

Decomposition by sources of the ξ inequality index

Scomposizione per fonti dell'indice di disuguaglianza ξ

Arcagni, A. and Zenga, M.M.

Abstract In this work we present the decomposition by sources of the inequality index ξ proposed by Zenga in 1984. The decomposition approach is inspired by the general one presented by Zenga et al. in 2012 and by Zenga in 2013 that allows to compare different point and synthetic inequality measures, such as the Gini and the Bonferroni indexes, and the Zenga index I .

Abstract *In questo lavoro si presenta la scomposizione per fonti dell'indice di disuguaglianza ξ proposto da Zenga nel 1984. L'approccio impiegato è ispirato a quello presentato da Zenga et al. nel 2012 e da Zenga nel 2013, il quale permette anche di confrontare diverse misure puntuali e globali di disuguaglianza, come quelle di Gini e Bonferroni, e l'indice di Zenga I .*

Key words: inequality, income, sources, decomposition

1 Introduction

A very important characteristic of an inequality index is its suitability in the decompositions by subgroups and by sources (see Rao [3], Radaelli [1], Radaelli and Zenga [2], Zenga et al. [7]). In particular Zenga et al. [7] presented an approach to decompose by sources the Zenga inequality index I starting from the decomposition of the point measure. This approach has been used by Zenga [6] to decompose also the Gini and Bonferroni indexes. This work uses the aforementioned approach to decompose the inequality index ξ proposed by Zenga in 1984 [4].

Arcagni, A.

Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano - Bicocca, Edificio U7, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy, e-mail: alberto.arcagni@unimib.it

Zenga, M.M.

Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano - Bicocca, Edificio U7, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy, e-mail: michele.zenga@unimib.it

Let Y be a continuous and nonnegative random variable with density function $f(y)$, expectation $E(Y) < \infty$, and with strictly increasing distribution function $F(y)$ and *first incomplete moment*

$$Q(y) = \int_0^y \frac{t}{E(Y)} f(t) dt.$$

The point inequality measure and the corresponding synthetic inequality measure are

$$Z(p) = \frac{y^*(p) - y(p)}{y^*(p)}, \quad \xi = \int_0^1 Z(p) dp.$$

where $y(p) = F^{-1}(p)$ and $y^*(p) = Q^{-1}(p)$.

For all the values of y in the support of Y and for all the values of p in $(0, 1)$, the following relations hold

$$Q(y) \leq F(y), \quad y^*(p) \geq y(p), \quad Z(p) \geq 0.$$

Note that the point measure is constant for the Log-Normal distribution [4].

2 Definitions and notation

Let $X_1, X_2, \dots, X_c$ be c variates (*income sources*) observable on each of the n units (individuals or households) of a finite population and let $Y = \sum_{j=1}^c X_j$ be the *total income*. The values of the c income sources referring to a single population unit, may be reported in a row of a *data matrix* and *the rows may be ordered according to the values of the variate Y* . The data matrix is represented in Table 1 edged by

- the row sums, i.e. the ordered values of the total income Y

$$y_{(i)} = \sum_{j=1}^c x_{ij} \quad \text{for } i = 1, 2, \dots, n;$$

- the column sums, i.e. the sources total amount

$${}_jT = \sum_{i=1}^n x_{ij} \quad \text{for } j = 1, 2, \dots, c;$$

- the overall income of the whole population

$$T = \sum_{j=1}^c {}_jT = \sum_{j=1}^c \sum_{i=1}^n x_{ij} = \sum_{i=1}^n y_{(i)}. \quad (1)$$

In this paper, for the sake of simplicity, n different values of the total income Y are considered

$$0 \leq y_{(1)} < \dots < y_{(i)} < \dots < y_{(n)}.$$

Table 1 Edged data matrix

	X_1	$\dots$	X_j	$\dots$	X_c	Y
1	x_{11}	$\dots$	x_{1j}	$\dots$	x_{1c}	$y_{(1)}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
i	x_{i1}	$\dots$	x_{ij}	$\dots$	x_{ic}	$y_{(i)}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
n	x_{n1}	$\dots$	x_{nj}	$\dots$	x_{nc}	$y_{(n)}$
	$1T$	$\dots$	jT	$\dots$	cT	T

The *empirical distribution function* $F(y)$ and the *first incomplete moment* $Q(y)$ are defined as follows

$$F(y) = (\text{number of } y_{(i)} \leq y)/n, \quad Q(y) = \sum_{y_{(i)} \leq y} y_{(i)} / (n\bar{y})$$

where $\bar{y}$ is the arithmetic mean of $y_{(1)}, y_{(2)}, \dots, y_{(n)}$. $F(y)$ and $Q(y)$ are step functions, therefore the definitions of their inverse functions must be generalized as follows

$$y(p) = F^{-1}(p) = \begin{cases} \inf\{y_{(i)} : F(y_{(i)}) \geq p\} & \text{for } 0 < p \leq 1 \\ \inf\{y_{(i)} : F(y_{(i)}) > p\} & \text{for } p = 0 \end{cases}$$

$$y^*(p) = Q^{-1}(p) = \begin{cases} \inf\{y_{(i)} : Q(y_{(i)}) \geq p\} & \text{for } 0 < p \leq 1 \\ \inf\{y_{(i)} : Q(y_{(i)}) > p\} & \text{for } p = 0. \end{cases}$$

Let $f(y_{(i)}) = 1/n$ and $q(y_{(i)}) = y_{(i)}/(n\bar{y})$. As described by Zenga [5], in order to obtain the values that $Z(p)$ assumes it is possible to use the *cograduation table* between the variables Y and Y^* . Y is the variable that assumes values $y_{(1)}, \dots, y_{(n)}$ with weights $f(y_{(1)}), \dots, f(y_{(n)})$ and Y^* is the variable that assumes the same values $y_{(1)}, \dots, y_{(n)}$ but with weights $q(y_{(1)}), \dots, q(y_{(n)})$.

The cograduation table gives the couples $(y(p_r), y^*(p_r))$ with weights $v_r > 0$, for $r = 1, 2, \dots, k$, to assign to

$$Z(p_r) = \frac{y^*(p_r) - y(p_r)}{y^*(p_r)} \quad (2)$$

with $p_r = \sum_{i=1}^r v_i$. Therefore the index ξ can be expressed as follows

$$\xi = \sum_{r=1}^k Z(p_r) v_r. \quad (3)$$

3 Decomposition by sources

From Equations (2) the point inequality measure can be decomposed as follows

$$\begin{aligned}
Z(p_r) &= \frac{y^*(p_r) - y(p_r)}{y^*(p_r)} = \frac{\sum_{j=1}^c x_j^*(p_r) - \sum_{j=1}^c x_j(p_r)}{y^*(p_r)} = \\
&= \sum_{j=1}^c \frac{x_j^*(p_r) - x_j(p_r)}{y^*(p_r)} = \sum_{j=1}^c H_r(X_j)
\end{aligned} \tag{4}$$

where

$$H_r(X_j) = \frac{x_j^*(p_r) - x_j(p_r)}{y^*(p_r)}$$

is the *contribution of the income source X_j to the point measure $Z(p_r)$* ,

$$x_j(p_r) = \{x_{tj} \text{ for } t : y_{(t)} = y(p_r)\} \tag{5}$$

is the value of the variable X_j in the data matrix in correspondence of the p_r -th quantile of the variable Y , and

$$x_j^*(p_r) = \{x_{tj} \text{ for } t : y_{(t)} = y^*(p_r)\} \tag{6}$$

is the value of the variable X_j in the data matrix in correspondence of the p_r -th quantile of the variable Y^* .

From Equation (3) and (4) it is easy to obtain the following result

$$\xi = \sum_{r=1}^k Z(p_r) v_r = \sum_{r=1}^k \sum_{j=1}^c H_r(X_j) v_r = \sum_{j=1}^c \sum_{r=1}^k H_r(X_j) v_r = \sum_{j=1}^c H(X_j) \tag{7}$$

where

$$H(X_j) = \sum_{r=1}^k H_r(X_j) v_r \quad \text{for } j = 1, 2, \dots, c$$

is the *contribution of the source X_j to the inequality synthetic index ξ of the total income Y* .

4 Example

The results of the previous section are illustrated by a 8×2 data matrix. The edged data matrix used in this example is represented in Table 2. Each observation has relative frequency $f(y_{(i)}) = 1/n = 1/8$. The cograduation table between the variables Y and Y^* is reported in Table 3, where the values $q(y_{(i)})$ are also reported. By ordering the nonzero cells $(y(p_r), y^*(p_r))$ from the bottom-left to the top-right, their values are the weights v_r , for $r = 1, 2, \dots, k$, where $k = 13$ is the number of nonzero cells. Finally, Table 4 allows to obtain the point and synthetic inequality measures and the corresponding contributions of each income source. In Figure 1 are represented the inequality curve $Z(p)$ and the contribution of X_1 to the inequality of the total income Y . The difference between the $Z(p)$ curve and the contribution of X_1

Table 2 Edged data matrix of the example

i	X_1	X_2	Y
1	12	8	20
2	22	10	32
3	20	28	48
4	45	23	68
5	37	55	92
6	64	76	140
7	100	60	160
8	110	130	240
jT	410	390	800

Table 3 Cograduation table between the variables Y and Y^* of the example

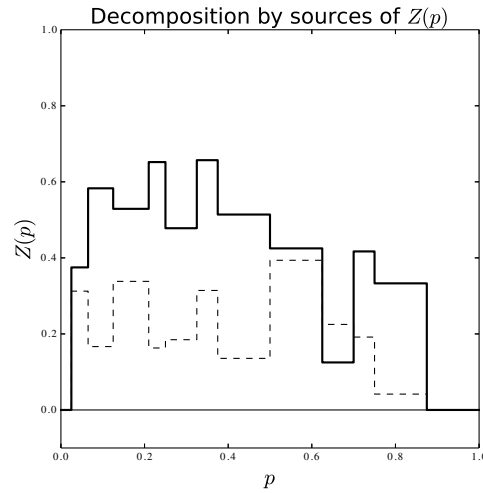
Y^*	Y	20	32	48	68	92	140	160	240	$q(y_j)$
240	-	-	-	-	-	-	0.05	0.125	0.125	0.300
160	-	-	-	-	-	0.125	0.075	-	-	0.200
140	-	-	-	0.05	0.125	-	-	-	-	0.175
92	-	-	0.04	0.075	-	-	-	-	-	0.115
68	-	-	0.085	-	-	-	-	-	-	0.085
48	0.06	-	-	-	-	-	-	-	-	0.060
32	0.04	-	-	-	-	-	-	-	-	0.040
20	0.025	-	-	-	-	-	-	-	-	0.025
$f(y_j)$	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1.000

Table 4 Summary of the example

r	v_r	p_r	$y(p_r)$	$y^*(p_r)$	$Z(p_r)$	$x_1(p_r)$	$x_1^*(p_r)$	$H_1(p_r)$	$x_2(p_r)$	$x_2^*(p_r)$	$H_2(p_r)$
1	0.025	0.025	20	20	0.000	12	12	0.0000	8	8	0.0000
2	0.040	0.065	20	32	0.375	12	22	0.3125	8	10	0.0625
3	0.060	0.125	20	48	0.583	12	20	0.1667	8	28	0.4167
4	0.085	0.210	32	68	0.529	22	45	0.3382	10	23	0.1912
5	0.040	0.250	32	92	0.652	22	37	0.1630	10	55	0.4891
6	0.075	0.325	48	92	0.478	20	37	0.1848	28	55	0.2935
7	0.050	0.375	48	140	0.657	20	64	0.3143	28	76	0.3429
8	0.125	0.500	68	140	0.514	45	64	0.1357	23	76	0.3786
9	0.125	0.625	92	160	0.425	37	100	0.3937	55	60	0.0312
10	0.075	0.700	140	160	0.125	64	100	0.2250	76	60	-0.1000
11	0.050	0.750	140	240	0.417	64	110	0.1917	76	130	0.2250
12	0.125	0.875	160	240	0.333	100	110	0.0417	60	130	0.2917
13	0.125	1.000	240	240	0.000	110	110	0.0000	130	130	0.0000
Synt.					0.3790			0.1852			0.1939

represents the contribution of X_2 to the inequality of the total income Y . We conclude the example by comparing the relative contribution of the sources to synthetic inequality index ξ with its share to the total amount of the income Y . From Equation (7) the relative contribution of the j -th source to synthetic inequality index ξ is evaluated as ${}_j\eta = H(X_j)/\xi$ and from Equation (1) the share of the j -th source to the total amount of the income Y is evaluated as ${}_j\gamma = {}_jT/T$. Therefore, we can observe that the source X_1 contributes for the ${}_1\eta = 48.02\%$ to the inequality of Y and its share is the ${}_1\gamma = 51.25\%$ of the total amount of the income Y .

Fig. 1 Decomposition by sources of $Z(p)$, bold line: $Z(p)$ curve, dashed line: X_1 contribution



5 Conclusions

In this work we proposed the decomposition by sources of the inequality index ξ . This approach allows to study the contributions of the sources to the point and synthetic inequality measures of the total income.

Observe that it is possible to obtain the same total income with different combinations of income sources, therefore the Equations (5) and (6) can represent sets with more than one element. The solution in such case and an application to actual data will be presented at the conference.

References

1. Radaelli, P.: On the decomposition by subgroups of the Gini index and Zenga's uniformity and inequality indexes. *Int. Stat. Rev.* **78**(1), 81–101, (2010).
2. Radaelli, P., Zenga, M.M.: On the decomposition of Gini mean difference and concentration ratio, *Stat. & Applicazioni*, **III**(2), 5–24, (2005)
3. Rao, V.: Two decomposition of concentration ratio. *J. of the R. Stat. Soc. Series A*, **132**, 418–425 (1969)
4. Zenga, M. M.: Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili di reddito. *Giornale degli Economisti e Annali di economia* **5-6**, 301–326 (1984)
5. Zenga, M.M.: Impiego delle tabelle di cograduazione per la determinazione dell'indice puntuale di concentrazione $Z(p)$. *Stat. Applicata*, **III**(3), 283–291 (1991)
6. Zenga, M.M.: Decomposition by sources of the Gini, Bonferroni and Zenga inequality indexes. *Stat. & Applicazioni*, **XI**(2) (2013)
7. Zenga, M.M., Radaelli, P., Zenga, Ma.: Decomposition of Zenga's inequality index by sources. *Stat. & Applicazioni*, **X**(1), 3–31, (2012)