Statistical estimation of gap of decomposability of the general poverty index

Haidara, Mohamed Cheikh
Université Gaston Berger, UFRSAT, LERSTAD
Route de Ngallèle
Saint-Louis (BP 234), Sénégal
E-mail: chaidara@ufrsat.org

Lo, Gane Samb
Université Gaston Berger, UFRSAT, LERSTAD
Route de Ngallèle
Saint-Louis (BP 234), Sénégal
E-mail: gane-samb.lo@ugb.edu.sn, ganesamblo@ufrsat.org

Introduction

We are concerned in this paper with the statistical estimation of the gap of decomposability of the class of the statistical poverty indices in general by significant confidence intervals. we would be able to handle seperated analyses in the subgroups and report the global case. Suppose that we have some statistic of the functional form $J_n = J(Y_1...Y_n)$ where $\mathcal{E} = \{Y_1, ..., Y_n\}$ is a random sample of the random variable Y defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and drawn from some specific population. Now suppose that this population is devided into K subgroups $S_1...S_K$ and let us, for each $i \in \{1, ..., K\}$, denote the subset of the random sample $\{Y_1..., Y_n\}$ coming from S_i by $\mathcal{E}_i = \{Y_{1,i}, ..., Y_{n_i,i}\}$ and then put $J_{n_i}(i) = J(Y_{1,i}, ..., Y_{n_i,i})$. The statistic J_n is said to be decomposable only whenever one always has

$$J_n = \frac{1}{n} \sum_{i=1}^{K} n_i J_{n_i}(i).$$

Since this property is very partical in welfare analyses, practitioner usually prefer decomposable indices such as the Foster-Greer-Thorbeck (FGT) [9] family and the Chakravarty family [7] however, the weighted measures, which are not decomposable, have very interesting properties in poverty analysis. Dismissing them only for non-decomposability would result in a disaster.

A brief reminder on Poverty measures

We consider a population of individuals or households, each of which having a random income or expenditure Y with distribution function $G(y) = \mathbb{P}(Y \leq y)$. In the sequel, we use Y as an income variable while it might be any positive random variable. An individual is classified as poor whenever his income or expenditure Y fulfills $Y < Z$, where Z is a specified threshold level.
Consider now a random sample $Y_1, Y_2, ...Y_n$ of size n of incomes, with empirical distribution function $G_n(y) = n^{-1}\#\{Y_i \leq y : 1 \leq i \leq n\}$. The sample is then equal to $Q_n = nG_n(Z)$. Given these preliminaries, we introduce measurable functions $A(p, q, z)$, $w(t)$, and $d(t)$ of $p, q \in \mathbb{N}$, and $z, t \in \mathbb{R}$. Set

$$B(Q_n) = \sum_{i=1}^{Q_n} w(i).$$

Let now $Y_{1,n} \leq Y_{2,n} \leq ... \leq Y_{n,n}$ be the order statistics of the sample $Y_1, Y_2, ...Y_n$ of Y. We consider general poverty indices (GPI) of the form

$$J_n = GPI_n = \frac{A(Q_n, n, , Z)}{nB(Q_n)} \sum_{j=1}^{Q_n} w(\mu_1 n + \mu_2 Q_n - \mu_3 j + \mu_4) d\left(\frac{Z - Y_{j,n}}{Z}\right),$$

where $\mu_1, \mu_2, \mu_3, \mu_4$ are constants. This global form of poverty indices was introduced in [1] (see also [2], [1] and [3]) as an attempt to unify the large number of poverty indices that have been introduced in the literature since the pioneering work of the Nobel Prize winner, Amartya Sen(1976) who first derived poverty measures (see [5]) from an axiomatic point of view. A survey of these indices is to be found in Zheng [8].

Statistical decomposability

From now on, we suppose in our study, that population of households is divided into K subgroups such that, for each $i \in \{1, ..., K\}$ that the probability that a randomly drawn household comes from the i^{th} subgroup is $p_i > 0$, with $p_1 + ... + p_K = 1$. Let us suppose that we draw a sample of size n from the population : $Y_1, ..., Y_n$ and let us denote those of the n_i^* observations coming from the i^{th} subgroup, $(1 \leq i \leq K)$ by $Y_{i,j}$, $j = 1, ..., n_i^*$. Let $J_{n_i^*}(G_i) = J_{n_i^*}(Y_{i,1}, ..., Y_{i,n_i^*})$ the empirical index measured on the i^{th} subgroup and $J_n(G)$ the global index. Clearly, decomposability implies for all $n \geq 1$,

$$gd_n = J_n - \frac{1}{n} \sum_{i=1}^{K} n_i^* J_{n_i^*} \equiv 0.$$

Surely, $n^* = (n_1^*, ...n_K^*)$ follows a multimonial law with parameters n and $p = (p_1, ..., p_K)$. Since each $p_i > 0$, we surely have that each $n_i^* \to \infty$ $a.s.$, as $n \to \infty$. We will have by (a) and by exact poverty index ([2]),

$$gd_n = J_n(G) - \frac{1}{n} \sum_{i=1}^{K} n_i^* J_{n_i^*}(G_i) \to_P gd = J(G) - \sum_{i=1}^{K} p_i J_i(G_i).$$

So we have the exact gap of decomposability gd. It follows that gd is zero if the distribution of the income is the same over all the population, that the more homogeneous the income is over the population, the lower the gap of decomposability gd is. Now we want to find the law of

$$gd_n^* = \sqrt{n}(gd_n - gd)$$

for a more accurate estimation of gd by confidence intervals. At this step, we have to precise our random scheme. We put a probability space $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ and put $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$. We draw the observations in the following way. In each trial, we draw a subgroup, the ith subgroup having the occuring probability p_i. And we put. We conclude that $\{Y_1, ..., Y_n\}$ is an independant sample drawn from $G(y)$ we can prouve that

$$G(y) = \sum_{i=1}^{K} p_i G_i(y),$$

the mixture of the distribution functions of the subgroups incomes. We readily see that conditionnally on $n^* \equiv (n_1^*, n_2^*, ..., n_K^*) = (n_1, n_2, ..., n_K) \equiv \overline{n}$ with $n_1 + n_2 + ... + n_K = n$, $\{Y_{i,j},\ 1 \leq j \leq n_i^*\}$ are independant r.v.'s with d.f. G_i.

Our results

The results stated here hold for a very large class of poverty measures summerized in the GPI. This is why we need the representation Theorem of the GPI in [4]. In fact we do not need here the complete form of [4], but a special case of it, based on the assumptions described below. For that, suppose that G_i $(1 \leq i \leq K)$, the distribution function of the income for the ith subgroup, and G the distribution function of the income for the global population. Let also

$$\gamma(x) = d(\frac{Z-x}{Z})1_{(x \leq Z)}$$

and

$$e(x) = 1_{(x \leq Z)}.$$

The following assumptions are required. (HD0) $G_0(Z) \in]0,1[$ for $G_0 \in \{G, G_1, ..., G_K\}$. (HD1) There exists a function $h(p,q)$ of $(p,q) \in \mathbb{N}^2$ and a function $c(s,t)$ of $(s,t) \in (0,1)^2$ such that, as $n \to +\infty$,

$$\max_{1 \leq j \leq Q} \left| A(n,Q)h^{-1}(n,Q)w(\mu_1 n + \mu_2 Q - \mu_3 j + \mu_4) - c(Q/n, j/n) \right| = o_P(n^{-1/2}).$$

(HD2) There exists a function $\pi(s,t)$ of $(s,t) \in \mathbb{R}^2$ such that
as $n \to +\infty$,

$$\max_{1 \leq j \leq Q} \left| w(j)h^{-1}(n,Q) - \frac{1}{n}\pi(Q/n, j/n) \right| = o_P(n^{-1/2}).$$

(HD3) The bivariate functions c and π have continuous partial differentials. (HD4) For a fixed x, the functions $y \to \frac{\partial c}{\partial y}(x,y)$ and $y \to \frac{\partial \pi}{\partial y}(x,y)$ are monotone. (HD5) G_0 is strictly increasing for any $G_0 \in \{G, G_1, ..., G_K\}$. (HD6) We have for any $G_0 \in \{G, G_1, ..., G_K\}$,

$$0 < H_c(G_0) = \int c(G_0(Z), G_0(y))\gamma(y)dG_0(y) < +\infty,$$

$$0 < H_\pi(G_0) = \int \pi(G_0(Z), G_0(y))e(y)dG_0(y) < +\infty,$$

$$J(G_0) = H_c(G_0)/H_\pi(G_0),$$

$$g_0 = H_\pi^{-1}(G_0)g_c - H_c(G_0)H_\pi^{-2}(G_0)g_\pi + K(G_0)e(\cdot),$$

with

$$g_c(\cdot) = c(G_0(Z), G_0(\cdot))\gamma(\cdot),$$

$$g_\pi = \pi(G_0(Z), G_0(\cdot))e(\cdot),$$

$$K(G_0) = H_\pi^{-1}(G_0)K_c(G_0) - H_c(G_0)H_\pi^{-2}(G_0)K_\pi(G_0),$$

with

$$K_c(G_0) = \int_0^1 \frac{\partial c}{\partial x}(G_0(Z), s)\gamma(G_0^{-1}(s))ds,$$

$$K_\pi(G_0) = \int_0^1 \frac{\partial \pi}{\partial x}(G_0(Z), s)e(G_0^{-1}(s))ds,$$

and

$$\nu_0 = H_\pi^{-1}(G_0)\nu_{c,0} - H_c(G_0)H_\pi^{-2}(G_0)\nu_{\pi,0},$$

where

$$\nu_{c,0}(y) = \frac{\partial c}{\partial y}(G_0(Z), G_0(y))\gamma(y),$$

$$\nu_{\pi,0}(y) = \frac{\partial \pi}{\partial y}(G_0(Z), G_0(y))e(y).$$

We now put,

$$l_i = (g - g_i)G_i^{-1},$$

$$c_i(t) = (p_i\nu - \nu_i)G^{-1}(t),$$

and let

$$A_1 = \sum_{i=1}^K p_i \mathbb{E}\mathbb{G}^2\left(i, (g-g_i)G_i^{-1}\right),$$

$$A_2 = \sum_{i=1}^K \int_0^1 \int_0^1 (s\wedge t - st)c_i(t)c_i(s)dsdt,$$

$$\begin{aligned} A_3 \ = \ & \sum_{i=1}^K \sum_{h\neq i}^K p_h^2 \int_0^1\int_0^1 \left[G_h(G_i^{-1}(s)) \wedge G_h(G_i^{-1}(t)) - G_h(G_i^{-1}(s))G_h(G_i^{-1}(t))\right] \times \\ & \nu(G_i^{-1}(s))\nu(G_i^{-1}(t))dtds + \\ & \sum_{i\neq j}^K p_i p_j \sum_{h\notin\{i,j\}}^K p_h \int_0^1\int_0^1 \left[G_h(G_i^{-1}(s)))\wedge G_h(G_j^{-1}(t))) - G_h(G_i^{-1}(s)))G_h(G_j^{-1}(t)))\right] \times \\ & \nu(G_i^{-1}(s))\nu(G_j^{-1}(t))dtds, \end{aligned}$$

$$B_1 = \sum p_i^K \int_0^1 \left\{\int_{-\infty}^{G_i^{-1}(s)} (g-g_i)(y)dG_i(y) - s\mathbb{E}(g-g_i)(Y^i)\right\} c_i(s)ds,$$

$$B_2 = \sum_{i\neq j}^K p_i p_j \int_0^1\int_0^1 [s\vee G_i(G_j^{-1}(t)) - sG_i(G_j^{-1}(t))]c_i(s)\nu(G_j^{-1}(t)dsdt,$$

$$B_3 = \sum_{i\neq j}^K p_i p_j \int_0^1 \left\{\int_{-\infty}^{G_j^{-1}(s)} (g-g_i)(y)dG_i(y) - G_i(G_j^{-1}(s))\times\mathbb{E}(g-g_i)(Y^i)\right\}\times\nu(G_j^{-1}(s))ds.$$

Theorem : Let (HD0)-(HD6) hold. Then

$$gd_{n,0}^* = \sqrt{n}(gd_n - gd_0) \rightsquigarrow \mathcal{N}(0, \vartheta_1^2 + \vartheta_3^2),$$

and

$$gd_n^* = \sqrt{n}(gd_n - gd) \rightsquigarrow \mathcal{N}(0, \vartheta_1^2 + \vartheta_2^2),$$

with

$$\vartheta_1^2 = A_1 + A_2 + A_3 + 2(B_1 + B_2 + B_3),$$

and

$$\vartheta_2^2 = \sum_{h=1}^{K} {F_h}^2 p_h - \left(\sum_{h=1}^{K} F_h p_h\right)^2,$$

for

$$F_h = J_h(G_h)/\sqrt{p_h} + \mathbb{E}g(Y^h) + \sum_{i=1}^{K} p_i \mathbb{E}G_h(Y^i)\nu(Y^i),$$

and

$$\vartheta_3^2 = \sum_{h=1}^{K} {M_h}^2 p_h - \left(\sum_{h=1}^{K} M_h p_h\right)^2,$$

for

$$M_h = \mathbb{E}g(Y^h) + \sum_{i=1}^{K} p_i \mathbb{E}G_h(Y^i)\nu(Y^i).$$

The Sen Case

The conditions (HD1), (HD2), (HD3) and (HD4) hold for this measure and we have here $c(x,y) = x-y$ and $\pi(x,y) = y/x$. Further when (HD0), (HD5) and (HD6) are true, the results of Theorem apply with

$$J(G_0) = 2\int_0^{G_0(Z)} \left(1 - \frac{s}{G_0(Z)}\right)\left(\frac{Z - G_0^{-1}(s)}{Z}\right) ds,$$

$$K(G_0) = 2\left(1 - \frac{1}{ZG_0(Z)}\int_0^{G_0(Z)} G_0^{-1}(s)ds\right) + \frac{J(G_0)}{G_0(Z)},$$

$$g_0(y) = 2\left\{\left[\left(1 - \frac{G_0(y)}{G_0(Z)}\right)\left(\frac{Z-y}{Z}\right) - 2\left(\frac{G_0(y)}{G_0(Z)}\right)\left(\frac{J(G_0)}{G_0(Z)}\right)\right] + K(G_0)\right\} 1_{(y \le Z)},$$

and

$$\nu_0(y) = \left[-\frac{2}{G_0(Z)}\left(\frac{Z-y}{Z}\right) - 4\frac{J(G_0)}{G_0(Z)^2}\right] 1_{(y \le Z)}.$$

Datadriven applications

In this note, we particularize our results to the Sen [5] as a example. The same can be done for the Shorrocks measure [10] or the kakwani family [6] of indices. We consider the Senegalese database ESAM 1 of 1996 which includes 3278 households. We first consider the geographical decomposition into the areas Kolda (ko), Dakar (Capital of Sénégal, dak), Ziguinchor (zi), Diourbel (di), Saint-Louis (stl), Tamba (ta), Kaolack (ka), Thiès (th), Louga (lg), Fatick (fa) :

o

(Area)	Senegal	ko	dakar	zi	di	stl	ta	ka	th	lg	fa
(Sen)%	34.71	51.66	22.73	39.13	40.16	37.51	47.47	37.91	41.31	34.53	42.22
(Size)	3278	198	1122	216	231	314	126	316	401	174	180

Computing ϑ_1^2, ϑ_2^2 and ϑ_3^2 with the estimations $p \approx n_i/n$, gives $\vartheta_1^2 + \vartheta_2^2 = 0.093195$, $\vartheta_1^2 + \vartheta_3^2 = 0.093224$ and $gd_n = 1.25450\ 10^{-3}$. This gives the 95%-confidence $J(G) \in [34.7\%, 34.71\%]$, $dg \in [-0.00919, 0.00117]$,

Secondly we consider the decomposition by the Household Chief gender :

Gender	Sénégal	Male	female
Sen Index	34.7 %	35.27 %	32.62 %
size	3278	2559	919

Computing ϑ^2 with the estimations $p \approx n_i/n$, gives $\vartheta_1^2 + \vartheta_2^2 = 1.87$, $\vartheta_1^2 + \vartheta_3^2 = 1.78$, and $gd_n = 1.496 \times 10^{-4}$. This gives the 95%-confidence :$dg \in [-0.00437, 0.0016]$,that $J(G) \in [34.696\%, 34.704\%]$, We get the same conclusion, that is the gap of decomposaiblity is been significanly very low. The sen measure is then pratically decomposable.

Conclusion

We just illustrated how apply our results for the Sen measure and the Senegalase database ESAM I. But It would be more interesting and instructive to conduct large scale datadriven studies for the West African databases for example, for several measures. It would also be interesting to see the influence of the Kakwani parameter k on the results. Theses studies are underway.

References

[1] Lo, G.S., Sall. S. and Seck, C.T.(2006). Une Théorie asymptotique des indicateurs de pauvreté. *Mathematical Reports of Academy of sciences.*, Vol **31**. (2), pp. 45-52.

[2] G. S. Lo, Estimation asymptotique des indices de pauvreté : modélisation continue et analyse spatio-temporelle de la pauvreté au Sénégal. *J. Afric. Commun. Sci. Techn.*, **4**, (2009), pp.341-377.

[3] Sall, S.T. and Lo, G.S., (2010). Uniform Weak Convergence of the time-dependent poverty Measure for Continuous Longitudinal Data. *Braz. J. Probab. Stat.*, 24, (3), 457-467

[4] Gane Samb and Sall, S.T.(2010). Asymptotic Representation Theorems for Poverty Indices. *Afrik. Statist.*, (5), pp. 238-244.

[5] Sen Amartya K.(1976). Poverty: An Ordinal Approach to Measurement. *Econometrica*, 44, 219-231.

[6] Kakwani, N.(1980). On a Class of Poverty Measures. *Econometrica*, 48, 437-446.

[7] S. R. Chakravarty .(1983). A new Poverty Index. *Mathematical Social Science,* 6, 307-313.

[8] Zheng, B.(1997). Aggregate Poverty Measures. *Journal of Economic Surveys*, 11 (2), 123-162.

[9] Foster, J., Greer, J. and Shorrocks, A.(1984). A class of Decomposable Poverty Measures. *Econometrica* 52, 761-766.

[10] A. Shorrocks. (1995). Revisiting the Sen Poverty Index. *Econometrica*, 63, 1225-1230.