Statistica & Applicazioni - Vol. XIX, n. 1, 2021, pp. 7-34
Published online november 2021

DECOMPOSITION BY SOURCES, BY SUBPOPULATIONS AND JOINT DECOMPOSITION BY SUBPOPULATIONS AND SOURCES OF GINI, BONFERRONI AND ZENGA 2007 INEQUALITY INDEXES

Igor Valli*
Michele Zenga**

SUMMARY

Recently, the authors have illustrated the decompositions by subpopulations of the Gini (1914), Bonferroni (1930) and Zenga (2007) inequality measures. These decompositions were illustrated by a numerical example involving non-overlapping subpopulations and by a numerical example involving overlapping subpopulations. In the present paper we illustrate the decomposition by sources, the decomposition by subpopulations and the joint decomposition by subpopulations and sources of the three cited indexes. These decompositions are applied to data from the 2014 central bank of Italy sample survey on household income and wealth and are performed using the R package ineqJD.

Keywords: Gini Index, Bonferroni Index, Zenga Index, Decompositions of Inequality Indexes, IneqJD.
DOI: 10.26350/999999_000043
ISSN: 18246672 (print) 2283-6659 (digital)

1. INTRODUCTION

The Gini $G(Y)$, Bonferroni $V(Y)$ and Zenga (2007) $I(Y)$ synthetic inequality measures are the weighted means of the corresponding point measures $G_h(Y)$, $V_h(Y)$ and $I_h(Y)$:

$$G(Y) = \sum_{h=1}^{r} G_h(Y) \cdot \frac{n_{h.}}{N}, \quad V(Y) = \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{N}, \quad I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N};$$

$\left\{(y_h, n_h) : h = 1, \ldots, r; \sum_{h=1}^{r} n_{h.} = N\right\}$ is the frequency distribution of the whole population. $(X_1, \ldots, X_j, \ldots, X_c)$ are the c variates (income sources) observable on each of the N units of the population and $Y = \sum_{j=1}^{c} X_j$. The whole population is split in k distinct subpopulations: $n_{.g} = \sum_{h=1}^{r} n_{hg}$ is the size of subpopulation g, $g = 1, \ldots, k$, n_{hg} is the frequency of y_h in the subpopulation g; $N = \sum_{g=1}^{k} n_{.g}$;

* ✉ igor.valli@gmail.com.
** Dipartimento di Statistica e Metodi Quantitativi - Università Milano-Bicocca - Piazza dell'Ateneo Nuovo, 1 - 20126 MILANO (e-mail: michele.zenga@unimib.it).

$n_{h.} = \sum_{g=1}^{k} n_{hg}$. Recently, the decompositions of the Zenga (2007) $I(Y)$ index, by subpopulations (Radaelli, 2008, 2010; Zenga, 2016a), by sources (Zenga, Radaelli and Zenga Ma., 2012), as well as the joint decomposition by subpopulations and sources (Zenga, 2015) have been obtained by the use of the new ''two-step'' approach. In the first step these Authors obtained the decomposition of $I_h(Y)$, then, second step, by averaging these decompositions, they obtained the corresponding decomposition of $I(Y)$. In addition, Zenga (2013) has used this two-step approach for the decomposition by sources of $V(Y)$ and $G(Y)$ indexes, too. Zenga and Valli (2016), and Zenga (2016b) obtained, by the use of the ''two-step'' approach, the decompositions by subpopulations of the Bonferroni $V(Y)$ and of the Gini $G(Y)$ indexes, respectively. Moreover, Zenga and Valli (2017, 2018) have obtained, respectively, the joint decomposition by subpopulations and sources of $V(Y)$ and $G(Y)$.

In this paper we illustrate the decompositions by sources, by subpopulations and the joint decomposition by subpopulations and sources, based on the two step approach, of the three inequality measures. These decompositions are applied to data from the 2014 central bank of Italy sample survey on household income and wealth (Bank of Italy, 2016) using the R package ''`ineqJD`'' (Arcagni and Valli, 2019). The rest of the paper is organized as follows. In Section 2 we give some definitions and notation, necessary in the case of frequency distribution of the total income Y, to the calculation of the point and the synthetic measures of the three inequality indexes. In particular, at each y_h, the whole population is split into two non-overlapping groups: the lower group $\{Y \leq y_k\}$ including the first $P_{h.} = \sum_{t=1}^{h} n_{t.}$ units and the corresponding upper group $\{Y > y_k\}$ including the remaining $N - P_h$ units. $M(Y)$, $\bar{M}_{h.}(Y)$, $\overset{+}{M}_{h.}(Y)$ are: the mean of the whole population, the mean of the lower group and the mean of the upper group. $P_{hg}(Y) = P_{hg} = \sum_{t=1}^{h} n_{tg}$ is the number of units of the lower group $\{Y \leq y_k\}$ of the subpopulation g, and $(n_{.g} - P_{hg})$ are the units of the upper group of the same subpopulation g. $M_g(Y)$, $\bar{M}_{hg}(Y)$ and $\overset{+}{M}_{hg}(Y)$ are: the mean, the lower mean and the upper mean of the subpopulation g. In Section 3 we give the relations:

1 $M(Y) = \sum_{\ell=1}^{k} M_\ell(Y) \cdot \frac{n_{.\ell}}{N}$;

2 $\bar{M}_{h.}(Y) = \sum_{\ell=1}^{k} \bar{M}_{h\ell}(Y) \cdot p(\ell|h)$, where $p(\ell|h) = P_{h\ell}/P_{h.}$;

3 $\overset{+}{M}_{h.}(Y) = \sum_{g=1}^{k} \overset{+}{M}_{hg}(Y) \cdot a(g|h)$, where for $h = 1, \ldots, r-1$,
$a(g|h) = (n_{.g} - P_{hg})/(N - P_{h.})$, and for $h = r$, $a(g|h) = n_{rg}/n_{r.}$;

4 $M(Y) = \sum_{j=1}^{c} M(X_j)$, $M_\ell(Y) = \sum_{j=1}^{c} M_\ell(X_j)$;

5 $\bar{M}_{h.}(Y) = \sum_{j=1}^{c} \bar{M}_{h.}(X_j)$, $\bar{M}_{h\ell}(Y) = \sum_{j=1}^{c} \bar{M}_{h\ell}(X_j)$;

6 $\overset{+}{M}_{h.}(Y) = \sum_{j=1}^{c} \overset{+}{M}_{h.}(X_j)$, $\overset{+}{M}_{h\ell}(Y) = \sum_{j=1}^{c} \overset{+}{M}_{h\ell}(X_j)$.

From these relations we obtained the following additive decompositions for $[M(Y) - \bar{M}_{h.}(Y)]$ and for $\left[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)\right]$:

a)

$$\left[M(Y) - \bar{M}_{h.}(Y)\right] = \sum_{j=1}^{c}\left[M(X_j) - \bar{M}_{h.}(X_j)\right]$$

$$= \sum_{\ell=1}^{k}\sum_{g=1}^{k}\left[M_g(Y) - \bar{M}_{h\ell}Y)\right] \cdot \frac{n_{.g}}{N} \cdot p(\ell|h)$$

$$= \sum_{\ell=1}^{k}\sum_{g=1}^{k}\sum_{j=1}^{c}\left[M_g(X_j) - \bar{M}_{h\ell}(X_j)\right] \cdot \frac{n_{.g}}{N} \cdot p(\ell|h);$$

b)

$$\left[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)\right] = \sum_{j=1}^{c}\left[\overset{+}{M}_{h.}(X_j) - \bar{M}_{h.}(X_j)\right]$$

$$= \sum_{\ell=1}^{k}\sum_{g=1}^{k}\left[\overset{+}{M}_{h}g(Y) - \bar{M}_{h\ell}(Y)\right] \cdot a(g|h) \cdot p(\ell|h)$$

$$= \sum_{\ell=1}^{k}\sum_{g=1}^{k}\sum_{j=1}^{c}\left[\overset{+}{M}_{hg}(X_j) - \bar{M}_{h\ell}(X_j)\right] \cdot a(g|h) \cdot p(\ell|h).$$

The three decompositions of the Bonferroni $V_h(Y)$ and Gini $G_h(Y)$ point indexes are essentially based on the relation in a), while the decompositions of Zenga (2007) $I_h(Y)$ index are based on the relations in b). In Section 4 the decompositions by sources of the three point index $V_h(Y)$, $I_h(Y)$ and $G_h(Y)$, and of the corresponding synthetic measures $V(Y)$, $I(Y)$ and $G(Y)$, are obtained. In Section 5 we illustrate the joint decompositions by subpopultions and sources of the point and synthetic measures of the three indexes. In Section 6 are obtained the decompositions of the point and the synthetic indexes in the sum of the contributions of each subpopulation ℓ, $\ell = 1, \dots, k$. In addition, each of these contributions is split in a within and a between components. In Section 7, we give some information to apply these decompositions using the R package `ineqJD`. Finally, in Section 8 we give some concluding remarks.

2. DEFINITIONS AND NOTATION

Let $\{0 \leq y_1 < \dots < y_h < \dots < y_r\}$ be the set of the distinct values assumed by Y over the k subpopulations and $\left\{n_{1.}, \dots, n_{h.}, \dots, n_{r.} : \quad \sum_{h=1}^{r} n_{h.} = N\right\}$ be the corresponding frequencies. It is possible to report the whole $r \times k$ distribution as in Table 1

TABLE 1. - *Bivariate $r \times k$ frequency distribution of the whole population partitioned into k subpopulations*

	Subpopulation					
	1	...	g	...	k	Total
y_1	n_{11}	...	n_{1g}	...	n_{1k}	$n_{1.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	
y_h	n_{h1}	...	n_{hg}	...	n_{hk}	$n_{h.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
y_r	n_{r1}	...	n_{rg}	...	n_{rk}	$n_{r.}$
Total	$n_{.1}$	...	$n_{.g}$	...	$n_{.k}$	N

At each y_h the whole subpopulation can be split into two non-overlapping groups: a lower group $\{Y \leq y_h\}$ including the first $P_{h.} = \sum_{t=1}^{h} n_{t.}$ units and the corresponding upper group $\{Y > y_h\}$ including the remaining $(N - P_{h.})$ units. Let: $Q_{h.}(Y) = \sum_{t=1}^{h} y_t \cdot n_{t.}$, $T(Y) = Q_{r.}$, $\bar{M}_{h.}(Y) = Q_{h.}/P_{h.}$, $(h = 1, \dots, r)$ be the mean of the lower group (lower mean), and

$$\overset{+}{M}_{h.}(Y) = \begin{cases} (T - Q_{h,})/(N - P_{h.}) & h \neq r \\ y_{r.} & h = r \end{cases}$$

be the mean of the upper group (upper mean). Moreover, $M(Y) = Q_{r.}/N$ is the mean of Y computed on the N units of the whole population.For the Bonferroni, Zenga (2007) and Gini indexes, the point and the synthetic measures are respectively given by:

$$V_h(Y) = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)}, \text{and } V(Y) = \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{N}; \tag{1}$$

$$I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)}, \text{ and } I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N}; \tag{2}$$

$$G_h(Y) = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} \cdot 2p_{h.} - \frac{n_{h.}}{N} \cdot \left(\frac{M(Y) - y_h}{M(Y)}\right), \text{ and} \tag{3}$$

$$G(Y) = \sum_{h=1}^{r} G_h(Y) \cdot \frac{n_{h.}}{N}. \tag{4}$$

In (3) $p_{h.} = P_{h.}/N$ is the relative frequency of the lower group $\{Y \leq y_h\}$. For the distribution $\{(y_h, n_{hg}), h = 1, \dots, r, g = 1, \dots, k\}$ of the subpopulation g: $P_{hg} = \sum_{t=1}^{h} n_{tg}$ is the number of units of the lower group $\{Y \leq y_h\}$ in the subpopulation g and $(n_{.g} - P_{hg})$ is the number of units of the corresponding upper group

$\{Y > y_h\}$ of the same subpopulation g. The ratio, $p(\ell|h) = P_{h\ell}/P_{h.}$, $(h = 1, \ldots, r,\ \ell = 1, \ldots k)$ gives the relative frequency of subpopulation ℓ in the lower group. Note that: $P_{h.} = \sum_{\ell=1}^{k} P_{h\ell}$ and $\sum_{\ell=1}^{k} p(\ell|h) = 1$. Analogously

$$a(g|h) = \begin{cases} (n_{.g} - P_{hg})/(N - P_{h.}) & h \neq r \\ n_{rg}/n_{r.} & h = r \end{cases}$$

is the relative frequency of subpopulation g in the upper group and $\sum_{g=1}^{k} a(g|h) = 1$. Moreover let: $Q_{hg}(Y) = \sum_{t=1}^{h} y_t n_{tg}$; $T_g(Y) = Q_{rg}(Y)$. $M_g(Y) = T_g/n_{.g}$ is the mean of subpopulation g, $\bar{M}_{hg}(Y)$ and $\overset{+}{M}_{hg}(Y)$ are the lower and the upper means of the subpopulation g. For more details on the calculation of $\overset{+}{M}_{hg}(Y)$ and $\bar{M}_{hg}(Y)$, see Zenga and Valli (2020).

3. PRELIMINARY CALCULATIONS

The mean $M(Y)$ is related to the k means $M_g(Y)$ of the k subpopulations by the relation

$$M(Y) = \sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{N} \tag{5}$$

Analogously, the lower mean $\bar{M}_{h.}(Y)$ and the upper mean $\overset{+}{M}_{h.}(Y)$ are related to the k lower means $\bar{M}_{h\ell}(Y)$ and the k upper means $\overset{+}{M}_{hg}(Y)$, by the following relations:

$$\bar{M}_{h.}(Y) = \sum_{\ell=1}^{k} \bar{M}_{h\ell}(Y) \cdot p(\ell|h) \tag{6}$$

$$\overset{+}{M}_{h.}(Y) = \sum_{g=1}^{k} \overset{+}{M}_{hg}(Y) \cdot a(g|h) \tag{7}$$

Now, let $(X_1, \ldots, X_j, \ldots, X_c)$ be the variates (income sources) observable on each of the N units of the population: $Y = \sum_{j=1}^{c} X_j$. Hence:

$$M(Y) = \sum_{j=1}^{c} M(X_j), \quad \text{and} \quad M_g(Y) = \sum_{j=1}^{c} M_g(X_j);$$

$$\bar{M}_{h.}(Y) = \sum_{j=1}^{c} \bar{M}_{h.}(X_j), \quad \text{and} \quad \bar{M}_{h\ell}(Y) = \sum_{j=1}^{c} \bar{M}_{h\ell}(X_j);$$

$$\overset{+}{M}_{h.}(Y) = \sum_{j=1}^{c} \overset{+}{M}_{h.}(X_j), \quad \text{and} \quad \overset{+}{M}_{hg}(Y) = \sum_{j=1}^{c} \overset{+}{M}_{hg}(X_j).$$

For the decomposition by sources and the joint decomposition by sources and subpopulations, we decompose first of all the numerators of the three point inequality measures. Using in the numerator of $V_h(Y)$ the relations $M(Y) = \sum_{j=1}^{c} M(X_j)$ and $\bar{M}_{h.}(Y) = \sum_{j=1}^{c} \bar{M}_{h.}(X_j)$ gives:

$$[M(Y) - \bar{M}_{h.}(Y)] = \sum_{j=1}^{c} \big[M(X_j) - \bar{M}_{h.}(X_j)\big]. \tag{8}$$

Using the relations (5) and (6) in the numerator of $V_h(Y)$ gives:

$$\begin{aligned}
[M(Y) - \bar{M}_{h.}(Y)] &= \sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{N} - \sum_{\ell=1}^{k} \bar{M}_{h\ell}(Y) \cdot p(\ell|h) \\
&= \sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{N} \sum_{\ell=1}^{k} p(\ell|h) - \sum_{\ell=1}^{k} \bar{M}_{h\ell}(Y) \cdot p(\ell|h) \sum_{g=1}^{k} \cdot \frac{n_{.g}}{N} \\
&= \sum_{\ell=1}^{k} \sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{N} \cdot p(\ell|h) - \sum_{\ell=1}^{k} \sum_{g=1}^{k} \bar{M}_{h\ell} \cdot p(\ell|h) \cdot \frac{n_{.g}}{N} \\
&= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \big[M_g(Y) - \bar{M}_{h\ell}(Y)\big] \cdot p(\ell|h) \cdot \frac{n_{.g}}{N}.
\end{aligned}$$

Finally, using the relations $M_g(Y) = \sum_{j=1}^{c} M_g(X_j)$ and $\bar{M}_{h\ell}(Y) = \sum_{j=1}^{c} \bar{M}_{h\ell}(X_j)$ in this latter double sum gives:

$$M(Y) - \bar{M}_{h.}(Y) = \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \big[M_g(X_j) - \bar{M}_{h\ell}(X_j)\big] \cdot p(\ell|h) \cdot \frac{n_{.g}}{N}. \tag{9}$$

Now, using the relations $\overset{+}{M}_{h.}(Y) = \sum_{j=1}^{c} \overset{+}{M}_{h.}(X_j)$ and $\bar{M}_{h.}(Y) = \sum_{j=1}^{c} \bar{M}_{h.}(X_j)$ in the numerator of $I_h(Y)$ gives:

$$\Big[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)\Big] = \sum_{j=1}^{c} \Big[\overset{+}{M}_{h.}(X_j) - \bar{M}_{h.}(X_j)\Big]. \tag{10}$$

Using (7) and (6) in the numerator of $I_h(Y)$ gives:

$$\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y) = \sum_{\ell=1}^{k} \sum_{g=1}^{k} \Big[\overset{+}{M}_{hg}(Y) - \bar{M}_{h\ell}(Y)\Big] \cdot a(g|h) \cdot p(\ell|h).$$

Finally, using the relations

$$\overset{+}{M}_{hg}(Y) = \sum_{j=1}^{c} \overset{+}{M}_{hg}(X_j) \quad \text{and} \quad \mathrm{M}_{h\ell}(\mathrm{Y}) = \sum_{\mathrm{j}=1}^{\mathrm{c}} \mathrm{M}_{h\ell}(\mathrm{X}_\mathrm{j})$$

in this latter double sum gives:

$$\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y) = \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \Big[\overset{+}{M}_{hg}(X_j) - \bar{M}_{h\ell}(X_j)\Big] \cdot a(g|h) \cdot p(\ell|h). \tag{11}$$

To obtain the decompositions of the Gini point index $G_h(Y)$ we need to decompose the difference $M(Y) - y_h$. The sums of the values of X_j and of Y observable on each of the $n_{h.}$ units with $Y = y_h$, are denoted with $S_{h.}(X_j)$ and $S_{h.}(Y)$, respectively; note that $S_{h.}(Y) = y_h \cdot n_{h.}$. The corresponding means are: $M_{h.}(X_j) = S_{h.}(X_j)/n_{h.}$ and $M_{h.}(Y) = S_{h.}(Y)/n_{h.}$. Now, from the relation $Y = \sum_{j=1}^{c} X_j$ we can write $y_h = M_{h.}(Y) = \sum_{j=1}^{c} M_{h.}(X_j)$. Thus, using this last relation and the relation $M(Y) = \sum_{j=1}^{c} M(X_j)$ we obtain:

$$M(Y) - y_h = \sum_{j=1}^{c} \left[M(X_j) - M_{h.}(X_j) \right]. \tag{12}$$

The value of y_h can be written as follows:

$$y_h = \sum_{\ell=1}^{k} y_h \cdot f(\ell|h)$$

where $f(\ell|h) = n_{h\ell}/n_{h.}$ and $\sum_{\ell=1}^{k} f(\ell|h) = 1$. Now, using this latter relation and (5) in $M(Y) - y_h$, gives:

$$\begin{aligned} M(Y) - y_h &= \sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{N} - \sum_{\ell=1}^{k} y_h \cdot f(\ell|h) \\ &= \sum_{g=1}^{k} M_g(Y) \cdot \frac{n_{.g}}{N} \cdot \sum_{\ell=1}^{k} f(\ell|h) - \sum_{\ell=1}^{k} y_h \cdot f(\ell|h) \cdot \sum_{g=1}^{k} \frac{n_{.g}}{N} \\ &= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \left[M_g(Y) - y_h \right] \cdot \frac{n_{.g}}{N} \cdot f(\ell|h). \end{aligned}$$

The sum of the values of X_j and Y, observable on each of the n_{hg} units with $Y = y_h$, are denoted by $S_{hg}(X_j)$ and $S_{hg}(Y) = y_h \cdot n_{hg}$, respectively. Obviously, $S_{hg}(Y) = \sum_{j=1}^{c} S_{hg}(X_j)$. In the case $n_{hg} > 0$,

$$\frac{S_{hg}(Y)}{n_{hg}} = y_h = M_{hg}(Y) \quad \text{and} \quad \frac{S_{hg}(X_j)}{n_{hg}} = M_{hg}(X_j)$$

are the mean of Y and X_j computed on these units. Obviously

$$y_h = M_{hg}(Y) = \sum_{j=1}^{c} M_{hg}(X_j). \tag{13}$$

In the case of $n_{hg} = 0$ we set $S_{hg}(X_j) = S_{hg}(Y) = M_{hg}(X_j) = M_{hg}(Y) = 0$, and the relation $M_{hg}(Y) = \sum_{j=1}^{c} M_{hg}(X_j)$ is extendend to the case $n_{hg} = 0$. For more details on this point see Zenga M, Valli I. (2018). Finally, using (13) and the relation $M_g(Y) = \sum_{j=1}^{c} M_g(X_j)$ in the last double summation, gives:

$$[M(Y) - y_h] = \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \left[M_g(X_j) - M_{h\ell}(X_j) \right] \cdot \frac{n_{.g}}{N} \cdot f(\ell|h). \tag{14}$$

4. DECOMPOSITION BY SOURCES

In this section, we present the decomposition by sources of the three point and synthetic indexes.Dividing both sides of (8) by $M(Y)$ gives:

$$V_h(Y) = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} = \sum_{j=1}^{c} V_h(X_j) \tag{15}$$

where,

$$V_h(X_j) = \frac{M(X_j) - \bar{M}_{h.}(X_j)}{M(Y)} \tag{16}$$

is the relative difference between the mean and the lower mean of X_j w.r.t. the ‘‘overall’’ mean of Y and gives the contribution of X_j to the point index $V_h(Y)$. From (1) and the results in (15) and (16) we can decompose the synthetic Bonferroni inequality index:

$$V(Y) = \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{N} = \sum_{h=1}^{r}\sum_{j=1}^{c} V_h(X_j) \cdot \frac{n_{h.}}{N} \tag{17}$$

$$= \sum_{j=1}^{c} V(X_j) \tag{18}$$

where

$$V(X_j) = \sum_{h=1}^{r} V_h(X_j) \cdot \frac{n_{h.}}{N}$$

denotes the contribution of the source X_j to the synthetic index $V(Y)$. Note that $V(X_j)$ is the weighted mean of $V_h(X_j)$, with weights $n_{h.}/N$.

Now dividing both sides of (10) by $\overset{+}{M}_{h.}(Y)$ gives:

$$I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)} = \sum_{j=1}^{c} B_h(X_j), \tag{19}$$

where

$$B_h(X_j) = \frac{\overset{+}{M}_{h.}(X_j) - \bar{M}_{h.}(X_j)}{\overset{+}{M}_{h.}(Y)} \tag{20}$$

is the relative difference between the upper mean and the lower mean of X_j w.r.t. the ‘‘overall’’ upper mean of Y and gives the contribution of X_j to the Zenga (2007) point index $I_h(Y)$.

From (2) and the results in (19) and (20) the decomposition by sources of the Zenga synthetic index is obtained:

$$I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N} = \sum_{h=1}^{r} \sum_{j=1}^{c} B_h(X_j) \cdot \frac{n_{h.}}{N} \tag{21}$$

$$= \sum_{j=1}^{c} B(X_j), \tag{22}$$

where

$$B(X_j) = \sum_{h=1}^{r} B_h(X_j) \cdot \frac{n_{h.}}{N}$$

is the weighted mean of $B_h(X_j)$ with weights $n_{h.}/N$ and denotes the contribution of the source X_j to the synthetic index $I(Y)$.

For the decomposition of the Gini point index (first step) we use in (3) the relations (8) and (12):

$$G_h(Y) = \sum_{j=1}^{c} \frac{M(X_j) - \bar{M}_{h.}(X_j)}{M(Y)} \cdot 2p_h - \sum_{j=1}^{c} \frac{M(X_j) - M_h(X_j)}{M(Y)} \cdot \frac{n_{h.}}{N} \tag{23}$$

$$= \sum_{j=1}^{c} \left[\frac{M(X_j) - \bar{M}_{h.}(X_j)}{M(Y)} \cdot 2p_{h.} - \frac{M(X_j) - M_h(X_j)}{M(Y)} \cdot \frac{n_{h.}}{N} \right] \tag{24}$$

$$= \sum_{j=1}^{c} \left[V_h(X_j) \cdot 2p_{h}. - A_h(X_j) \cdot \frac{n_{h.}}{N} \right] \tag{25}$$

$$= \sum_{j=1}^{c} C_h(X_j), \tag{26}$$

where

$$A_h(X_j) = \frac{M(X_j) - M_h(X_j)}{M(Y)} \quad \text{and} \tag{27}$$

$$C_h(X_j) = \frac{M(X_j) - \bar{M}_{h.}(X_j)}{M(Y)} \cdot 2p_{h.} - \frac{M(X_j) - M_h(X_j)}{M(Y)} \cdot \frac{n_{h.}}{N} \tag{28}$$

$$= V_h(X_j) \cdot 2p_{h.} - A_h(X_j) \cdot \frac{n_{h.}}{N} \tag{29}$$

is the contribution of X_j to the point index $G_h(Y)$.

The decomposition by sources of the Gini synthetic index (second step) is then obtained putting (26) in (4):

$$G(Y) = \sum_{h=1}^{r} G_h(Y) \cdot \frac{n_{h.}}{N} = \sum_{h=1}^{r} \sum_{j=1}^{c} C_h(X_j) \cdot \frac{n_{h.}}{N} \tag{30}$$

$$= \sum_{j=1}^{c} C(X_j) \tag{31}$$

where,

$$C(X_j) = \sum_{h=1}^{r} C_h(X_j) \cdot \frac{n_{h.}}{N} \tag{32}$$

is the contribution of the source X_j to the synthetic index $G(Y)$.

5. JOINT DECOMPOSITION BY SUBPOPULATIONS AND SOURCES

Dividing both sides of (9) by $M(Y)$ gives the following basic $k \times k \times c$ joint decomposition by subpopulations and sources of the Bonferroni point index:

$$V_h(Y) = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} = \sum_{l=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \frac{M_g(X_j) - \bar{M}_{h\ell}(X_j)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{N} \tag{33}$$

$$= \sum_{l=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} V_{h\ell g}(X_j), \tag{34}$$

where

$$V_{h\ell g}(X_j) = \frac{M_g(X_j) - \bar{M}_{h\ell}(X_j)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{N} \tag{35}$$

is the contribution of X_j to $V_h(Y)$ that derives from the comparison between the lower mean $\bar{M}_{hl}(X_j)$ and the mean $M_g(X_j)$. In other words, $V_{h\ell g}(X_j)$ is one of the $k \times k \times c$ joint contributions to the point measure $V_h(Y)$.

Putting (33) in (1), gives the following $k \times k \times c$ joint decomposition by subpopulations and sources of the Bonferroni $V(Y)$ index.

$$V(Y) = \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{N}$$

$$= \sum_{h=1}^{r} \left[\sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \frac{M_g(X_j) - \bar{M}_{h\ell}(X_j)}{M(Y)} \cdot p(\ell|h) \cdot \frac{n_{.g}}{N} \right] \cdot \frac{n_{h.}}{N} \tag{36}$$

$$= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \left[\sum_{h=1}^{r} V_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N} \right] \tag{37}$$

$$= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} V_{.\ell g}(X_j), \tag{38}$$

where

$$V_{.\ell g}(X_j) = \sum_{h=1}^{r} V_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N} \tag{39}$$

is the weighted mean of $V_{h\ell g}(X_j)$ with weights $n_{h.}/N$.

The application of the ''two-step'' approach to the Zenga (2007) index gives the following results. For the point index $I_h(Y)$, dividing both sides of (11) by $\overset{+}{M}_{h.}(Y)$ gives (first step):

$$\begin{aligned} I_h(Y) &= \frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)} \\ &= \sum_{\ell=1}^{k}\sum_{g=1}^{k}\sum_{j=1}^{c} \frac{\overset{+}{M}_{hg}(X_j) - \bar{M}_{h\ell}(X_j)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(g|h) \qquad (40) \\ &= \sum_{\ell=1}^{k}\sum_{g=1}^{k}\sum_{j=1}^{c} B_{h\ell g}(X_j), \qquad (41) \end{aligned}$$

where,

$$B_{h\ell g}(X_j) = \frac{\overset{+}{M}_{hg}(X_j) - \bar{M}_{h\ell}(X_j)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(g|h) \qquad (42)$$

is the contribution of X_j to $I_h(Y)$ that is given by the product of the relative difference $\frac{\overset{+}{M}_{hg}(X_j) - \bar{M}_{h\ell}(X_j)}{\overset{+}{M}_{h.}(Y)}$ and $p(\ell|h) \cdot a(g|h)$.

The decomposition of the Zenga (2007) synthetic index is obtained (second step) putting (40) in (2):

$$\begin{aligned} I(Y) &= \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N} \\ &= \sum_{h=1}^{r}\left[\sum_{\ell=1}^{k}\sum_{g=1}^{k}\sum_{j=1}^{c} \frac{\overset{+}{M}_{hg}(X_j) - \bar{M}_{h\ell}(X_j)}{M(Y)} \cdot p(\ell|h) \cdot a(g|h)\right] \cdot \frac{n_{h.}}{N} \qquad (43) \\ &= \sum_{\ell=1}^{k}\sum_{g=1}^{k}\sum_{j=1}^{c}\left[\sum_{h=1}^{r} B_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N}\right] \qquad (44) \\ &= \sum_{\ell=1}^{k}\sum_{g=1}^{k}\sum_{j=1}^{c} B_{.\ell g}(X_j), \qquad (45) \end{aligned}$$

where

$$B_{.\ell g}(X_j) = \sum_{h=1}^{r} B_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N} \qquad (46)$$

is the weighted mean of $B_{h\ell g}(X_j)$ with weights $n_{h.}/N$.

Using in (3) the relations (9) and (14) gives the following $k \times k \times c$ joint decomposition of the Gini $G_h(Y)$ point index:

$$
\begin{aligned}
G_h(Y) &= \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} \cdot 2p_{h.} - \frac{n_{h.}}{N} \cdot \left(\frac{M(Y) - y_h}{M(Y)} \right) \\
&= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \left[\frac{M_g(X_j) - \bar{M}_{h\ell}(X_j)}{M(Y)} \cdot \frac{n_{.g}}{N} \cdot p(\ell|h) \cdot 2p_{h.} + \right. \\
&\qquad \left. - \frac{M_g(X_j) - M_{h\ell}(X_j)}{M(Y)} \cdot \frac{n_{.g}}{N} \cdot f(\ell|h) \cdot \frac{n_{h.}}{N} \right] \qquad (47) \\
&= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \left[V_{h\ell g}(X_j) \cdot 2p_{h.} - A_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N} \right] \qquad (48) \\
&= \sum_{j=1}^{c} \sum_{\ell=1}^{k} \sum_{g=1}^{k} C_{h\ell g}(X_j). \qquad (49)
\end{aligned}
$$

In (48),

$$
V_{hlg}(X_j) = \frac{M_g(X_j) - \bar{M}_{h\ell}(X_j)}{M(Y)} \cdot \frac{n_{.g}}{N} \cdot p(\ell|h)
$$

is the contribution of X_j to $V_h(Y)$ that derives from the comparison between the lower mean $\bar{M}_{h\ell}(X_j)$ and the mean $M_g(X_j)$,

$$
A_{h\ell g}(X_j) = \frac{M_g(X_j) - M_{h\ell}(X_j)}{M(Y)} \cdot \frac{n_{.g}}{N} \cdot f(\ell|h) \qquad (50)
$$

is the contribution of X_j to $G_h(Y)$ that derives from the comparison between $M_{h\ell}(X_j)$ and $M_g(X_j)$. Finally, in (49)

$$
\begin{aligned}
C_{h\ell g}(X_j) = & \frac{M_g(X_j) - \bar{M}_{h\ell}(X_j)}{M(Y)} \cdot \frac{n_{.g}}{N} \cdot p(\ell|h) \cdot 2p_{h.} + \qquad (51) \\
& - \frac{M_g(X_j) - M_{h\ell}(X_j)}{M(Y)} \cdot \frac{n_{.g}}{N} \cdot f(\ell|h) \cdot \frac{n_{h.}}{N}
\end{aligned}
$$

is the contribution of X_j to $G_h(Y)$ that derives from the comparisons of $M_g(X_j)$ w.r.t. $\bar{M}_{h\ell}(X_j)$ and $M_{h\ell}(X_j)$.

The joint decomposition of the Gini synthetic index is obtained putting (second step) the decomposition (49) in (4):

$$
\begin{aligned}
G(Y) &= \sum_{h=1}^{r} G_h(Y) \cdot \frac{n_{h.}}{N} \\
&= \sum_{h=1}^{r} \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} C_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N} \qquad (52) \\
&= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} \left[\sum_{h=1}^{r} C_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N} \right] \qquad (53)
\end{aligned}
$$

$$= \sum_{\ell=1}^{k} \sum_{g=1}^{k} \sum_{j=1}^{c} C_{.\ell g}(X_j), \tag{54}$$

where

$$C_{.\ell g}(X_j) = \sum_{h=1}^{r} C_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N} \tag{55}$$

is the weighted mean of $C_{h\ell g}(X_j)$ with weights $n_{h.}/N$.

6. DECOMPOSITION OF THE POINT AND THE SYNTHETIC INDEXES IN THE SUM OF k CONTRIBUTIONS, ONE FOR EACH SUBPOPULATION

As we have seen, using the ''two-step'' approach in the frequency distribution framework, we have decomposed in a similar way the three aforementioned inequality indexes. So, let $\Lambda_h(\cdot)$ and $\Lambda(\cdot)$ a generic point and the synthetic inequality measures respectively, choosen between those presented (Bonferroni, Zenga (2007) and Gini indexes). In analogy with what reported, let $\Lambda_{h\ell g}(X_j)$ the finer decomposition element obtained (e.g. $V_{h\ell g}(X_j)$ or $B_{hlg}(X_j)$ or $C_{h\ell g}(X_j)$). In these terms:

$$\Lambda(Y) = \sum_{h=1}^{r} \Lambda_h(Y) \cdot \frac{n_{h.}}{N} \tag{56}$$

$$= \sum_{h=1}^{r} \sum_{j=1}^{c} \Lambda_h(X_j) \cdot \frac{n_{h.}}{N} \tag{57}$$

$$= \sum_{h=1}^{r} \sum_{j=1}^{c} \sum_{\ell=1}^{k} \sum_{g=1}^{k} \Lambda_{h\ell g}(X_j) \cdot \frac{n_{h.}}{N}. \tag{58}$$

Since the decompositions proposed are additive, summing over a dimension provide elements of particular significance:

$$\sum_{j=1}^{c} \sum_{\ell=1}^{k} \sum_{g=1}^{k} \Lambda_{h\ell g}(X_j) = \sum_{j=1}^{c} \sum_{\ell=1}^{k} \Lambda_{h\ell.}(X_j) = \sum_{\ell=1}^{k} \Lambda_{h\ell.}(Y) = \Lambda_h(Y), \tag{59}$$

where

$$\Lambda_{h\ell.}(X_j) = \sum_{g=1}^{k} \Lambda_{h\ell g}(X_j), \tag{60}$$

$$\Lambda_{h\ell.}(Y) = \sum_{j=1}^{c} \Lambda_{h\ell.}(X_j)$$

Note that $\Lambda_{h\ell.}(Y)$ is the contribution of subpopulation ℓ to the point index $\Lambda_h(Y)$ and $\Lambda_{h\ell}(X_j)$ is the contribution of X_j to $\Lambda_{h\ell}(Y)$. In the case of Zenga (2007) index (59) gives:

$$I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)} = \sum_{\ell=1}^{k} B_{h\ell.}(Y) \tag{62}$$

where

$$B_{h\ell.}(Y) = \frac{\overset{+}{M}_{h.}(Y) - \overset{-}{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \tag{63}$$

is the contribution of subpopulation ℓ to the point index $I_h(Y)$.

Since in $\Lambda_{h\ell g}(\cdot)$ are involved comparisons between subpopulations ℓ and g (mean, upper mean and lower mean), $(\ell, g = 1, \dots, k)$, when $\ell = g$ the element $\Lambda_{h\ell\ell}(\cdot)$ is obtained. In this way, we can rewrite the contribution of the subpopulation ℓ to the point measure as:

$$\Lambda_{h\ell.}(\cdot) = \Lambda_{h\ell\ell}(\cdot) + \sum_{g \neq \ell} \Lambda_{h\ell g}(\cdot) = \Lambda_{h\ell W}(\cdot) + \Lambda_{h\ell B}(\cdot) \tag{64}$$

where $\Lambda_{h\ell W}(\cdot)$ denotes the within part and $\Lambda_{h\ell B}(\cdot)$ denotes the between part of $\Lambda_{h\ell.}(\cdot)$, respectively. Hence:

$$\Lambda(\cdot) = \sum_{\ell=1}^{k} \left\{ \sum_{h=1}^{r} \Lambda_{h\ell.}(\cdot) \cdot \frac{n_{h.}}{N} \right\} = \sum_{\ell=1}^{k} \left\{ \sum_{h=1}^{r} [\Lambda_{h\ell W}(\cdot) + \Lambda_{h\ell B}] \cdot \frac{n_{h.}}{N} \right\} \tag{65}$$

$$= \sum_{\ell=1}^{k} [\Lambda_{.\ell W} + \Lambda_{.\ell B}] = \Lambda_{..W}(\cdot) + \Lambda_{..B}(\cdot), \tag{66}$$

where

$$\Lambda_{.\ell W}(\cdot) = \sum_{h=1}^{r} \Lambda_{h\ell W}(\cdot) \cdot \frac{n_{h.}}{N}, \quad \text{and} \quad \Lambda_{.\ell B} = \sum_{h=1}^{r} \Lambda_{hlB}(\cdot) \cdot \frac{n_{h.}}{N} \tag{67}$$

denote the within and the between parts of the contribution of the subpopulation ℓ to the synthetic index respectively, and

$$\Lambda_{..W}(\cdot) = \sum_{\ell=1}^{k} \Lambda_{.\ell W}(\cdot) \quad \text{and} \quad {}_{..\mathrm{B}} = \sum_{1=1}^{\mathrm{k}} {}_{.\ell\mathrm{B}}(\cdot) \tag{68}$$

denote the within and the between parts of the synthetic index $\Lambda(\cdot)$.

7. APPLICATION

The results of the decompositions here proposed are applied to data from the 2014 Central Bank of Italy sample survey on household income and wealth (Bank of Italy, 2016), using the R package `ineqJD`. The package is written using the general results reported in formulas from (56) to (68) because in this setting it is possible to pass in as input object of different shape and, starting from the finer decomposition allowed by data, to obtain the desired elements of the decomposition by aggregation. We will perform step-by-step the decompositions showing which, and in which way, the package functions work, displaying at each step results and output. First of all we need to load the data. In this application data are derived from a survey carried out by Banca d'Italia where $N = 8156$ household are sampled reporting the net disposable income Y, the income sources X_j (X_1 =payrol income, X_2 =pensions and net transfers, X_3 =net self employment income and X_4 =property incomes), the membership area G (i.e. the subpopulations: 1=North, 2=Centre, 3=South) and the sample weights w_i.

```
> # install.packages('ineqJD')
> library('ineqJD')
> library(latex2exp) # for plot labels
> head(data)

         wi G X1     X2 X3         X4          Y
1 0.1723482 3  0 -26600  0 6600.00000 -20000.000
2 1.2084208 1  0 -16600  0 5004.99591 -11595.004
3 0.3419902 2  0  -4900  0   25.80806  -4874.192
4 1.1877978 3  0 -10000  0 8412.90403  -1587.096
5 1.2608120 3  0   -600  0    0.00000   -600.000
6 0.3657721 3  0      0  0    0.00000      0.000
> dim(data)

[1] 8156 7
```

To apply the proposed decompositions we need to move in the frequency distribution framework and compute the cumulative frequencies P_{hl}, the cumulative sum of each source $Q_{hl}(X_j)$. The function `dataProcessing()` computes these quantities and keeps all the pieces of information needed for the continuation like the distinct values of Y and the labels of subpopulations and sources. The function input are the data on which we want compute the inequality index (units argument), the group membership (groups argument) and the sample weights (weights argument). In the example at hand we have to consider not only the matrix of income sources but also the vectors of group membership and sample weights, respectively. If we had as data only the vector Y of incomes without sources, group membership and sample weights it would be enough to pass the vector Y in the units argument because without any additional parameter the arguments groups and weights are set automatically equal to 1 to mean that all the statistical units have the same weigths and belong to the same population.

```
> data_r <- dataProcessing(units=data[,c('X1','X2', 'X3', 'X4')],
+                          groups = data[,'G'],
+                          weights = data[,'wi'])
> str(data_r)

List of 3
 $ yh : num [1:7429] -20000 -11595 -4874 -1587 -600 ...
 $ Phl : num [1:7429, 1:3] 0 1.21 1.21 1.21 1.21 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ h: chr [1:7429] "1" "2" "3" "4" ...
  .. ..$ l: chr [1:3] "1" "2" "3"
 $ Qhlk: num [1:7429, 1:3, 1:4] 0 0 0 0 0 0 0 0 0 0 ...
  ..- attr(*, "dimnames")=List of 3
  .. ..$ h: chr [1:7429] "1" "2" "3" "4" ...
  .. ..$ l: chr [1:3] "1" "2" "3"
  .. ..$ k: chr [1:4] "X1" "X2" "X3" "X4"
 - attr(*, "class")= chr "dataProcessed"
```

The output of `dataProcessing()` are the vector of distinct values of the variate Y (`yh`), the matrix of cumulative frequencies (`Phl`) and the array of the cumulative sum of each source (`Qhlk`). We can now proceed to the evaluation and the decompositions of the inequality indexes. For this aim we need to call the function that takes the name from the Author who introduced the inequality measures here exposed.

```
> G_index <- gini(data_r)
> B_index <- bonferroni(data_r)
> Z_index <- zenga(data_r)
```

Since the outline of the mentioned functions is similar we display the structure of only one of these chosen randomly.

```
> str(Z_index)

List of 3
 $ index : chr "Zenga"
 $ decomposition: num [1:3, 1:3, 1:7429, 1:4] 0 0 0 0 0 ...
  ..- attr(*, "dimnames")=List of 4
  .. ..$ l1: chr [1:3] "1" "2" "3"
  .. ..$ l2: chr [1:3] "1" "2" "3"
  .. ..$ h : chr [1:7429] "1" "2" "3" "4" ...
  .. ..$ k : chr [1:4] "X1" "X2" "X3" "X4"
 $ dataProcessed:List of 3
  ..$ yh : num [1:7429] -20000 -11595 -4874 -1587 -600 ...
  ..$ Phl : num [1:7429, 1:3] 0 1.21 1.21 1.21 1.21 ...
```

```
 .. ..- attr(*, "dimnames")=List of 2
 .. .. ..$ h: chr [1:7429] "1" "2" "3" "4" ...
 .. .. ..$ l: chr [1:3] "1" "2" "3"
 ..$ Qhlk: num [1:7429, 1:3, 1:4] 0 0 0 0 0 0 0 0 0 0 ...
 .. ..- attr(*, "dimnames")=List of 3
 .. .. ..$ h: chr [1:7429] "1" "2" "3" "4" ...
 .. .. ..$ l: chr [1:3] "1" "2" "3"
 .. .. ..$ k: chr [1:4] "X1" "X2" "X3" "X4"
 ..- attr(*, "class")= chr "dataProcessed"
- attr(*, "class")= chr "decomposition"}}
```

These functions are of class `decompositions` and keep all the information and all the possible decompositions that can be computed on the data. To obtain the main decompositions useful for a standard user, as for the object of class `lm`, we run the `summary()` command.

```
> ineq_decompositions <- summary(Z_index); ineq_decompositions

Synthetic Zenga inequality index: 0.7

Joint contributions:
, , s = X1

  l2
l1       1      2      3
  1 0.0600 0.0317 0.0843
  2 0.0188 0.0102 0.0277
  3 0.0204 0.0110 0.0306

, , s = X2

  l2
l1       1      2      3
  1 0.0228 0.0107 0.0279
  2 0.0144 0.0067 0.0171
  3 0.0084 0.0041 0.0107

, , s = X3

  l2
l1       1      2      3
  1 0.0323 0.0150 0.0352
  2 0.0088 0.0042 0.0101
  3 0.0126 0.0059 0.0136

, , s = X4
```

```
  l2
l1       1       2       3
  1 0.0342 0.0141 0.0403
  2 0.0139 0.0058 0.0168
  3 0.0075 0.0028 0.0095

### GROUPS ###
  l2
l1       1       2       3
  1 0.1493 0.0715 0.1876
  2 0.0558 0.0269 0.0717
  3 0.0490 0.0238 0.0644

Within:
     1       2       3
0.1493 0.0269 0.0644

Between:
     1       2       3
0.1048 0.0953 0.2593

Marginal:
     1       2       3
0.2541 0.1222 0.3237

### GROUPS AND SOURCES ###
  s
l        X1     X2      X3      X4
  1 0.0992 0.0457 0.0537 0.0556
  2 0.0528 0.0215 0.0252 0.0227
  3 0.1425 0.0557 0.0588 0.0666

### SOURCES ###
    X1     X2      X3     X4
0.2945 0.1229 0.1377 0.1449
```

The output shows:

- the synthetic index
- the contribution of each subpopulation and of each source to the synthetic index (the `Joint contributions` section)
- the subpopulations contributions $Z_{.lg}(Y)$, the whithin $Z_{.lW}(Y)$ and the between components $Z_{.lB}(Y)$ and in the `Marginal` subsection the subpopulation contribution $Z_{.l.}(Y)$ (the `GROUPS` section) of the synthetic index

- the subpopulation contributions to the synthetic index derived from each source $Z_{.l.}(X_j)$ (the `GROUPS AND SOURCES` section)
- the contribution of each source to the synthetic index

Once obtained the decompositions generated from the function `gini()`, `bonferroni()` or `zenga()` we can display the inequality curve[1] We first call the function `inequallityCurves()` to generate step-functions representing the point measures (Figure 1) or the subpopulations (Figure 2) or sources (Figure 3) point contributions and then plot it. To display the inequality curve (the point measures) referred to the whole population, as in Figure 1, the commands are the following[2] Note that in the example at hand, the notation $I_{(p_{h\cdot})}(Y)$, $B_{(p_{h\cdot})\ell.}(Y)$ and $B_{(p_{h\cdot})}(X_j)$ are used instead of $I_h(Y)$, $B_{h\ell.}(Y)$ and $B_h(X_j)$, respectively, because these point measures are related to the population proportion $p_{h.} = P_{h.}/N$.

```
> ineq_curves <- inequalityCurves(Z_index)
> plot(ineq_curves, ylim=c(0,1), main='',
       ylab=TeX('I_{(p_{h.})}(Y)'), xlab=TeX('p_{h.}'))
```

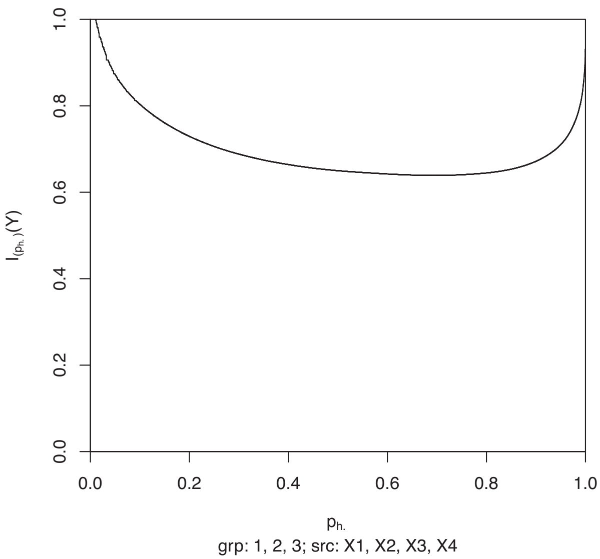

FIGURE 1. - *Zenga inequality curve $I_{(p_{h\cdot})}(Y)$. The value of the synthetic index $I(Y)$ is the area below the curve*

[1] Note that by default the plot reports the labels of subpopulations (`grp`) and sources (`src`). In the example at hand, the subpopulations labels are 1 (North), 2 (Center) and 3 (South and Island), while the sources (`src`) labels are X_1, X_2, X_3, X_4.

[2] The labels in the plots of this paper are generated using the R package `latex2exp` (Meschiari, 2021).

If the overall inequality can be decomposed, as in the application at hand, we could be interested to plot the inequality curve derived from the subpopulation $l = 1$ or from the subpopulations $l = 1$ and $l = 2$ and add it to the overall inequality curve (Figure 2). As before we need to generate the step-functions related to the interested subpopulations and then plot it again. Note that if all the subpopulations contributions are considered, we re-obtain the inequality curve of the whole population.

```
> plot(ineq_curves, ylim=c(0,1), main='', xlab=TeX('p_{h.}'),
ylab='')
> contrib1 <- inequalityCurves(Z_index, l = 1)
> contrib12 <- inequalityCurves(Z_index, l = 1:2)
> plot(contrib1, add = TRUE)
> plot(contrib12, add = TRUE)
> text(0.5, c(0.27, 0.4, 0.74),
  labels = c(TeX('B_{(p_{h.})1.}(Y)'),
  TeX('B_{(p_{h.})1.}(Y)+B_{(p_{h.})2.}(Y)'),
  TeX('B_{(p_{h.})1.}(Y)+B_{(p_{h.})2.}(Y)+B_{(p_{h.})3.}(Y)=I_{(-
  p_{h.})}(Y)')))
```

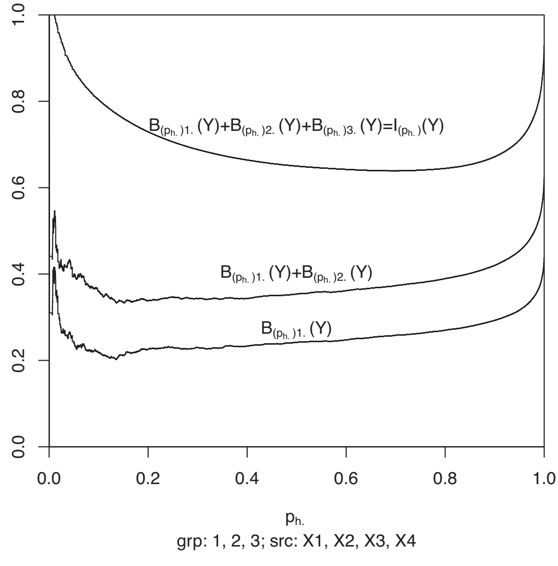

FIGURE 2. - *Cumulative subpopulations contributions $B_{(p_{h}.)\ell.}(Y)$ for the Zenga index $I_{(p_{h}.)}(Y)$. Note that $B_{(p_{h}.)1.}(Y)$ is the contribution of the "North" ($\ell = 1$), $B_{(p_{h}.)2.}(Y)$ is the contribution of the "Center" ($\ell = 2$) and $B_{(p_{h}.)3.}(Y)$ is the contribution of "South and Island" ($\ell = 3$). For more details see (62) and (63)*

Similarly, we can plot the inequality derived from two or more sources (Figure 3). In this case the scheme is similar for the previous case but we have to remember to change the input parameter `l` (that denote the subpopulation) in `k` (that denote the source). Note that if all the sources contributions are considered, we re-obtain the inequality curve of Y.

```
>  plot(ineq_curves,  ylim=c(0,1),  main=’’,  xlab=TeX(’p_{h.}’),
ylab=’’)
> contrib1 <- inequalityCurves(Z_index, k = 1)
> contrib12 <- inequalityCurves(Z_index, k = 1:2)
> contrib123 <- inequalityCurves(Z_index, k = 1:3)
> contrib1234 <- inequalityCurves(Z_index, k = 1:4)
> plot(contrib1, add = TRUE)
> plot(contrib12, add = TRUE)
> plot(contrib123, add = TRUE)
> plot(contrib1234, add = TRUE)
> text(0.55, c(0.33, 0.44, 0.56, 0.74),
   labels = c(TeX(’B_{(p_{h.})}(X_{1})’),
   TeX(’B_{(p_{h.})}(X_{1})+B_{(p_{h.})}(X_{2})’),
   TeX(’B_{(p_{h.})}(X_{1})+B_{(p_{h.})}(X_{2})+B_{(p_{h.})}(-
   X_{3})’),
   TeX(’B_{(p_{h.})}(X_{1})+B_{(p_{h.})}(X_{2})+B_{(p_{h.})}(-
   X_{3})+ B_{(p_{h.})}(X_{4})=I_{(p_{h.})}(Y)’)))
```

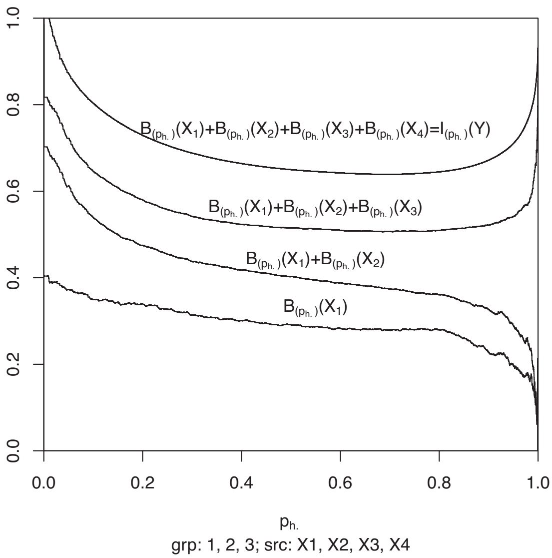

FIGURE 3. - *Cumulative sources contributions $B_{(p_{h\cdot})}(X_j)$ for the Zenga index $I_{(p_{h\cdot})}(Y)$. Note that $B_{(p_{h\cdot})}(X_1)$ is the contribution of the source X_1 (payroll income), $B_{(p_{h\cdot})}(X_2)$ is the contribution of the source X_2 (pensions and net transfers), $B_{(p_{h\cdot})}(X_3)$ is the contribution of the source X_3 (self employment income), $B_{(p_{h\cdot})}(X_4)$ is the contribution of the source X_4 (property income). For more details see (19) and (20)*

8. CONCLUDING REMARKS

Let, $\{0 \leq y_1 < \ldots < y_h < \ldots < y_r\}$ be the set of the r distinct values assumed by the variate Y, $Y \geq 0$, over k distinct subpopulations and

$$\left\{ n_{1.}, \ldots, n_{h.}, \ldots, n_{r.}; \sum_{h=1}^{r} n_{h.} = N \right\}$$

be the corresponding frequencies. Let $\{X_1, \ldots, X_j, \ldots, X_c\}$ be non-negative variates (income sources) observable on each of the N units of the population and $Y = \sum_{j=1}^{c} X_j$. Let $G(Y)$, $V(Y)$ and $I(Y)$ be the Gini (1914), Bonferroni (1930) and Zenga (2007) synthetic indexes. These three synthetic indexes are the following weighted means of the corresponding point measures $G_h(Y)$, $V_h(Y)$ and $I_h(Y)$:

$$G(Y) = \sum_{h=1}^{r} G_h(Y) \cdot \frac{n_{h.}}{N}, \quad V(Y) = \sum_{h=1}^{r} V_h(Y) \cdot \frac{n_{h.}}{N}, \quad I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N}.$$

In this paper we illustrate the decompositions, by sources, by subpopulations and the joint decomposition by subpopulations and sources of the three inequality measures based on the new convenient ''two-step'' approach. In the first step are obtained, using well known properties of the arithmetic mean, the decomposition of the point measures, then (second step), these decompositions are extended to the corresponding synthetic indexes.

For the decompositions of the three indexes plays an important role the $r \times k$ bivariate distribution (Table 1) of the N units according to the k subpopulations and to the r distinct values of Y. In particular: n_{hg} is the frequency of y_h in the subpopulation g, $g = 1, \ldots, k, n_{.g}$ is the size of subpopulation g and $N = \sum_{g=1}^{k} n_{.g}$. At each y_h the whole population is split into two non-overlapping groups: a lower group $(Y \leq y_h)$ including the first $P_{h.} = \sum_{t=1}^{h} n_{t.}$ units, and the correspondig upper group $(Y > y_h)$ including the remaining $(N - P_{h.})$ units. $M(Y)$, $\bar{M}_{h.}(Y)$ and $\overset{+}{M}_{h.}(Y)$ are the aritmethic means of: the whole population, the lower group and the upper group, respectively. The Bonferroni $V_h(Y)$, Zenga (2007) $I_h(Y)$ and Gini $G_h(Y)$ point indexes are given by:

$$V_h(Y) = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)},$$

$$I_h(Y) = \frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)}{\overset{+}{M}_{h.}(Y)},$$

$$G_h(Y) = \frac{M(Y) - \bar{M}_{h.}(Y)}{M(Y)} \cdot 2p_{h.} - \frac{n_{h.}}{N} \cdot \frac{M(Y) - y_h}{M(Y)}.$$

For the decomposition by sources of the Bonferroni and Zenga point indexes we need for each variate X_j: the mean $M(X_j)$ of the whole population, the mean $\bar{M}_{h.}(X_j)$ of the lower group $(Y \leq y_h)$ and the mean $\overset{+}{M}_{h.}(X_j)$ of the upper group. By the use of the relations: $M(Y) = \sum_{j=1}^{c} M(X_j)$, $\bar{M}_{h.}(Y) = \sum_{j=1}^{c} \bar{M}(X_j)$ and $\overset{+}{M}_{h.}(Y) = \sum_{j=1}^{c} \overset{+}{M}_{h.}(X_j)$ we obtain the following decompositions by sources of $V_h(Y)$ and $I_h(Y)$:

$$V_h(Y) = \sum_{j=1}^{c} V_h(X_j), \quad \text{where} \quad V_h(X_j) = \frac{M(X_j) - \bar{M}_{h.}(X_j)}{M(Y)};$$

$$I_h(Y) = \sum_{j=1}^{c} B_h(X_j), \quad \text{where} \quad B_h(X_j) = \frac{\overset{+}{M}_{h.}(X_j) - \bar{M}_{h.}(X_j)}{\overset{+}{M}_{h.}(Y)};$$

$V_h(X_j)$ and $B_h(X_j)$ are the contributions of X_j to the Bonferroni and Zenga (2007) point measures, respectively. For the decomposition of $G_h(Y)$ we need the decomposition of $[M(Y) - y_h]$, too. The sums of the values of X_j, $(j = 1, \ldots, c)$, and of Y observed on each of the $n_{h.}$ units with $Y = y_h$, are denoted $S_{h.}(X_j)$ and $S_{h.}(Y)$, respectively; note that $S_{h.}(Y) = y_h \cdot n_{h.}$. The corresponding means are: $M_{h.}(X_j) = S_{h.}(X_j)/n_{h.}$ and $M_{h.}(Y) = S_{h.}(Y)/n_{h.} = y_{h.}$. Obviously, $S_{h.}(Y) = \sum_{j=1}^{c} S_{h.}(X_j)$ and $y_h = \frac{1}{n_{h.}} \sum_{j=1}^{c} S_{h.}(X_j) = \sum_{j=1}^{c} M_{h.}(X_j)$.

Now, utilizing this last relation and the relation $M(Y) = \sum_{j=1}^{c} M(X_j)$ we obtain the following additive decomposition:

$$[M(Y) - y_h] = \sum_{j=1}^{c} \left[M(X_j) - M_{h.}(X_j)\right].$$

Then, putting in $G_h(Y)$ the last decomposition and the decomposition by sources of $V_h(Y)$, we obtain the decomposition by sources of the Gini (1914) point measures:

$$G_h(Y) = \sum_{j=1}^{c} C_h(X_j),$$

where,

$$C_h(X_j) = \frac{M(X_j) - \bar{M}_{h.}(X_j)}{M(Y)} \cdot 2p_{h.} - \frac{n_{h.}}{N} \cdot \frac{M(X_j) - M_{h.}(X_j)}{M(Y)}$$

is the contribution of X_j to the Gini point measure.

Finally, putting (second step) the above illustrated decompositions of $V_h(Y)$, $I_h(Y)$ and $G_h(Y)$ in the corresponding expression of the synthetic measures gives the following decomposition by sources:

$$V(Y) = \sum_{h=1}^{r} \left[\sum_{j=1}^{c} V_h(X_j)\right] \cdot \frac{n_{h.}}{N} = \sum_{j=1}^{c} \sum_{h=1}^{r} V_h(X_j) \cdot \frac{n_{h.}}{N} = \sum_{j=1}^{c} V(X_j),$$

$$I(Y) = \sum_{h=1}^{r} \left[\sum_{j=1}^{c} B_h(X_j)\right] \cdot \frac{n_{h.}}{N} = \sum_{j=1}^{c} \sum_{h=1}^{r} B_h(X_j) \cdot \frac{n_{h.}}{N} = \sum_{j=1}^{c} B(X_j),$$

$$G(Y) = \sum_{h=1}^{r} \left[\sum_{j=1}^{c} C_h(X_j)\right] \cdot \frac{n_{h.}}{N} = \sum_{j=1}^{c} \sum_{h=1}^{r} C_h(X_j) \cdot \frac{n_{h.}}{N} = \sum_{j=1}^{c} C(X_j);$$

where $V(X_j)$, $B(X_j)$ and $C(X_j)$ are the contributions of the source X_j to the corresponding synthetic indexes.

It is interesting to evaluate the relative contribution of X_j to the point and the synthetic inequality measures of the three indexes.

The relative contributions of X_j to:

1. the point indexes $I_h(Y)$, $V_h(Y)$, $G_h(Y)$ are respectively given by:

$$\beta_{h.}(X_j) = \frac{B_h(X_j)}{I_h(Y)} = \frac{\overset{+}{M}_{h.}(X_j) - \bar{M}_{h.}(X_j)}{\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)}$$

$$\omega_{h.}(X_j) = \frac{V_h(X_j)}{V_h(Y)} = \frac{M(X_j) - \bar{M}_{h.}(X_j)}{M(Y) - \bar{M}_{h.}(Y)}$$

$$\lambda_{h.}(X_j) = \frac{C_h(X_j)}{G_h(Y)} = \frac{\left[M(X_j) - \bar{M}_{h.}(X_j)\right] \cdot 2p_{h.} - \left[M(X_j) - M_{h.}(X_j)\right] \cdot \frac{n_{h.}}{N}}{\left[M(Y) - \bar{M}_{h.}(Y)\right] \cdot 2p_{h.} - \left[M(Y) - y_h\right] \cdot \frac{n_{h.}}{N}};$$

2. the synthetic indexes $I(Y)$, $V(Y)$ and $G(Y)$ are respectively given by:

$$\beta(X_j) = \frac{B(X_j)}{I(Y)} = \frac{\sum_{h=1}^{r} \beta_{h.}(X_j) \cdot I_h(X_j) \cdot \frac{n_{h.}}{N}}{I(Y)}$$

$$= \sum_{h=1}^{r} \beta_{h.}(X_j) \cdot \frac{I_h(X_j) \cdot n_{h.}}{I(Y) \cdot N}$$

$$\omega(X_j) = \frac{V(X_j)}{V(Y)} = \sum_{h=1}^{r} \omega_{h.}(X_j) \cdot \frac{V_h(X_j) \cdot n_{h.}}{V(Y) \cdot N}$$

$$\lambda(X_j) = \frac{C(X_j)}{G(Y)} = \sum_{h=1}^{r} \lambda_{h.}(X_j) \cdot \frac{G_h(X_j) \cdot n_{h.}}{G(Y) \cdot N}.$$

It is important to remark that recently, Zenga (2013) has shown that the relative contributions of X_j to the three point indexes are equal. For more details on this point, see Zenga (2013). It is also important to remark that in the same paper it is shown that the value of the contribution $C(X_j)$ to the synthetic Gini index obtained using the ‘‘two-step’’ approach is equal to those obtained by the different approach of Rao (1969), Lehrman and Yitzhaki (1984, 1985) and Radaelli and Zenga (2002, 2005).

It is worth to remark that the decomposition by sources can be extended to each of the k subpopulations. $M_\ell(Y)$, $\bar{M}_{h\ell}(Y)$ and $\overset{+}{M}_{h\ell}(Y)$ are respectively the mean, the lower mean and the upper mean of subpopulation ℓ. Obviously:

$$M_\ell(Y) = \sum_{j=1}^{c} M_\ell(X_j), \quad \bar{M}_{h\ell}(Y) = \sum_{j=1}^{c} \bar{M}_{h\ell}(X_j), \quad \overset{+}{M}_{h\ell}(Y) = \sum_{j=1}^{c} \overset{+}{M}_{h\ell}(X_j).$$

The three point inequality indexes of subpopulation ℓ , $\ell = 1, \ldots, k$, are:

$$V_{h\ell}(Y) = \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M_\ell(Y)},$$

$$I_{h\ell}(Y) = \frac{\overset{+}{M}_{h\ell}(Y) - \bar{M}_{h\ell}(Y)}{\overset{+}{M}_{h\ell}(Y)},$$

$$G_{h\ell}(Y) = \frac{M_\ell(Y) - \bar{M}_{h\ell}(Y)}{M_\ell(Y)} \cdot 2p_{h\ell} - \frac{M_\ell(Y) - y_h}{M_\ell(Y)} \cdot \frac{n_{h\ell}}{n_{.\ell}}; \quad p_{h\ell} = \frac{P_{h\ell}}{n_{.\ell}}.$$

Thus, the three synthetic measures of the subpopulation ℓ are:

$$V_{.\ell}(Y) = \sum_{h=1}^{r} V_{h\ell} \cdot \frac{n_{h\ell}}{n_{.\ell}}, \quad I_{.\ell}(Y) = \sum_{h=1}^{r} I_{h\ell} \cdot \frac{n_{h\ell}}{n_{.\ell}}, \quad G_{.\ell}(Y) = \sum_{h=1}^{r} G_{h\ell} \cdot \frac{n_{h\ell}}{n_{.\ell}}.$$

The decomposition by sources of $V_{h\ell}(Y)$ and $V_{.\ell}(Y)$ are given by: $V_{h\ell}(Y) = \sum_{j=1}^{c} V_{h\ell}(X_j)$, where $V_{h\ell}(X_j) = \frac{M_\ell(X_j) - \bar{M}_{h\ell}(X_j)}{M_\ell(Y)}$ is the contribution of X_j to the point index $V_{h\ell}(Y)$ of subpopulation ℓ; $V_{.\ell}(Y) = \sum_{j=1}^{c} V_{.\ell}(X_j)$, where $V_{.\ell}(X_j) = \sum_{h=1}^{r} V_{h\ell}(X_j) \cdot n_{h\ell}/n_{.\ell}$ is the contribution of X_j to the Bonferroni synthetic index $V_{.\ell}(Y)$ of subpopulation ℓ. Analogously: $I_{h\ell}(Y) = \sum_{j=1}^{c} B_{h\ell}(X_j)$, where $B_{h\ell}(X_j) = \frac{\overset{+}{M}_{h\ell}(X_j) - \bar{M}_{h\ell}(X_j)}{\overset{+}{M}_{h\ell}(Y)}$ is the contribution of X_j to the point index $I_{h\ell}(Y)$ of subpopulation ℓ; $I_{.\ell}(Y) = \sum_{j=1}^{c} B_{.\ell}(X_j)$, where $B_{.\ell}(X_j) = \sum_{h=1}^{r} B_{h\ell}(X_j) \cdot n_{h\ell}/n_{.\ell}$ is the contribution of X_j to the Zenga (2007) synthetic index $I_{.\ell}(Y)$ of subpopulation ℓ. Finally, $G_{h\ell}(Y) = \sum_{j=1}^{c} C_{h\ell}(X_j)$, where,

$$C_{h\ell}(X_j) = \frac{M_\ell(X_j) - \bar{M}_{h\ell}(X_j)}{M_\ell(Y)} \cdot 2p_{h\ell} - \frac{M_\ell(X_j) - M_{h\ell}(X_j)}{M_\ell(Y)} \cdot \frac{n_{h\ell}}{n_{.\ell}}$$

is the contribution of X_j to the Gini $G_{h\ell}(Y)$ point index of subpopulation ℓ, and $G_{.\ell}(X_j) = \sum_{j=1}^{c} C_{.\ell}(X_j)$, where $C_{.\ell}(X_j) = \sum_{h=1}^{r} C_{h\ell}(X_j) \cdot \frac{n_{h\ell}}{n_{.\ell}}$ is the contribution of X_j to the Gini synthetic index $G_{.\ell}(Y)$ of the subpopulation ℓ. It can be useful to evaluate for each subpopulation ℓ the relative contributions of X_j to the three synthetic indexes. For more details on this point see Zenga and Valli (2018), Pasquazzi and Zenga (2018), Zenga and Jedrzejczak (2020).

The decomposition by subpopulations and the joint decomposition by subpopulations and sources of the Bonferroni, Zenga and Gini indexes are essentially based on the decompositions of $[M(Y) - \bar{M}_{h.}(Y)]$, $[\overset{+}{M}_{h.}(Y) - \bar{M}_{h.}(Y)]$ and $[M(Y) - y_h]$ reported in Section 3.

One of the most important result of the decompositions by subpopulations of the three indexes obtained with the two step approach is that in the first step the Zenga $I_h(Y)$, the Bonferroni $V_h(Y)$ and the Gini $G_h(Y)$ point indexes are decomposed in the sum of k contributions, one for each subpopulation:

$$I_h(Y) = \sum_{\ell=1}^{k} B_{h\ell.}(Y), \quad V_h(Y) = \sum_{\ell=1}^{k} V_{h\ell.}(Y), \quad G_h(Y) = \sum_{\ell=1}^{k} C_{h\ell.}(Y);$$

where

$$B_{h\ell.}(Y) = \frac{\overset{+}{M}_{h.}(Y) - \bar{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h),$$

$$V_{h\ell.}(Y) = \frac{M_(Y) - \bar{M}_{h\ell}(Y)}{M(Y)} \cdot p(\ell|h),$$

$$C_{h\ell.}(Y) = V_{h\ell.}(Y) \cdot 2p_{h.} - \frac{M(Y) - y_h}{M(Y)} \cdot f(\ell|h) \cdot \frac{n_{h.}}{N}.$$

In addition, $B_{h\ell.}(Y)$, $V_{h\ell.}(Y)$ and $C_{h\ell.}(Y)$ are split into a within and a between components.

In particular, for $B_{h\ell.}(Y)$ we have: $B_{h\ell.}(Y) = B_{h\ell W}(Y) + B_{h\ell B}(Y)$; where

$$B_{h\ell W}(Y) = \frac{\overset{+}{M}_{h\ell}(Y) - \bar{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(\ell|h)$$

is the within component of $B_{h\ell.}(Y)$, and

$$B_{h\ell B}(Y) = \sum_{g \neq \ell} \frac{\overset{+}{M}_{hg}(Y) - \bar{M}_{h\ell}(Y)}{\overset{+}{M}_{h.}(Y)} \cdot p(\ell|h) \cdot a(g|h)$$

is the between component of $B_{h\ell.}(Y)$. Thus, the within and the between components of $I_h(Y)$ are, $I_h(Y) = \sum_{\ell=1}^{k} B_{h\ell.}(Y) = \sum_{\ell=1}^{k}[B_{h\ell W}(Y) + B_{h\ell B}(Y)] = B_{h.W}(Y) + B_{h.B}(Y)$.

Now, putting the decomposition $I_h(Y) = \sum_{\ell=1}^{k} B_{h\ell.}(Y)$ in $I(Y) = \sum_{h=1}^{r} I_h(Y) \cdot \frac{n_{h.}}{N}$ gives the following interesting $r \times k$ bivariate decomposition of the synthetic index $I(Y)$.

$$I(Y) = \sum_{h=1}^{r} \sum_{\ell=1}^{k} B_{h\ell.}(Y) \cdot \frac{n_{h.}}{N}.$$

From this bivariate decomposition we obtain

$$I(Y) = \sum_{\ell=1}^{k} \sum_{h=1}^{r} B_{h\ell.}(Y) \cdot \frac{n_{h.}}{N} = \sum_{\ell=1}^{k} B_{.\ell.}(Y).$$

Note that $B_{.\ell.}(Y) = \sum_{h=1}^{r} B_{h\ell.}(Y) \cdot n_{h.}/N$ is the contribution of the subpopulation ℓ to the synthetic index $I(Y)$, and is equal to the weighted mean of $B_{h\ell.}(Y)$ with weights $n_{h.}/N$.These results are extended to the Bonferroni and the Gini synthetic indexes. For more details on this point see Zenga and Valli (2020). Moreover, these results are extended to the joint decomposition of the three indexes. In the second part of the present paper the results of the decompositions are applied to data from 2014 central bank of Italy sample survey on household income and wealth (Bank of Italy, 2016), using the R package `ineqJD` (Arcagni and Valli, 2019). $N = 8156$ households were sampled recording the net disposable income Y, the income sources (X_1 = payroll incomes, X_2 = pensions and net transfers, X_3 = net self emplyment income and X_4 = property incomes). The subpopulations are the three geographical Italian areas North ($\ell = 1$), Centre ($\ell = 2$) and South ($\ell = 3$).

REFERENCES

Arcagni A., Valli I. (2019). *ineqJD: Inequality Joint Decomposition*. R package version 1.0. Url: https://CRAN.R-project.org/

Bank of Italy (2016). Household income and wealth in 2014. Supplements to the Statistical Bulletin - Sample Survey.

Bonferroni C. (1930). *Elementi di Statistica Generale*. Seeber, Firenze.

Gini C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, A.A. 1913-1914 Tomo LXXIII - Parte Seconda.

Lerman R., Yitzhaki S. (1984). A note on the calculation and the interpretation of the Gini index. *Economic Letters*, **15**, 363-368.

Lerman R., 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.

Meschiari S. (2021). *lateX2exp. R package version 0.5.0.* Url: https://github.com/stefano-meschiari/latex2exp

Pasquazzi L., Zenga M. (2018). Components of Gini, Bonferroni, and Zenga inequality indexes for EU income data. *Journal of Official statistics*, **34(1)**, 149-180.

Rao V. (1969). Two decompositions of concentration ratio. *Journal of the Royal Statistical Society A*, **132**, 418-425.

Radaelli P., 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 Nuechatel, Switzerland.

Radaelli P., Zenga M. (2005). On the decomposition of the Gini mean difference and concentration ratio. *Statistica & Applicazioni*, **III**(2), 5-24.

Radaelli P. (2008). A subgroup decomposition of Zenga's uniformity and inequality index. *Statistica & Applicazioni*, **VI**(2), 117-136.

Radaelli P. (2010). On the Decomposition by Subgroups of the Gini Index and Zenga's Uniformity and Inequality Indexes. *International Statistical Review*, **78**(1), 81-101.

Zenga M. (2007). Inequality curve and inequality index based on the ratios between lower and upper means. *Statistica & Applicazioni*, **V**(1), 3-27.

Zenga M., Radaelli P., Zenga Ma.(2012). Decomposition of Zenga's inequality index by sources. *Statistica & Applicazioni*, **X**(1), 3-31.

Zenga M. (2013). Decomposition by sources of the Gini, Bonferroni and Zenga inequality indexes. *Statistica & Applicazioni*, **XI**(2), 131-158.

Zenga M. (2015). Joint decomposition by subpopulations and sources of the Zenga 2007 inequality index $I(Y)$. *Statistica & Applicazioni*, **XIII**(2), 163-195.

Zenga M. (2016a). Decomposition by subpopulations of the point and synthetic Zenga (2007) inequality indexes. *Metron*, **74**(3), 375-405.

Zenga M. (2016b). Decomposition by subpopulations of the point and the synthetic Gini inequality indexes. *Statistica & Applicazioni*, **XIV**(1), 49-82.

Zenga M., Valli I. (2016). On the decomposition by subpopulations of the point and synthetic Bonferroni inequality measures. *Statistica & Applicazioni*, **XIV**(1), 3-28.

Zenga M., Valli I. (2017). Joint decomposition by subpopulations and sources of the point and synthetic Bonferroni inequality measures. *Statistica & Applicazioni*, **XV**(2), 83-120.

Zenga M., Valli I. (2018). Joint decomposition by subpopulations and sources of the point and the synthetic Gini inequality indexes. *Statistica & Applicazioni*, **XVI**(1), 31-79.

Zenga M., Jedrzejczak A. (2020). Decomposition of the Zenga inequality index I(Y) into the contributions of Macro Regions and income components - An application to data from Poland and Italy, *Argumenta Oeconomica*, **1**(44), 101-125.

Zenga M., Valli I. (2020). Decomposition by subpopulations of Gini, Bonferroni and Zenga inequality measures. *Statistica & Applicazioni*, **XVIII**(1), 69-113.