Decomposition of Zenga's inequality index I

Paolo Radaelli

Dipartimento di Metodi Quantitativi

per le Scienze Economiche ed Aziendali

Università degli Studi di Milano-Bicocca

Key words: inequality, uniformity, subgroups decomposition.

1 Introduction

This paper concerns some notes about the inequality index based on the ratios between lower and upper arithmetic means recently proposed by Zenga [2006].

Beside some definitions and notation given in section 2, we show (section 3) that the inequality measure proposed fulfills the population replication principle.

In the remainder of the paper our purpose is to consider a population partitioned in c subgroups. In this case usually the common aim is to evaluate the within and between subgroups contributions to the overall uniformity (inequality). The first decomposition, derived in section 4, allows to evaluate the contribution of each subgroup to the overall uniformity, this contribution comprises both the within and the between sources.

The key point of the main decomposition of the overall uniformity proposed in section 5 is the introduction of the uniformity point index (5.1) which can be evaluated either within the same subgroup or between two different subgroups such that a within/between subgroups uniformity decomposition arises.

Moreover section 5.1 provides an interesting representation of the previous decomposition: the overall uniformity index is obtained as a weighted average of two components related, respectively, to the within and to the between contributions. This representation allows to obtain an analogous decomposition of the overall inequality index (section 6).

Finally section 7 is devoted to conclusions and further developments.

2 Definitions and notation

Given a distribution frequency:

$$\left\{(x_i, n_i):\ i = 1, \ldots, r;\ 0 \leqslant x_1 < \ldots < x_r;\ \sum n_i = N\right\} \tag{2.1}$$

of a non negative variable X, let:

$$N_i = \sum_{t=1}^{i} n_t \qquad i = 1, \ldots, r \tag{2.2}$$

$$R_i = N - N_{i-1} = \sum_{t=i}^{r} n_t \qquad i = 1, \ldots, r \tag{2.3}$$

$$Q_i = \sum_{t=1}^{i} x_t n_t \qquad i = 1, \ldots, r \tag{2.4}$$

$$T = Q_r = \sum_{i=1}^{r} x_i n_i > 0 \tag{2.5}$$

$$M = \frac{T}{N} \tag{2.6}$$

$$W_i = T - Q_{i-1} = \sum_{t=i}^{r} x_t n_t > 0 \qquad i = 1, \ldots, r. \tag{2.7}$$

Given $i \in \{1, \ldots, r\}$, suppose to split the distribution into two groups: a lower group $\{(x_1, n_1), \ldots, (x_i, n_i)\}$ and an upper group $\{(x_i, n_i), \ldots, (x_r, n_r)\}$; recently Zenga [2006] has proposed to measure the point inequality between these two groups by means of the ratio of the correspondent arithmetic means:

$$\bar{M}_i = \frac{Q_i}{N_i} = \frac{1}{N_i} \sum_{t=1}^{i} x_t n_t \qquad i = 1, \ldots, r \tag{2.8}$$

$$\overset{+}{M}_i = \frac{W_i}{R_i} = \frac{1}{N - N_{i-1}} \sum_{t=i}^{r} x_t n_t \qquad i = 1, \ldots, r. \tag{2.9}$$

The point inequality index proposed is:

$$I_i = \frac{\overset{+}{M}_i - \overset{-}{M}_i}{\overset{+}{M}_i} = 1 - \frac{\overset{-}{M}_i}{\overset{+}{M}_i} = 1 - U_i \qquad i = 1, \ldots, r \tag{2.10}$$

where:

$$U_i = \frac{\overset{-}{M}_i}{\overset{+}{M}_i} = \frac{Q_i}{N_i} \frac{R_i}{W_i} \qquad i = 1, \ldots, r \tag{2.11}$$

measures the uniformity between the lower and the upper group i.e. the value $U_i \cdot 100$ gives the percentage of $\overset{-}{M}_i$ in terms of $\overset{+}{M}_i$.

Both I_i and U_i lie in the interval $[0; 1]$; in particular $I_i = 0$ means no inequality between lower and upper groups and $I_i = 1$ means maximum inequality (i.e. lower group mean is null).

The author also proposes an inequality diagram in the unit square and derives the synthetic inequality measure as the weighted arithmetic mean of the point measures I_i with weights $\frac{n_i}{N}$

$$I = \sum_{i=1}^{r} I_i \cdot \frac{n_i}{N}. \tag{2.12}$$

The minimum value of I is 0 and it is reached in the case of no inequality ($r = 1$ in the original distribution).

In the case of maximum inequality $\{(x_1 = 0, n_1 = N - 1), (x_2 = T, n_2 = 1)\}$ the index I reaches its maximum value $1 - \frac{1}{N^2}$.

One of the main feature of the point and synthetic inequality measures proposed is the ease of interpretation.

3 The synthetic inequality measure I and the population replication principle

Zenga [2006] proves that the index I fulfills the elementary properties which inequality measures are usually assumed to possess; in particular:

1. in the case of absence of inequality $I = 0$;
2. in the case of maximum inequality the value of the index must be equal to an increasing function C_N of N such that: $\lim_{N\to\infty} C_N = 1$;
3. scale independence;
4. equal additions (subtractions) decrease (increase) I;
5. Pigou-Dalton transfers principle.

In this section we deal with the population replication principle[1] which requires the inequality of a given distribution to be the same as that of the distribution obtained by replicating any number of times each individual value in the initial distribution. In other words: if $\{x_1, \ldots, x_N\}$ is the initial distribution, an inequality synthetic measure fulfills the population replication principle if its value evaluated on the initial distribution is equal to the one evaluated on the distribution obtained by replicating k times the initial one, where k is a positive integer.
In our case replicating k times the initial distribution (2.1) leads to the frequency distribution:

$$\left\{(x_i, k \cdot n_i) : \; i = 1, \ldots, r; \; 0 \leqslant x_1 < \ldots < x_r; \; \sum k \cdot n_i = k \cdot N\right\}. \tag{3.1}$$

The values of the arithmetic means of the lower group (2.8) and of the upper group (2.9) are obviously unaffected by the replication and thus the point inequality indexes (2.10) remain unchanged.

[1]This property is also known as: principle of proportionate additions to persons (Dalton [1920, p. 357]); symmetry axiom for population or Dalton population principle (see Deutsch and Silber [1999]); invariance to the population replication (see Zenga [1986]).

The value of the synthetic inequality measure I evaluated on the replicated distribution (3.1) is consequently the same as the one obtained in the initial distribution given that the weights n_i/N in (2.12) are unchanged.
In conclusion we can say that the inequality measure I fulfills the population replication principle.
We conclude this section by remembering that others inequality measure such as the Bonferroni index or the widespread Gini index do not fulfill the population replication principle, in fact Zenga [1986] proved that if a normalized inequality measure assumes value one if and only if there is maximum inequality, then it does not fulfill the population replication principle.

4 Population's subgroups contribution to the overall uniformity

Suppose to observe a non negative variable X on c different subgroups and let $\{x_1, \ldots, x_i, \ldots, x_r\}$ denote the distinct values assumed by the variable X on all the c subgroups.
It is possible to represent the whole distribution in the following table:

	subgroup					
X	1	$\ldots$	j	$\ldots$	c	tot
x_1	n_{11}	$\ldots$	n_{1j}	$\ldots$	n_{1c}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	n_{i1}	$\ldots$	n_{ij}	$\ldots$	n_{ic}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_r	n_{r1}	$\ldots$	n_{rj}	$\ldots$	n_{rc}	$n_{r\cdot}$
tot	$n_{\cdot 1}$	$\ldots$	$n_{\cdot j}$	$\ldots$	$n_{\cdot c}$	N

Table 4.1: Table of the whole distribution of the c groups.

where n_{ij} denotes the frequency of the value x_i in subgroup j (note that n_{ij} is zero if the variable X does not take the value x_i in the j-th subgroup).
Let us define, for the j-th group distribution $\{(x_i, n_{ij}) : \ i = 1, \dots, r\}$:

$$ {}_jN_i = \sum_{k=1}^{i} n_{kj} \quad \text{the cumulative frequencies} \tag{4.1} $$

$$ {}_jR_i = n_{\cdot j} - {}_jN_{i-1} = \sum_{k=i}^{r} n_{kj} \quad \text{the retro cumulative frequencies} \tag{4.2} $$

$$ {}_jQ_i = \sum_{k=1}^{i} x_k n_{kj} \quad \text{the cumulative sum of units} \tag{4.3} $$

$$ {}_jW_i = {}_rQ_i - {}_jQ_{i-1} = \sum_{k=i}^{r} x_k n_{kj} \quad \text{the retro cumulative sum of units.} \tag{4.4} $$

The point uniformity indexes for the j-th group are defined as:

$$ {}_jU_i = \begin{cases} \dfrac{{}_j\overset{-}{M}_i}{{}_j\overset{+}{M}_i} = \dfrac{{}_jQ_i}{{}_jN_i}\,\dfrac{{}_jR_i}{{}_jW_i} & \text{if}^2\, {}_jN_i > 0 \text{ and } {}_jR_i > 0; \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, \dots, r, \tag{4.5} $$

so that the uniformity index for this group is given by the weighted average of the ratios (4.5) with weights given by the correspondent relative frequencies in the group:

$$ {}_jU = \sum_{i=1}^{r} {}_jU_i \, \frac{n_{ij}}{n_{\cdot j}} \tag{4.6} $$

and then the related inequality index is:

$$ {}_jI = 1 - {}_jU. \tag{4.7} $$

For the marginal distribution of X, $\{(x_i, n_{i\cdot}) : \ i = 1, \dots, r\}$ we have, with a slight

[2] Lower and upper groups must be non-empty.

modification of the notation introduced in section 2:

$$N_i = \sum_{k=1}^{i} n_{k\cdot} = \sum_{j=1}^{c} {}_jN_i \tag{4.8}$$

$$R_i = N - N_{i-1} = \sum_{k=i}^{r} n_{k\cdot} = \sum_{j=1}^{c} {}_jR_i \tag{4.9}$$

$$Q_i = \sum_{k=1}^{i} x_k n_{k\cdot} = \sum_{j=1}^{c} {}_jQ_i \tag{4.10}$$

$$W_i = Q_r - Q_{i-1} = \sum_{k=i}^{r} x_k n_{k\cdot} = \sum_{j=1}^{c} {}_jW_i. \tag{4.11}$$

The overall uniformity indexes are thus given by:

$$U_i = \frac{Q_i}{N_i}\,\frac{R_i}{W_i} \qquad i = 1, \ldots, r \tag{4.12}$$

and the correspondent synthetic uniformity and inequality measures are:

$$U = \sum_{i=1}^{r} U_i\,\frac{n_{i\cdot}}{N} \tag{4.13}$$

and

$$I = 1 - U. \tag{4.14}$$

We are now interested in examining the link between the uniformity (inequality) in the overall distribution and the values obtained for population subgroups.
For a fixed $i \in \{1, \ldots, r\}$ we have, from equations (4.8-4.12):

$$\begin{aligned} U_i =& \frac{Q_i}{N_i}\,\frac{R_i}{W_i} = \left\{\sum_{j=1}^{c} {}_jQ_i\right\}\,\frac{R_i}{W_iN_i} = \left\{\sum_{j=1}^{c} \frac{{}_jQ_i}{{}_jW_i}\,{}_jW_i\right\}\,\frac{R_i}{W_iN_i} \\ =& \left\{\sum_{j=1}^{c}\left[\frac{{}_jQ_i}{{}_jW_i}\frac{{}_jR_i}{{}_jN_i}\right]\frac{{}_jN_i}{{}_jR_i}{}_jW_i\right\}\,\frac{R_i}{W_iN_i}. \end{aligned} \tag{4.15}$$

The quantity within square brackets in (4.15) is the i-th uniformity index (4.5) of the j-th subgroup, thus:

$$U_i = \left\{ \sum_{j=1}^{c} {}_jU_i \frac{{}_jN_i}{{}_jR_i} {}_jW_i \right\} \frac{R_i}{W_i N_i}$$

$$= \sum_{j=1}^{c} {}_jU_i \frac{{}_jW_i}{W_i} \frac{{}_jN_i}{N_i} \frac{R_i}{{}_jR_i}. \tag{4.16}$$

Equation (4.16) decomposes the overall uniformity point index U_i in a weighted sum of the subgroups uniformity point indexes (4.5); the weights are proportional to the j-th subgroup shares of the retro cumulative sums of units (4.4) and of the cumulative frequencies (4.1) and inverse proportional to the share of the retro cumulative frequencies (4.2).

This decomposition allows to evaluate the contribution of each of the c subgroups to the overall point uniformity by considering the shares:

$$\frac{1}{U_i} \left[{}_jU_i \frac{{}_jW_i}{W_i} \frac{{}_jN_i}{N_i} \frac{R_i}{{}_jR_i} \right] \quad j = 1, \dots, c. \tag{4.17}$$

Considering now the overall uniformity index U, by substituting (4.16) in (4.13) we obtain:

$$U = \sum_{i=1}^{r} U_i \frac{n_{i\cdot}}{N} = \sum_{i=1}^{r} \sum_{j=1}^{c} {}_jU_i \frac{{}_jW_i}{W_i} \frac{{}_jN_i}{N_i} \frac{R_i}{{}_jR_i} \frac{n_{i\cdot}}{N}$$

$$= \sum_{j=1}^{c} \left[\sum_{i=1}^{r} {}_jU_i \frac{{}_jW_i}{W_i} \frac{{}_jN_i}{N_i} \frac{R_i}{{}_jR_i} \frac{n_{i\cdot}}{N} \right]. \tag{4.18}$$

Once more the contribution of each of the c subgroups to the overall uniformity can be evaluated by the shares:

$$\frac{1}{U} \left[\sum_{i=1}^{r} {}_jU_i \frac{{}_jW_i}{W_i} \frac{{}_jN_i}{N_i} \frac{R_i}{{}_jR_i} \frac{n_{i\cdot}}{N} \right] \quad j = 1, \dots, c. \tag{4.19}$$

5 On the within/between subgroups decomposition of the overall uniformity

One common application of inequality measures involves the study of the relationship between the overall inequality value and the ones obtained for population subgroups. The ideal decomposition is the one that allows to obtain the overall inequality as the sum of only two components measuring, respectively, the within and the between subgroups inequality. An inequality measure is called aggregative if the knowledge of the inequality measures of the subgroups and their aggregative characteristic (mean and numerousness) suffices to compute the overall inequality. This aggregative property requires the within term to be a weighted sum of the subgroups inequality values and the between term to be a function of only the means and the number of individuals in each group. It is well known (Shorrocks [1980]) that this ideal decomposition can be reached only for measures belonging to the one parameter (α) family of generalized entropy inequality measures[3]:

$$
\begin{aligned}
I^{(\alpha)} &= \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{M}\right)^{\alpha} - 1\right] \qquad (\alpha \neq 0, \alpha \neq 1)\\
I^{(0)} &= -\frac{1}{N}\sum_{i=1}^{N}\log\frac{x_i}{M}\\
I^{(1)} &= \frac{1}{N}\sum_{i=1}^{N}\frac{x_i}{M}\log\frac{x_i}{M}
\end{aligned}
$$

studied by Deutsch and Silber [1999] and Zenga [1986]. For other measures this representation is not obtainable: the Gini concentration ratio decomposition, for instance, includes a third term (see Dagum [1997]); the Bonferroni index decomposition studied by Tarsitano [1990] includes a third term as well and does not depend only on the aggregative characteristic of the subgroups.

In this section we will obtain a within-between subgroups decomposition of the overall uniformity by considering the configuration scheme reported in table 4.1.

Focusing the attention on the i-th value x_i, it is possible to compute the point uniformity index ${}_jU_i$ defined in equation (4.5) within the j-th subgroup by the ratio of the correspondent lower and upper means.

As illustrated in section 2, the ratio ${}_jU_i$ measures the uniformity between the lower group (values $x \leqslant x_i$) and the upper one (values $x \geqslant x_i$) in, we need to specify it

[3] We set, for simplicity, $r = N;\ n_i = 1,\ i = 1, \ldots, N$ in the frequency distribution (2.1).

now, the j-th subgroup.

The novelty is that the lower group of the j-th subgroup (values $x \leqslant x_i$ in subgroup j) can also be compared with the correspondent upper group in the h-th subgroup (values $x \geqslant x_i$ in subgroup h) by the cross ratio $\frac{{}_j\overset{-}{M}_i}{{}_h\overset{+}{M}_i}$.

So we can define the point uniformity indexes:

$$
{}_{j,h}U_i = \begin{cases} \frac{{}_j\overset{-}{M}_i}{{}_h\overset{+}{M}_i} = \frac{{}_jQ_i}{{}_jN_i}\frac{{}_hR_i}{{}_hW_i} & \text{if}^4\, {}_jN_i > 0 \text{ and } {}_hR_i > 0; \\ 0 & \text{otherwise} \end{cases} \qquad j,h = 1,\ldots,c;\; i = 1,\ldots,r. \tag{5.1}
$$

For $j = h$, (5.1) is a comparison within the same subgroup. For $j \neq h$, (5.1) is a comparison between different subgroups (j and h) and should be interpreted as a point cross uniformity index. Clearly, for a fixed i and j, we have one within index and $(c-1)$ cross indexes.

This separation will allow us to split the overall uniformity into two components concerning, respectively, a within and a between source.

The overall uniformity point index in correspondence with the value x_i is given by:

$$
U_i = \frac{\overset{-}{M}_i}{\overset{+}{M}_i} \tag{5.2}
$$

and the overall lower mean $\overset{-}{M}_i$ can be obtained as the weighted average of the lower means ${}_j\overset{-}{M}_i \quad (j = 1,\ldots,c)$ of the c subgroups with weights given by the correspondent cumulative frequencies ${}_jN_i$ (the numerousness of the correspondent groups):

$$
\overset{-}{M}_i = \frac{Q_i}{N_i} = \frac{1}{N_i}\sum_{j=1}^{c} {}_j\overset{-}{M}_i\; {}_jN_i. \tag{5.3}
$$

[4]The comparison is possibile when either the lower group in the j-th subgroup and the upper group in the h-th subgroup are non-empty.

Substituting (5.3) in (5.2) and remembering that $\overset{+}{M}_i = W_i/R_i$ we get[5]:

$$U_i = \sum_j {}_j\overset{-}{M}_i \; \frac{{}_jN_i}{N_i}\frac{R_i}{W_i}. \tag{5.4}$$

The weight assigned to the j-th subgroup lower mean ${}_j\overset{-}{M}_i$ in (5.4) can be rewritten as follows:

$$\begin{aligned}\frac{{}_jN_i}{N_i}\frac{R_i}{W_i} &= \frac{{}_jN_i}{N_i}\frac{1}{W_i}\left(\sum_h {}_hR_i\right) = \frac{{}_jN_i}{N_i}\frac{1}{W_i}\left(\sum_h \frac{{}_hR_i}{{}_hW_i}{}_hW_i\right)\\ &= \frac{{}_jN_i}{N_i}\left[\frac{1}{W_i}\left(\sum_h \frac{1}{{}_h\overset{+}{M}_i}{}_hW_i\right)\right]\end{aligned} \tag{5.5}$$

where we observe that the quantity within square brackets is the reciprocal of the weighted harmonic mean of the upper means ${}_j\overset{+}{M}_i$ with weights ${}_jW_i$.
Substituting (5.5) into (5.4) we obtain:

$$\begin{aligned}U_i &= \sum_j {}_j\overset{-}{M}_i \; \frac{{}_jN_i}{N_i}\frac{1}{W_i}\left(\sum_h \frac{1}{{}_h\overset{+}{M}_i}{}_hW_i\right)\\ &= \sum_j {}_j\overset{-}{M}_i \; \frac{{}_jN_i}{N_i}\frac{1}{W_i}\left(\frac{{}_1W_i}{{}_1\overset{+}{M}_i} + \ldots + \frac{{}_cW_i}{{}_c\overset{+}{M}_i}\right)\\ &= \sum_j\sum_h \left(\frac{{}_j\overset{-}{M}_i}{{}_h\overset{+}{M}_i}\right) \; \frac{{}_jN_i}{N_i}\frac{{}_hW_i}{W_i}.\end{aligned} \tag{5.6}$$

The ratio within curve brackets in (5.6) is the point uniformity index (5.1), consequently:

$$U_i = \sum_j\sum_h {}_{j,h}U_i \; \frac{{}_jN_i}{N_i}\frac{{}_hW_i}{W_i}. \tag{5.7}$$

[5]Equation (5.4) can be also obtained from (4.16):

$$\begin{aligned}U_i &= \sum_j {}_jU_i \; \frac{{}_jW_i}{W_i} \; \frac{{}_jN_i}{N_i} \; \frac{R_i}{{}_jR_i} = \sum_j \frac{{}_j\overset{-}{M}_i}{{}_j\overset{+}{M}_i} \; \frac{{}_jW_i}{W_i} \; \frac{{}_jN_i}{N_i} \; \frac{R_i}{{}_jR_i}\\ &= \sum_j \frac{{}_j\overset{-}{M}_i}{{}_jW_i/{}_jR_i} \; \frac{{}_jW_i}{W_i} \; \frac{{}_jN_i}{N_i} \; \frac{R_i}{{}_jR_i} = \sum_j {}_j\overset{-}{M}_i \; \frac{{}_jN_i}{N_i}\frac{R_i}{W_i}.\end{aligned}$$

Equation (5.7) points out that the overall uniformity U_i, in correspondence with the value x_i, can be expressed as a weighted average of the uniformities ${}_{j,h}U_i$ since it is easy to verify that:

$$\sum_j \sum_h \frac{{}_jN_i}{N_i} \frac{{}_hW_i}{W_i} = \sum_j \frac{{}_jN_i}{N_i} \sum_h \frac{{}_hW_i}{W_i} = 1.$$

To our purpose, we now split the second summation in (5.7) by considering on one hand the value of the index $h = j$ and on the other the $(c-1)$ indexes $h \neq j$:

$$U_i = \sum_j {}_{j,j}U_i \frac{{}_jN_i}{N_i} \frac{{}_jW_i}{W_i} + \sum_j \sum_{h \neq j} {}_{j,h}U_i \frac{{}_jN_i}{N_i} \frac{{}_hW_i}{W_i}. \tag{5.8}$$

The first summation in (5.8) involves all the point uniformity indexes obtained within the same subgroup while the second one regards the cross ratios ${}_{j,h}U_i$ $(h \neq j)$, that is the uniformity indexes resulting when the comparison takes place between the lower mean of the j-th subgroup and the correspondent upper mean of another subgroup. Consequently the first summation can be interpreted as a measure of the within subgroups component of the overall uniformity index U_i while the second one measures the between subgroups contribution.
The decomposition obtained for each of the indexes U_i can be applied to obtain an analogous decomposition of the overall synthetic uniformity measure (4.13), in fact:

$$U = \sum_{i=1}^{r} U_i \frac{n_{i\cdot}}{N} = \sum_i \sum_j \sum_h {}_{j,h}U_i \frac{{}_jN_i}{N_i} \frac{{}_hW_i}{W_i} \frac{n_{i\cdot}}{N} \tag{5.9}$$

$$= \sum_i \sum_j {}_{j,j}U_i \frac{{}_jN_i}{N_i} \frac{{}_jW_i}{W_i} \frac{n_{i\cdot}}{N} + \sum_i \sum_j \sum_{h \neq j} {}_{j,h}U_i \frac{{}_jN_i}{N_i} \frac{{}_hW_i}{W_i} \frac{n_{i\cdot}}{N} \tag{5.10}$$

highlighting the within and the between contributions to the overall uniformity.

Example 5.1

Consider the frequency distribution of three groups ($j = 1, 2, 3$) reported in table 5.1(a).
Tables 5.1(b)-(e) report all the quantities necessary to calculate the point uniformity indexes ${}_{j,h}U_i$ as defined in (5.1): results are reported in table 5.2.

(a) n_{ij}

x_i	subgroup 1	2	3	Tot
1	1	3	2	6
2	2	6	1	9
5	3	8	6	17
6	0	4	4	8
8	4	5	3	12
11	2	3	3	8
12	2	7	5	14
13	3	2	4	9
18	1	9	1	11
20	2	3	1	6
Tot	20	50	30	100

(b) ${}_jN_i$

x_i	subgroup 1	2	3	Tot
1	1	3	2	6
2	3	9	3	15
5	6	17	9	32
6	6	21	13	40
8	10	26	16	52
11	12	29	19	60
12	14	36	24	74
13	17	38	28	83
18	18	47	29	94
20	20	50	30	100

(c) ${}_jR_i$

x_i	subgroup 1	2	3	Tot
1	20	50	30	100
2	19	47	28	94
5	17	41	27	85
6	14	33	21	68
8	14	29	17	60
11	10	24	14	48
12	8	21	11	40
13	6	14	6	26
18	3	12	2	17
20	2	3	1	6

(d) ${}_jQ_i$

x_i	subgroup 1	2	3	Tot
1	1	3	2	6
2	5	15	4	24
5	20	55	34	109
6	20	79	58	157
8	52	119	82	253
11	74	152	115	341
12	98	236	175	509
13	137	262	227	626
18	155	424	245	824
20	195	484	265	944

(e) ${}_jW_i$

x_i	subgroup 1	2	3	Tot
1	195	484	265	944
2	194	481	263	938
5	190	469	261	920
6	175	429	231	835
8	175	405	207	787
11	143	365	183	691
12	121	332	150	603
13	97	248	90	435
18	58	222	38	318
20	40	60	20	120

Table 5.1: Frequency distribution n_{ij}, cumulative frequencies ${}_jN_i$, retro cumulative frequencies ${}_jR_i$, cumulative sums of units ${}_jQ_i$ and retro cumulative sums of units ${}_jW_i$.

	Group 1			Group 2			Group 3			
x_i	1	2	3	1	2	3	1	2	3	U_i
1	0.10256	0.10331	0.11321	0.10256	0.10331	0.11321	0.10256	0.10331	0.11321	0.10593
2	0.16323	0.16286	0.17744	0.16323	0.16286	0.17744	0.13058	0.13028	0.14195	0.16034
5	0.29825	0.29140	0.34483	0.28947	0.28283	0.33469	0.33801	0.33025	0.39080	0.31471
6	0.26667	0.25641	0.30303	0.30095	0.28938	0.34199	0.35692	0.34320	0.40559	0.31964
8	0.41600	0.37235	0.42705	0.36615	0.32773	0.37588	0.41000	0.36698	0.42089	0.37093
11	0.43124	0.40548	0.47177	0.36653	0.34464	0.40098	0.42326	0.39798	0.46304	0.39479
12	0.46281	0.44277	0.51333	0.43343	0.41466	0.48074	0.48209	0.46122	0.53472	0.45628
13	0.49848	0.45493	0.53725	0.42648	0.38922	0.45965	0.50147	0.45766	0.54048	0.45080
18	0.44540	0.46547	0.45322	0.46662	0.48764	0.47480	0.43698	0.45666	0.44465	0.46862
20	0.48750	0.48750	0.48750	0.48400	0.48400	0.48400	0.44167	0.44167	0.44167	0.47200

Table 5.2: Point uniformity indexes ${}_{j,h}U_i$ $(j, h = 1, 2, 3;\ i = 1, \ldots, 10)$ and, in the last column, the overall point uniformity indexes U_i.

Consider for example the value ${}_{2,2}U_3 = 0.28283$*: it derives from a comparison within the same subgroup (the second one) and can be interpreted by saying that the mean of the values of the second group no greater than* $x_3 = 5$ *represents the* 28.283% *of the mean of the values (of the second group) no lower than* 5*. Whereas, the index* ${}_{2,3}U_3 = 0.33469$ *has been obtained by comparing the lower group* $(x \leqslant 5)$ *of the second subgroup with the upper group* $(x \geqslant 5)$ *of the third subgroup; it means that the correspondent lower mean of the subgroup 2 represents the* 33.469% *of the correspondent upper mean in subgroup 3.*
The values in the last column of table 5.2 are the overall uniformity point indexes (5.2); for example:

$$U_3 = \frac{\overset{-}{M}_3}{\overset{+}{M}_3} = \frac{Q_3}{N_3}\frac{R_3}{W_3} = \frac{109}{32}\frac{85}{920} = 0.31471$$

denotes that the overall (lower) mean of the values $x \leqslant 5$ *represents the* 31.471% *of the overall (upper) mean of the values* $x \geqslant 5$*.*
According to (5.7), each index U_i *can be obtained as the weighted average of the*

correspondent ${}_{j,h}U_i$ $(j,h=1,2,3)$ *that is the values reported in the correspondent row of table 5.2; for example:*

$$\begin{aligned}U_3 =&{}_{1,1}U_3\frac{{}_1N_3}{N_3}\frac{{}_1W_3}{W_3}+{}_{1,2}U_3\frac{{}_1N_3}{N_3}\frac{{}_2W_3}{W_3}+\ldots+{}_{3,2}U_3\frac{{}_3N_3}{N_3}\frac{{}_2W_3}{W_3}+{}_{3,3}U_3\frac{{}_3N_3}{N_3}\frac{{}_3W_3}{W_3}\\ =&0.29825\frac{6}{32}\frac{190}{920}+0.29140\frac{6}{32}\frac{469}{920}+\ldots+0.33025\frac{9}{32}\frac{469}{920}+0.39080\frac{9}{32}\frac{261}{920}\\ =&0.31471.\end{aligned}$$

The index U_i *can be further decomposed, according to (5.8), into a within and a between subgroups source: results are reported in table 5.3. In the same table, for each* U_i*, the percentage of the within and between components are reported.*

	within		between		
x_i	$\sum_j {}_{j,j}U_i \frac{{}_jN_i}{N_i}\frac{{}_jW_i}{W_i}$	%	$\sum_j\sum_{h\neq j} {}_{j,h}U_i \frac{{}_jN_i}{N_i}\frac{{}_hW_i}{W_i}$	%	U_i
1	0.04061	38.333	0.06532	61.667	0.10593
2	0.06482	40.426	0.09552	59.574	0.16034
5	0.11933	37.917	0.19538	62.083	0.31471
6	0.12290	38.451	0.19674	61.549	0.31964
8	0.13618	36.713	0.23475	63.287	0.37093
11	0.14467	36.645	0.25012	63.355	0.39479
12	0.17178	37.647	0.28450	62.353	0.45628
13	0.16208	35.955	0.28871	64.045	0.45080
18	0.20216	43.140	0.26646	56.860	0.46862
20	0.17558	37.200	0.29642	62.800	0.47200
Tot	0.13771	38.225	0.22256	61.775	0.36027

Table 5.3: Decomposition of the uniformity indexes U_i according to equation (5.8).

The last row of table 5.3 reports the decomposition of the overall uniformity index $U=0.36027$ *according to equation (5.10): it means that, following the approach here proposed, the overall uniformity is due for the* 38.225% *to the uniformity within subgroups and for the* 61.775% *to the uniformity between subgroups.*

5.1 Another representation of the uniformity decomposition

In this section we propose another expression for the decomposition introduced in the previous section.

Equation (5.8) can also be written as:

$$\begin{aligned} U_i =& \frac{\sum_j {}_{j,j}U_i \, \frac{{}_jN_i}{N_i}\frac{{}_jW_i}{W_i}}{\sum_j \frac{{}_jN_i}{N_i}\frac{{}_jW_i}{W_i}} \sum_j \frac{{}_jN_i}{N_i}\frac{{}_jW_i}{W_i} + \frac{\sum_j \sum_{h\neq j} {}_{j,h}U_i \, \frac{{}_jN_i}{N_i}\frac{{}_hW_i}{W_i}}{\sum_j \sum_{h\neq j} \frac{{}_jN_i}{N_i}\frac{{}_hW_i}{W_i}} \sum_j \sum_{h\neq j} \frac{{}_jN_i}{N_i}\frac{{}_hW_i}{W_i} \\ =& {}_WU_i \; A_i + {}_BU_i \; (1 - A_i) \end{aligned} \tag{5.11}$$

where:

$$A_i = \sum_j \frac{{}_jN_i}{N_i}\frac{{}_jW_i}{W_i} \tag{5.12}$$

denotes the sum of the weights ascribable to the (within) point uniformity indexes ${}_{j,j}U_i$ while $1 - A_i$ is the whole weight assigned to the uniformity cross indexes ${}_{j,h}U_i$ $(h \neq j)$ evaluated between two different subgroups.

The appeal of equation (5.11) lies in the representation of the point uniformity index U_i as the weighted average of two quantities related, respectively, to the within and between contribution.

An analogous expression holds for the overall uniformity index:

$$\begin{aligned} U =& \sum_i U_i \; \frac{n_{i.}}{N} \\ =& \sum_i \left[{}_WU_i \; A_i + {}_BU_i \; (1 - A_i) \right] \; \frac{n_{i.}}{N} \\ =& \frac{\sum_i {}_WU_i \; A_i \frac{n_{i.}}{N}}{\sum_i A_i \frac{n_{i.}}{N}} \sum_i A_i \frac{n_{i.}}{N} + \frac{\sum_i {}_BU_i \; (1 - A_i) \frac{n_{i.}}{N}}{\sum_i (1 - A_i) \frac{n_{i.}}{N}} \sum_i (1 - A_i) \frac{n_{i.}}{N} \\ =& {}_WU \; A + {}_BU \; (1 - A) \end{aligned} \tag{5.13}$$

where $A = \sum_i A_i \frac{n_{i.}}{N}$ denotes the average of the A_i (equation 5.12) weighted by the relative frequency of the i-th value and can be thus interpreted as the overall weight referable to the uniformity indexes evaluated within the same subgroup.

Conversely $1 - A$ can be perceived as the overall weight ascribable to the uniformity indexes evaluated between different subgroups.

Example 5.2

Table 5.4 reports the representation of each U_i*, evaluated on the frequency distribution of the three groups of example 5.1, following equation (5.11). The last row of the table provides the analogous representation for the overall uniformity* U *as in equation (5.13).*

	within		between		
x_i	${}_WU_i$	A_i	${}_BU_i$	$1 - A_i$	U_i
1	0.10565	0.38436	0.10611	0.61564	0.10593
2	0.16000	0.40512	0.16057	0.59488	0.16034
5	0.30649	0.38933	0.31995	0.61067	0.31471
6	0.31427	0.39108	0.32309	0.60892	0.31964
8	0.35743	0.38100	0.37924	0.61900	0.37093
11	0.38015	0.38056	0.40378	0.61944	0.39479
12	0.44445	0.38649	0.46373	0.61351	0.45628
13	0.43052	0.37649	0.46304	0.62351	0.45080
18	0.48037	0.42085	0.46009	0.57915	0.46862
20	0.47886	0.36667	0.46803	0.63333	0.47200
	${}_WU$	A	${}_BU$	$1 - A$	U
	0.35361	0.38945	0.36452	0.61055	0.36027

Table 5.4: Decomposition of the uniformity indexes U_i according to equation (5.11).

It is interesting to consider the situation in which the c subgroups are all equal[6] i.e. we have the replication of the same distribution c times.
In this case the evaluation of an uniformity index within the same subgroup or between different subgroups leads to the same result:

$$ {}_{j,j}U_i = {}_{j,h}U_i = {}_{h,j}U_i = {}_{h,h}U_i = U_i \qquad j, h = 1\ldots, c;\ i = 1, \ldots, r. \tag{5.14}$$

Moreover:

$$ \frac{{}_jN_i}{N_i} = \frac{{}_jW_i}{W_i} = \frac{1}{c} \quad j = 1\ldots, c;\ i = 1, \ldots, r \tag{5.15}$$

consequently the within and the between terms in decomposition (5.8) are, respectively:

$$\sum_j {}_{j,j}U_i\ \frac{{}_jN_i}{N_i}\frac{{}_jW_i}{W_i} = \frac{1}{c}\ U_i$$
$$\sum_j \sum_{h\neq j} {}_{j,h}U_i\ \frac{{}_jN_i}{N_i}\frac{{}_hW_i}{W_i} = \frac{c-1}{c}\ U_i$$

such that the uniformity index U_i is split into two parts according to the following arguments: the possibile c^2 choices of two subgroups from the c given can be separated into the c in which we consider two times the same subgroup and the $c(c-1)$ couples formed by different subgroups. From equation (5.14) each of the couples leads to the same uniformity index thus, a share $c/c^2 = 1/c$ of the overall uniformity index U_i is ascribable to the within component and a share $c(c-1)/c^2 = (c-1)/c$ is ascribable to the between component. The same discussion holds for the decomposition of the overall uniformity index reported in equation (5.10).
Now from equation (5.15) we get:

$$A_i = \sum_j \frac{{}_jN_i}{N_i}\frac{{}_jW_i}{W_i} = \frac{1}{c}$$

[6]In the table 4.1 of the whole distribution of the c subgroups we have:

$$n_{ij} = n_{ih} \text{ and } n_{i\cdot} = c\ n_{ij} \quad j, h = 1\ldots, c;\ i = 1, \ldots, r.$$

so in equation (5.11) we have:

$$ {}_W U_i = \frac{\frac{1}{c}\,U_i}{\frac{1}{c}} = U_i \qquad {}_B U_i = \frac{\frac{c-1}{c}\,U_i}{1-\frac{1}{c}} = U_i. $$

and in equations (5.11) and (5.13):

$$ {}_W U_i = {}_B U_i = U_i \qquad\qquad {}_W U = {}_B U = \sum_i U_i \frac{n_{i.}}{N} = U. $$

In conclusion we observe that in the case of perfect equality of the c subgroups, decompositions (5.8) and (5.10) depend on c and in particular the shares of the within and between components are, respectively, $1/c$ and $(c-1)/c$ and they tend, respectively, to 0 and to 1 as the number of the subgroups tends to infinity.
On the other hand the terms ${}_W U_i = {}_B U_i$ and ${}_W U = {}_B U$ in equations (5.11) and (5.13) do not depend on the number c of the subgroups considered; in these expressions the effect of c is switched on the weights A_i and A.

6 The decomposition of the inequality index

The decomposition (5.11) derived in the previous section refers to the point uniformity index U_i and can be applied to obtain an analogous decomposition of the point inequality index I_i:

$$ \begin{aligned} I_i =& 1 - U_i = 1 - [{}_W U_i \; A_i + {}_B U_i \; (1 - A_i)] \\ =& (1 - {}_W U_i) \; A_i + (1 - {}_B U_i) \; (1 - A_i) \\ =& {}_W I_i \; A_i + {}_B I_i \; (1 - A_i) \,. \end{aligned} \tag{6.1} $$

On the basis of (6.1) the point inequality index I_i is expressed as weighted average of the components ${}_W I_i = 1 - {}_W U_i$ and ${}_B I_i = 1 - {}_B U_i$ related, respectively, to the within and between subgroups contributions to the inequality.
The overall inequality index $I = 1 - U$ can be decomposed in the same way by

considering (5.13):

$$\begin{aligned} I =& 1 - U = 1 - [{}_W U \; A + {}_B U \; (1 - A)] \\ =& (1 - {}_W U) \; A + (1 - {}_B U) \; (1 - A) \\ =& {}_W I \; A + {}_B I \; (1 - A) \,. \end{aligned} \tag{6.2}$$

Equation (6.2) represents the overall inequality index I as a weighted average of a within and a between term.

If we are interested in obtaining an expression that involves the sum of the within and between terms, instead of their weighted average, a third term comes up and we have the following decompositions for the point inequality indexes:

$$I_i = {}_W I_i + {}_B I_i - [{}_W I_i \; (1 - A_i) + {}_B I_i \; A_i] \qquad i = 1, \ldots, r \tag{6.3}$$

and for the overall inequality index:

$$I = {}_W I + {}_B I - [{}_W I \; (1 - A) + {}_B I \; A] \,. \tag{6.4}$$

We thus observe that in order to obtain the overall inequality index it is necessary to subtract a non negative term from the sum of the within and between sources: this term is again a weighted average of ${}_W I$ and ${}_B I$ but the weights are interchanged with respect to the decomposition (6.2). The presence of the third term in the decomposition (6.4) is necessary because the sum of the first two should exceed the value of the overall inequality index I and its maximum value $1-1/N^2$. For example: in the particular situation, considered in section 5.1, in which the c subgroups are all equal we have:

$$I = (1 - U) < {}_W I + {}_B I = (1 - U) + (1 - U) = 2 \; (1 - U) \,.$$

7 Conclusions and further developments

In this paper we propose a decomposition for the inequality index based on the ratios between lower and upper arithmetic means proposed by Zenga [2006] in a subgroups framework.

The decomposition scheme adopted, at first for the overall uniformity, is different from the usual one which expects the within component to be a function of the uniformities (inequalities) evaluated in the subgroups and the between component to arise as a comparison among some convenient aggregate characteristics of the subgroups such as the arithmetic means. In particular we defined the point uniformity indexes in a way that allows a point comparison either within or between subgroups. The within and between components are subsequently derived as a weighted average of these ratios. The choice of this way of proceeding should be justified by thinking that, in a sense, the kind of inequality index adopted suggests a subgroups decomposition approach and, in our opinion, our proposal seems to be easy and convenient for the particular inequality index considered. In fact the decomposition mainly involves the point uniformity indexes which are crucial in the definition of the index itself.

Moreover the possibility to express the decomposition as a weighted average of two terms related, respectively, to the within and to the between contribution should represent an advantage.

A deeper analysis of the decomposition is needed by means of applications and comparison with other inequalities index decomposition proposals. For the latter we are particularly interested in the decomposition of the Gini concentration ratio proposed by Dagum [1997] which appears to follow an approach close to the one here proposed and in the three terms decomposition of the normalized family of indexes $H^{(\alpha)}$ proposed by Zenga [1986].

References

Dagum, C. (1997). Decomposition and Interpretation of Gini and the Generalized Entropy Inequality Measures. In *ASA Proceedings of the Business and Economic Statistics Section*, pages 200–205. American Statistical Association.

Dalton, H. (1920). The Measurement of the Inequality of Incomes. *The Economic Journal*, 30:348–361.

Deutsch, J. and Silber, J. (1999). Inequality Decomposition by Population Subgroups and the Analysis of Interdistributional Inequality. In Silber, J. and Sen, A., editors, *Handbook of Income Inequality Measurement*, pages 363–403. Kluwer Academic Press.

Shorrocks, A. (1980). The Class of Additively Decomposable Inequality Measures. *Econometrica*, 48(3):613–626.

Tarsitano, A. (1990). Concentration Curves and Concentration Indexes Derived from them. In Dagum, C. and Zenga, M., editors, *The Bonferoni Index of Income Inequality*. Springer-Verlag. Proceedings of the Second International Conference on Income Distribution by Size: Generation, Distribution, Measurement and Applications, held at the University of Pavia, Italy, September 28-30, 1989.

Zenga, M. (1986). On the Normalization, the Invariance to the Population Replication and the Additive Decomposition in the $\left\{I^{(\alpha)}\right\}$ Class of Inequality Measures. *Metron*, 44:131–152.

Zenga, M. (2006). Inequality Curve and Inequality Index Based on the Ratios Between Lower and Upper Arithmetic Means. Rapporto di Ricerca 115, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali - Università degli Studi di Milano-Bicocca.