Statistica & Applicazioni - Vol. XVIII, n. 1, 2020, pp. 61-67

THE JOINT DECOMPOSITION OF THE PIETRA INDEX

Francesco Porro*
Mariangela Zenga*

SUMMARY

In this paper, a multi-decomposition of the Pietra index is presented. This innovative methodology allows to achieve a relevant task, since it combines simultaneously the two most celebrated kinds of decomposition: by sources and by subpopulations. The key result of the proposed procedure is the detail level of decomposition: it allows to split the value of the index, by assessing the contribution of each source in each subpopulation. It is worth noting that this final result is not reached by all the decomposition procedures, since most of them do not permit to identify the contributions at so detailed level. The proposed joint decomposition is a sort of generalization, since from it two decompositions by sources and by subpopulations of the Pietra index, already proposed in the literature, can be obtained. Beyond the methodological details, an application based on the Survey Household Income and Wealth 2018 – carried out by Bank of Italy – is provided in order to clarify the advantages of the procedure.

Keywords: Pietra Index, Schutz Index, Hoover Index, Joint Decomposition, Decomposition by Subpopulations, Decomposition by Sources.
DOI: 10.26350/999999_000029
ISSN: 18246672 (print)

1. INTRODUCTION

The Pietra index is a very well-known index with more than a century of history. It was introduced in the literature in the early of 1900 (Pietra, 1915), but it has been ‘‘rediscovered’’ many times: indeed, it coincides with the Ricci index (Ricci, 1916), with the Hoover index (Hoover, 1936), and with the Schutz coefficient (Schutz, 1951). Some other papers refer to the Pietra index also as the Robin Hood index. Perhaps the reason of such large notoriety and longevity is its very intuitive interpretation, since the Pietra index represents the portion of the total amount to be redistributed from the owners with more than the mean, to the others, in order to obtain the situation of minimum inequality (where all the units have the same amount). In the literature, two proposals of decomposition by subpopulation of the Pietra index can be found in Frosini (2012) and in Habib (2012). More recently,

* Dipartimento di Statistica e Metodi Quantitativi - Università di Milano-Bicocca - Piazza dell’Ateneo Nuovo, 1 - 20126 MILANO
(e-mail: francesco.porro1@unimib.it; mariangela.zenga@unimib.it).

Porro and Zenga (2020) proposed two procedures to decompose the Pietra index, the former by subpopulations, the latter by sources.

2. DEFINITIONS AND INITIAL SETTINGS

Let Y be a non-negative statistical variable on a population of size N. Let $y_1 < y_2 < \cdots < y_r$ denote the distinct values assumed by Y with frequencies $n_{1\cdot}, n_{2\cdot}, \cdots, n_{r\cdot}$, respectively. Obviously, $\sum_{h=1}^{r} n_{h\cdot} = N$. Let $M(Y) = \sum_{h=1}^{r} y_h \frac{n_{h\cdot}}{N}$ be the arithmetic mean of Y and let

$$S_{M(Y)} = \sum_{h=1}^{r} |y_h - M(Y)| \frac{n_{h\cdot}}{N}$$

be the Mean Absolute Deviation (MAD) of Y from $M(Y)$. The inequality index P proposed in Pietra (1915) is defined as:

$$\mathcal{P} = \frac{S_{M(Y)}}{2M(Y)}.$$

An alternative expression for the Pietra index that will be useful for the decomposition procedure, is now described. For each y_h (with $h = 1, 2, \ldots, r$) the population can be split in two non-overlapping groups, the lower group and the upper one: the first one, denoted by $\{Y \leq y_h\}$, containing the first $P_{h\cdot} = \sum_{t=1}^{h} n_{t\cdot}$ units; the second one, denoted by $\{Y > y_h\}$, with the other $N - P_{h\cdot}$ units. By defining $\bar{h}$ as $\bar{h} = max\{h\colon\ y_h \leq M(Y)\}$ and by considering the definition of lower group, the Pietra index can be computed as:

$$\mathcal{P} = \frac{M(Y) - \bar{M}_{\bar{h}\cdot}(Y)}{M(Y)} p_{\bar{h}\cdot},$$

where $$p_{\bar{h}\cdot} = \frac{P_{\bar{h}\cdot}}{N} = \frac{\sum_{h=1}^{\bar{h}} n_{h\cdot}}{N} \qquad \text{and} \qquad \bar{M}_{\bar{h}\cdot}(Y) = \frac{\sum_{h=1}^{\bar{h}} y_h n_{h\cdot}}{P_{\bar{h}}}$$

are the cumulative relative frequency corresponding to $y_{\bar{h}}$, and the lower mean at $\bar{h}$, that is the arithmetic mean of lower group $\{Y \leq M(Y)\}$, respectively. It is worth remarking that the ratio $\frac{M(Y) - \bar{M}_{h\cdot}(Y)}{M(Y)}$ is the pointwise inequality measure of Bonferroni index, at $\bar{h}$. See Zenga and Valli (2016) for further details.

3. THE JOINT DECOMPOSITION

The two decompositions proposed by Porro and Zenga (2020) are now reported.

THEOREM 1 (Decomposition by subpopulations)
Let the population be divided into k (with $k \geq 2$) subpopulations, as showed in Ta-

ble 1. The value $n_{\cdot g}$ denotes the size of subpopulation g, and n_{hg} denotes the frequency of y_h in the subpopulation g, (with $h = 1, ...r$; $g = 1, ..., k$). Then the following decomposition of the Pietra index holds:

$$\mathcal{P} = \sum_{l=1}^{k}\sum_{g=1}^{k} \frac{M_g(Y) - \bar{M}_{\bar{h}l}(Y)}{M(Y)} \frac{n_{\cdot g}}{N} p(l|\bar{h}) \cdot p_{\bar{h}\cdot}$$
$$= \sum_{l=1}^{k}\sum_{g=1}^{k} V_{\bar{h}lg}(Y) p_{\bar{h}\cdot} = \sum_{l=1}^{k} \frac{M(Y) - \bar{M}_{\bar{h}l}(Y)}{M(Y)} p_{\bar{h}\cdot} = \sum_{l=1}^{k} V_{\bar{h}l\cdot}(Y) p_{\bar{h}\cdot}$$

where $p(l|\bar{h}) = \frac{p_{\bar{h}l}}{p_{\bar{h}\cdot}}$ is the relative frequency of the subpopulation l in the lower group $\{Y \leq M(Y)\}$, $M_g(Y)$ the arithmetic mean of Y in the subpopulation g, $\bar{M}_{\bar{h}l}(Y)$ the lower mean of Y in the subpopulation l, evaluated at $\bar{h}$. The quantity $V_{\bar{h}l\cdot}(Y) \cdot p_{\bar{h}\cdot}$ is the contribution to $\mathcal{P}$ due to subpopulation l.

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

	Subpopulations				
Y	1	2	$\cdots$	k	
y_1	n_{11}	n_{12}	$\cdots$	n_{1k}	$n_{1.}$
y_2	n_{21}	n_{22}	$\cdots$	n_{2k}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
y_r	n_{r1}	n_{r2}	$\cdots$	n_{rk}	$n_{r.}$
	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.k}$	

The result of this decomposition is a $(k \times k)$-matrix with the entries $V_{\bar{h}lg}(Y)p_{\bar{h}\cdot}$, showing how the value of $\mathcal{P}$ is split across the subpopulations. As described in Porro and Zenga (2020), the sum of the entries of column l represents the contribution to $\mathcal{P}$ related to subpopulation l. Also the classical decomposition in Within and Between components can easily achieved: the former is the sum of the entries in the main diagonal, the latter is the sum of all other ones.

THEOREM 2 (Decomposition by sources)
Let the variable Y be the sum of c (with $c \geq 2$) sources $Y = \sum_{j=1}^{c} X_j$. Then the following decomposition of the Pietra index holds:

$$\mathcal{P} = \sum_{j=1}^{c} \frac{M(X_j) - \bar{M}_{\bar{h}\cdot}(X_j)}{M(Y)} p_{\bar{h}\cdot} = \sum_{j=1}^{c} W_{\bar{h}\cdot}(X_j) p_{\bar{h}\cdot},$$

where $M(X_j)$ and $\bar{M}_{\bar{h}\cdot}(X_j)$ are the arithmetic mean and the lower mean at $\bar{h}$ of the *source X_j, respectively. The quantity $W_{\bar{h}\cdot}(X_j)p_{\bar{h}\cdot}$ represents the contribution of the source X_j to $\mathcal{P}$.*

By combining these two procedures, a new joint decomposition, simultaneously by subpopulations and by sources, can be introduced.

THEOREM 3 (Joint decomposition)
Let the population be divided into k subpopulations, and let the variable Y be the sum of c sources. Then the following joint decomposition of the Pietra index holds:

$$\mathcal{P} = \sum_{l=1}^{k}\sum_{g=1}^{k}\sum_{j=1}^{c}\frac{M_g(X_j) - \bar{M}_{\bar{h}l}(X_j)}{M(Y)} \cdot \frac{n_{\cdot g}}{N} \cdot p(l|\bar{h}) \cdot p_{\bar{h}\cdot}$$

$$= \sum_{l=1}^{k}\sum_{j=1}^{c}\frac{M(X_j) - \bar{M}_{\bar{h}l}(X_j)}{M(Y)} \cdot p(l|\bar{h}) \cdot p_{\bar{h}\cdot} = \sum_{l=1}^{k}\sum_{j=1}^{c} V_{\bar{h}l\cdot}(X_j)p_{\bar{h}\cdot},$$

where $V_{\bar{h}l\cdot}(X_j)p_{\bar{h}\cdot}$ represents the contribution to $\mathcal{P}$ of the source X_j in the subpopulation l.

This joint decomposition states that the Pietra index can be split in c $(k \times k)$-matrices: each one shows how the inequality due to the corresponding source is partitioned among the subpopulations. Analogously as before, for each matrix, the sum of the values in the main diagonal provides the Within component to the inequality due to the corresponding source, and the sum of all the remaining values furnishes the Between component. It is worth pointing out that the two aforementioned decompositions (by subpopulations and by sources) introduced in Porro and Zenga (2020) can be re-obtained as ''marginal'' decompositions from the joint one. This joint decomposition allows to assess the contribution to the total inequality related to each source, due to each subpopulation: such detail level permits to better describe the investigated phenomenon, also by remarking any different behavior of the sources.

4. AN APPLICATION TO HOUSEHOLD INCOMES

An application of the proposed joint decomposition procedure is now described. The dataset, provided by Bank of Italy, has been collected in 2016 and published in 2018. It comes from the Survey on Household Income and Wealth (SHIW), which covers 7420 households composed of 16462 individuals. The analyzed variable Y is the *Household Net Disposable Income*, that is the sum of four sources: X_1 = *Payroll income*, X_2 = *Pensions and net transfers*, X_3 = *Net self employment income*, X_4 = *Property incomes*. The households are grouped (by subpopulations) according to their geographical area. The considered areas are $k = 3$: *North* (*Subpop* $= 1$) with 43.1% of households, *Center* (*Subpop* $= 2$) with 21.6%, and *South and Islands* (*Subpop* $= 3$) with the remaining 35.3%. The computations provides $\mathcal{P} = 0.4033$, revealing a medium level of inequality. The four matrices of the joint decomposition are reported in Table 2: in their last row the partition by subpopulations of the contribution of each source is displayed. It is worth remarking that the *South and Islands* is the subpopulation that more counts in the contributions of all the sources, with the percentages equal to 47.7%, 54.3%, 42.9%, and 51.1%, respectively. In Table 3 the decomposition by subpopulations for the *Household Net Disposable Income* is shown: the contributions of each subpopulation are quite in line with the ones related to the single sources. The most relevant discrepancy is related to the source *Net self employment income* (X_3) where the percentages are quite different, meaning that for that source the inequality distribution is quite different with respect to the inequality distribution of Y. From the Table 3 the decomposition in Within and Between components can be achieved: the values of the two components are 0.1355 (33.6%), and 0.2678 (66.4%), respectively. From an analysis of data in Table 4 it is possible to argue that the source *Payroll income* X_1 is the most relevant for the inequality in all the three subpopulations, with a remarkable percentage of 47.9% in the subpopulation *Center*. The behavior of the sources is reflected also by the variable Y, as it can be observed in the last column.

TABLE 2. - *The 4* (3×3)*-matrices of the joint decomposition, one for each source*

Source X_1		*Subpop.*		
		1	2	3
Subpop.	1	0.0309	0.0181	0.0431
	2	0.0125	0.0075	0.0181
	3	0.0103	0.0072	0.0177
Total		0.0537 (32.5%)	0.0328 (19.8%)	0.0789 (47.7%)

Source X_2		*Subpop.*		
		1	2	3
Subpop.	1	0.0185	0.0086	0.0284
	2	0.0091	0.0042	0.0111
	3	0.0014	0.0001	0.0074
Total		0.0290 (31.6%)	0.0129 (14.1%)	0.0499 (54.3%)

Source X_3		*Subpop.*		
		1	2	3
Subpop.	1	0.0160	0.0079	0.0181
	2	0.0057	0.0028	0.0064
	3	0.0046	0.0021	0.0049
Total		0.0263 (38.4%)	0.0128 (18.7%)	0.0294 (42.9%)

Source X_4		*Subpop.*		
		1	2	3
Subpop.	1	0.0168	0.0067	0.0227
	2	0.0088	0.0036	0.0118
	3	0.0023	0.0003	0.0052
Total		0.0279 (36.0%)	0.0100 (12.9%)	0.0397 (51.1%)

TABLE 3. - *The decomposition by subpopulations matrix for the total inequality of the Household Net Disposable Income (Y)*

Income Y		*Subpop.* 1	2	3	
Subpop.	1	0.0822	0.0413	0.1123	
	2	0.0361	0.0181	0.0504	
	3	0.0186	0.0091	0.0352	
Total		0.1369 (33.9%)	0.0685 (17.0%)	0.1979 (49.1%)	$\mathcal{P}$ = 0.4033

TABLE 4. - *The joint decomposition and the decomposition by sources matrix for the total inequality.*

Source	*Subpopulations* 1	2	3	*Contribution to* $\mathcal{P}$: $V_{\bar{h}l\cdot}(X_j)p_{\bar{h}\cdot}$
X_1	0.0537 (39.2%)	0.0328 (47.9%)	0.07892 (39.9%)	0.1654 (41.0%)
X_2	0.029 (21.2%)	0.0129 (18.8%)	0.0499 (25.2%)	0.0918 (22.8%)
X_3	0.0263 (19.2%)	0.0128 (18.7%)	0.0294 (14.9%)	0.0685 (17.0%)
X_4	0.0279 (20.4%)	0.0100 (14.6%)	0.0397 (20.1%)	0.0776 (19.2%)
				$\mathcal{P}$ = 0.4033

5. CONCLUSIONS

In this paper a new procedure to jointly decompose by subpopulations and by sources the Pietra index is proposed. This new procedure has the important advantage to assess the contribution to the Pietra index due to each subpopulation, related to each source. The application of such methodology to the dataset considered in the application, allows to state -among the other remarks- that for a specific subpopulation (*Center*), the contributions of the four sources are different from those related to the entire sample: in particular this holds for the sources *Payroll income* X_1 (47.9% versus 41%) and *Property incomes* X_4 (14.6% versus 19.2%). Without a powerful methodological tool with a very high detail level, like the proposed joint decomposition, this peculiarity is difficult to be revealed.

ACKNOWLEDGMENTS

The first author wants to thank the URBiNAT project (which has received funding from the EU Horizon H2020 research and innovation program under grant agreement No 776783). Some ideas developed in this paper arose for the first time during the participation in such project. Interesting and useful discussions with some project partners helped to better identify the research questions. Nevertheless, the concepts and conclusions presented reflect the authors view only.

REFERENCES

Frosini B.V. (2012). Approximation and decomposition of Gini, Pietra-Ricci and Theil inequality measures. *Empirical Economics*, **43**, 175-197.

Habib E.A. (2012). On the decomposition of the Schutz coefficient: an exact approach with an application. *Electronic Journal of Applied Statistical Analysis*, **5**(2), 187-198.

Hoover E.M. Jr. (1936). The Measurement of Industrial Localization. *The Review of Economics and Statistics*, **18**(4), 162-171.

Pietra G. (1915). Delle relazioni tra gli indici di variabilità. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, **74**(2).

Porro F., Zenga Ma. (2020). Decomposition by sources and by subpopulations of the Pietra index: two applications to professional football teams in Italy. *Submitted*.

Ricci U. (1916). L'indice di variabilità e la curva dei redditi. *Giornale degli Economisti e Rivista di Statistica*, **53**, 177-228.

Schutz R. (1951). On the Measurement of Income Inequality. *The American Economic Review*, **41**(1), 107-122.

Zenga M., Valli I. (2016). On the decomposition by subpopulations of the point and synthetic Bonferroni inequality measures. *Statistica & Applicazioni*, **14**(1), 3-28.