given that we consider the individual weighted differences $\left(x_{ij}-x_{(ij)}\right)i$, while on the contrary, in (32) the interpretation is not clear since the term to be subtracted depends on the ratios $\overset{*}{R}(X_j)\big/R\left(X_j\right)$. Furthermore, we do not know whether it is advisable to compute "concentration ratios" on values in non ascending order and obtain values which may be negative for a measure that by definition (and traditional use in literature) should lie in the interval [0; 1]. In other words the measure $\overset{*}{R}(X_j)$ can be hardly explained.

5.2 *The Decomposition Proposed by Lerman and Yitzhaki*

Lerman and Yitzhaki (1984 and 1985) propose a decomposition of the overall Gini concentration ratio by income sources. In particular the authors show that each source's contribution to the Gini coefficient may be viewed as the product of three factors:

- the source's Gini coefficient;
- the source's share of total income;
- the Gini correlation between the source and the rank of total income.

The point of departure is the relationship between the Gini's mean difference and the covariance; this relation was pointed out by De Vergottini (1950)[3] in a paper concerning a general expression for concentration indexes[4] despite it is frequently ascribed to Stuart (1954).[5]

This relation states that the Gini's mean difference with repetition of a variable *X* is equal to four times the covariance between the variable and its rank:

$$\Delta'(X) = 4Cov\left[X, F_X(X)\right] \tag{33}$$

where F_X denotes the cumulative distribution of *X*.

For a variate X that takes the values $x_1, \ldots, x_i, \ldots, x_N$ on the *N* units of a finite population (33) can be rewritten as:

$$\begin{aligned}\Delta'(X) &= \frac{1}{N^2}\sum_i \sum_l \left|x_i - x_l\right| = 4\ Cov\left[X; \frac{r(X)}{N}\right] \\ &= \frac{2}{N^2}\sum_i x_i\left[2\ r(x_i) - N - 1\right] = \frac{2}{N^2}\sum_i x_{(i)}\left(2\ i - N - 1\right)\end{aligned} \tag{34}$$

where $r(x_i)$ is the rank of *i*-th value.

In this framework, Yitzhaki and Olkin (1991) (see also Olkin and Yitzhaki (1992) and Schechtman and Yitzhaki (1999)) define, for two random variables *X* and *Y* with continuous distribution functions F_X and F_Y, respectively, and a continuous bivariate distribution $F_{X,Y}$, the Gini covariance between *X* and *Y* as:

$$G\,\text{cov}\left(X, Y\right) \equiv Cov\left[X; F_Y(Y)\right]. \tag{35}$$

A measure of association between *X* and *Y*, see Schechtman and Yitzhaki

[3] See also Zenga (1987, p. 47).

[4] This approach has been followed also in Dancelli (1987).

[5] See for example David (1968), Lerman and Yitzhaki (1984) and Balakrishnan and Rao (1998, p. 497).

(1987) for details, can be defined as:

$$\Gamma\left(X,Y\right) \equiv \frac{Cov\left[X;F_Y(Y)\right]}{Cov\left[X;F_X(X)\right]} \tag{36}$$

and between *Y* and *X* as:

$$\Gamma\left(Y,X\right) \equiv \frac{Cov\left[Y;F_X(X)\right]}{Cov\left[Y;F_Y(Y)\right]}. \tag{37}$$

In our framework, the mean difference (with repetition) of the variate $Y = \sum_j X_j$ is, according to (33):

$$\begin{aligned} \Delta'(Y) &= 4Cov\left[Y,F_Y(Y)\right] \\ &= 4\sum_j Cov\left[X_j,F_Y(Y)\right] \\ &= 4\sum_j \frac{Cov\left[X_j,F_Y(Y)\right]}{Cov\left[X_j,F_{X_j}(X_j)\right]}\, Cov\left[X_j,F_{X_j}(X_j)\right] \\ &= \sum_j \Gamma_j\ \Delta'(X_j) \end{aligned} \tag{38}$$

where:

$$\Gamma_j = \Gamma\left[X_j,Y\right] = \frac{Cov\left[X_j,F_Y(Y)\right]}{Cov\left[X_j,F_{X_j}(X_j)\right]} \tag{39}$$

is the Gini correlation (36) between the *j*-th component X_j and the sum Y.
It must be observed that (39) is equivalent to the ratio between Rao's (28) and (27):

$$\Gamma_j = \frac{\overset{*}{R}(X_j)}{R\left(X_j\right)}. \tag{40}$$

In order to compare (38) with the decomposition here proposed, we rewrite (20) for the mean difference with repetition:

$$\Delta'(Y) = \sum_j \Delta'(X_j) - \frac{4}{N^2}\sum_j \sum_i \left(x_{(ij)} - x_{ij}\right) i \tag{41}$$

and (38) as:

$$\Delta'(Y) = \sum_j \Delta'(X_j) - \sum_j \Delta'(X_j)\left(1-\Gamma_j\right) \tag{42}$$

Clearly:[6]

$$\frac{4}{N^2}\sum_j \sum_i \left(x_{(ij)} - x_{ij}\right) i = \sum_j \Delta'(X_j)\left(1-\Gamma_j\right).$$

For a fixed *j*:

$$\frac{4}{N^2}\sum_i \left(x_{(ij)} - x_{ij}\right) i = \Delta'(X_j)\left(1-\Gamma_j\right) \tag{43}$$

vanishes if and only if $\Gamma_j = +1$ that is when there is a perfect positive Gini correlation between X_j and Y or, equivalently, when X_j and Y are cograduated.
This comparison highlights that the term to be subtracted to the sum of the mean differences of the variates X_j, ($j = 1,\dots,c$) in order to obtain the mean difference of Y in (41) can be interpreted as a measure of departure from the situation of perfect positive Gini correlation between each variate X_j and the sum Y.

Dividing both terms in (38) by $2M_1(Y)$, we obtain the decomposition for the Gini's concentration ratio:

6

$$\begin{aligned}\sum_j \Delta'(X_j)\left(1-\Gamma_j\right) &= \sum_j 4Cov\left[X_j; F_{X_j}(X_j)\right]\left(1 - \frac{Cov\left[X_j; F_Y(Y)\right]}{Cov\left[X_j; F_{X_j}(X_j)\right]}\right)\\ &= 4\sum_j \left\{Cov\left[X_j; F_{X_j}(X_j)\right] - Cov\left[X_j; F_Y(Y)\right]\right\}\\ &= \frac{2}{N^2}\sum_j \left\{\sum_i x_{ij}\left[2r(x_{ij}) - N - 1\right] - \sum_i x_{ij}\left[2r(y_i) - N - 1\right]\right\}\\ &= \frac{4}{N^2}\sum_j \sum_i x_{ij}\left[r(x_{ij}) - r(y_i)\right] = \frac{4}{N^2}\sum_j \sum_i \left(x_{(ij)} - x_{ij}\right) i.\end{aligned}$$

$$
\begin{aligned}
R(Y) &= \frac{\Delta'(Y)}{2M_1(Y)} = \sum_j \Gamma_j \; \frac{\Delta'(X_j)}{2M_1(Y)} \\
&= \sum_j \Gamma_j \; \frac{\Delta'(X_j)}{2M_1(X_j)} \frac{M_1(X_j)}{M_1(Y)} \\
&= \sum_j \Gamma_j \; R(X_j) \; \omega_j
\end{aligned} \tag{44}
$$

where ω_j is the *j*-th income component share (25).

In order to point out the relation between Rao's and Lerman and Yitzhaki's decompositions, we observe that (44) can be rewritten as:

$$
R(Y) = \sum_j R(X_j) \; \omega_j - \sum_j R(X_j) \; \omega_j \left(1 - \Gamma_j\right) \tag{45}
$$

that, in view of (40), becomes:

$$
R(Y) = \sum_j R\left(X_j\right) \; \omega_j - \sum_j R\left(X_j\right) \; \omega_j \left[1 - \frac{\overset{*}{R}(X_j)}{R\left(X_j\right)}\right] \tag{46}
$$

which is Rao's decomposition (30).

6. Decomposition of the Mean Difference of a Linear Combination

In this section we provide an extension of the decomposition shown in section 3 to the more general case of a linear combination of variates.
Let

$$
Y = \alpha_1 X_1 + \ldots + \alpha_c X_c = \sum_j \alpha_j X_j \tag{47}
$$

denotes the linear combination of the c variates X_j with coefficients $\alpha_j \neq 0 \quad (j = 1, \ldots, c)$.
If we introduce:

$$
Z_j = \alpha_j X_j \quad j = 1, \ldots, c
$$

Y in (47) is simply the sum of the "new" variates Z_j $(j=1,\dots,c)$ so we can decompose Gini's mean difference of Y according to (20) as:

$$\Delta(Y)=\Delta(Z_1+\dots+Z_c)=\sum_j \Delta(Z_j)-\frac{4}{N(N-1)}\sum_j\sum_i\left(z_{(ij)}-z_{ij}\right)i \quad (48)$$

where, for each j, the values z_{ij} are sorted according to the their increasing total $y_{(i)}$ and the values $z_{(ij)}$ are sorted themselves.

In order to express $\Delta(Y)$ as a function of the original variates X_j we observe that:

- $$\Delta(Z_j)=|\alpha_j|\Delta(X_j) \quad j=1,\dots,c; \quad (49)$$
- $$z_{ij}=\alpha_j x_{ij}=\begin{cases}\alpha_j x_{(ij)} & \text{if } \alpha_j>0\\ \alpha_j x_{(N-i+1\ j)} & \text{if } \alpha_j<0\end{cases} \quad j=1,\dots,c; \quad i=1,\dots,N. \quad (50)$$

If we define the function:

$$\begin{aligned} i_j &= \frac{N+1}{2}+\left[\frac{2i-N-1}{2}\right]\text{sgn}(\alpha_j) \\ &= \begin{cases} i & \text{if } \alpha_j>0 \\ N-i+1 & \text{if } \alpha_j<0\end{cases} \quad j=1,\dots,c; \quad i=1,\dots,N \end{aligned} \quad (51)$$

where:

$$\text{sgn}(k)=\begin{cases}+1 & \text{if } k>0\\ -1 & \text{if } k<0\end{cases}$$

it is possible to get a unique expression for (50):

$$z_{ij}=\alpha_j x_{(i_j j)} \quad j=1,\dots,c; \quad i=1,\dots,N. \quad (52)$$

Finally (48) can be rewritten with respect to the original variates X_j as follows:

$$\Delta(Y)=\sum_j|\alpha_j|\Delta(X_j)-\frac{4}{N(N-1)}\sum_j\alpha_j\sum_i\left(x_{(i_j j)}-x_{ij}\right)i. \quad (53)$$

7. Decomposition of Gini's Concentration Ratio of a Linear Combination

As in section 4 we assume that the c variates Z_j are non-negative and that their mean values are positive:

$$M_1(Z_j) = \frac{1}{N}\sum_i z_{ij} > 0 \quad j = 1,\ldots,c \tag{54}$$

so that:

$$M_1(Y) = \sum_j M_1(Z_j) > 0. \tag{55}$$

This means that, with respect to the original variates X_j, we should have, for each $j = 1,\ldots,c$:

$$\begin{array}{ll} \alpha_j > 0 & \text{if } X_j \geq 0 \\ \alpha_j < 0 & \text{if } X_j < 0 \end{array}.$$

From now on we suppose, without loss of generality, that $\alpha_j > 0$ and $X_j \geq 0$ for $j = 1,\ldots,c$.
Gini's concentration ratio of Y can be decomposed according to (26) as follows:

$$R(Y) = \sum_j R\left(Z_j\right)\,\omega_j - \frac{2\sum_j \sum_i \left(z_{(ij)} - z_{ij}\right) i}{M_1(Y) N(N-1)} \tag{56}$$

where:

$$\omega_j = \frac{M_1(Z_j)}{M_1(Y)}$$

denotes the share of the variate Z_j on the sum Y.

With respect to the original variates X_j we have:

$$R(Z_j) = \frac{\Delta(Z_j)}{2M_1(Z_j)} = \frac{\alpha_j \Delta(X_j)}{2\alpha_j M_1(Z_j)} = R(X_j) \quad j = 1,\dots,c$$

and:

$$M_1(Y) = \sum_j M_1(Z_j) = \sum_j \alpha_j M_1(X_j).$$

Finally we can rewrite decomposition (56) as follows:

$$R(Y) = \sum_j R\left(X_j\right)\, \omega_j - \frac{2\sum_j \alpha_j \sum_i \left(x_{(ij)} - x_{ij}\right)\, i}{N(N-1)\sum_j \alpha_j M_1\left(X_j\right)}. \qquad (57)$$

8. Concluding remarks

In this paper we show an easy subtractive decomposition for the Gini's mean difference $\Delta(Y)$ of a variate Y obtained as the sum of *c* variates. In this decomposition the uniform ranking (cograduation) matrix (5) plays a central role given that Gini's mean difference of the sum is no greater than the sum of Gini's mean differences of the variates added up with equality only if the data matrix (3) is a uniform ranking (cograduation) matrix.
By utilizing the decomposition of $\Delta(Y)$ we get a simple analogous decomposition for the Gini's concentration ratio $R(Y)$.
Furthermore the decomposition here proposed for $R(Y)$ is compared with the decompositions proposed by Rao (1969) and Lerman and Yitzhaki (1984; 1985).
Finally in sections 6 and 7 the decompositions of the Gini's mean difference and concentration ratio are extended to the more general case of a linear combination of variates.

Acknowledgement

This work is partially supported by F.A.R. 2003 – Università degli Studi di Milano-Bicocca.

References

Balakrishnan, N. and Rao, C. (1998). *Order Statistics: Theory & Methods*, vol. 16 of Handbook of Statistics. North-Holland.

Borroni, C. and Cazzaro, M. (2005). Some Developments about a New Nonparametric Test Based on Gini's Mean Difference. (presented in *International Conference in Memory of Two Eminent Social Scientists: C. Gini and M. O. Lorenz. Their impact in the XX-th century development of probability, statistics and economics* – 23-26 May - Siena).

Borroni, C. and Zenga, M. (2003). A Test of Concordance Based on the Distributive Compensation Ratio. *Rapporto di Ricerca 51, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali* – Università degli Studi di Milano-Bicocca.

Dancelli, L. (1987). In Tema di Relazioni e di Discordanze fra Indici di Variabilità e di Concentrazione. In Zenga, M., editor, *La Distribuzione Personale del Reddito:Problemi di formazione, di ripartizione e di misurazione*. Vita e Pensiero.

David, H. (1968). Gini's Mean Difference Rediscovered. *Biometrika*, **55**, 573–575.

De Vergottini, M. (1950). Sugli Indici di Concentrazione. *Statistica*, **X**, 445–454.

Gini, C. (1914). Sulla Misura della Concentrazione e della Variabilità dei Caratteri. In *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*. Anno Accademico 1913-1914 -, Tomo LXXIII - Parte seconda, 1201–1248. Venezia -Premiate Officine Grafiche C. Ferrari.

Hardy, G., Littlewood, J., and Pólya, G. (1952). *Inequalities*. Cambridge University Press, 2nd edition.

Kakwani, N. (1980). *Income Inequality and Poverty. Methods of Estimation and Policy Applications*. Oxford University Press.

Lerman, R. and Yitzhaki, S. (1984). A Note on The Calculation and Interpretation of the Gini Index. *Economics Letters*, **15**, 363–368.

Lerman, R. and Yitzhaki, S. (1985). Income Inequality Effects by Income Source: A New Approach and Applications to the United States. *The Review of Economics and Statistics*, **67**, 151–156.

Maffenini, W. (2003). Osservazioni sull'Indice di Compensazione Distributiva. *Rapporto di Ricerca 57, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali* - Università degli Studi di Milano-Bicocca. (Forthcoming in *Statistica*).

Olkin, I. and Yitzhaki, S. (1992). Gini Regression Analysis. *International Statistical Review*, **60**, 185–196.

Radaelli, P. and Zenga, M. (2002). Decomposition of the Mean Difference of the Sum of K Variables. Abstract of contributed papers. Fourth International Conference on the L_1-Norm and Related Methods - University of Nuechâtel, Switzerland.

Rao, V. (1969). Two Decompositions of Concentration Ratio. *Journal of the Royal Statistical Society*. Series A (General), **132**, 418–425.

Schechtman, E. and Yitzhaki, S. (1987). A Measure of Association Based on Gini's Mean Difference. *Communications in Statistics*, **16**, 207–231.

Schechtman, E. and Yitzhaki, S. (1999). On the Proper Bounds of the Gini Correlation. *Economics Letters*, **63**, 133–138.

Stuart, A. (1954). The Correlation Between Variate-Values and Ranks in Samples From a Continuous Distribution. *The British Journal of Statistical Psychology*, **VII**, 37–44.

Yitzhaki, S. and Olkin, I. (1991): Concentration Indeces and Concentration Curves. In Mosler, K. and Scarsini, M., editors, *Stochastic Orders and Decision Under Risk*, volume 19 of Lecture Notes - Monograph Series. Institute of Mathematical Statistics - Hayward California.

Zenga, M. (1987). Concentration Measures. In Naddeo, A., editor, *Italian Contributions to the Methodology of Statistics*, 42–51. Società Italiana di Statistica - Cleup Padova.

Zenga, M. (2003). Distributive Compensation Ratio Derived from the Decomposition of the Mean Difference of a Sum. *Statistica & Applicazioni*, **I**(1), 19–28.

Zenga, M. and Radaelli, P. (2002). Decomposition of the Mean Difference of the Sum of Variates. *Rapporto di Ricerca 42, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali* - Università degli Studi di Milano-Bicocca.