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 separated 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 \( E = \{Y_1, ..., Y_n\} \) is a random sample of the random variable \( Y \) defined on a probability space \((\Omega, A, \mathbb{P})\) and drawn from some specific population. Now suppose that this population is divided 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 \( E_i = \{Y_{1,i}, ..., Y_{n,i}\} \) and then put \( J_n(i) = J(Y_{1,i}, ..., Y_{n,i}) \). The statistic \( J_n \) is said to be decomposable only whenever one always has

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

Since this property is very particular 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 \( G_n(y) = \frac{1}{n} \sum_{i=1}^{n} I(Y_i \leq y) \). 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 \( J_n^*(G_i) = J_n^*(Y_{i,1}, ..., Y_{i,n_i^*}) \) the empirical index measured on the \( i^{th} \) subgroup and \( J_n(G) \) the global index. Clearly, decomposability implies for all \( n \geq 1 \),
\[ gd_n = J_n - \frac{1}{n} \sum_{i=1}^{K} n_i^* J_n^*(G_i) \equiv 0. \]
Surely, \( n^* = (n_1^*, ..., n_K^*) \) follows a multinomial 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]),
\[ gd_n = J_n(G) - \frac{1}{n} \sum_{i=1}^{K} n_i^* J_n^*(G_i) \to_p gd = J(G) - \sum_{i=1}^{K} p_i J_i(G_i). \]
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 \( (\Omega_1 \times \Omega_2, A_1 \otimes A_2, P_1 \otimes P_2) \) and put \( P = P_1 \otimes P_2 \). We draw the observations in the following way. In each trial, we draw a subgroup, the \( i^{th} \) subgroup having the occurring probability \( p_i \). And we put. We conclude that \( \{Y_1, ..., Y_n\} \) is an independant sample drawn from \( G(y) \) we can prove 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 \( n^* \equiv (n_1^*, n_2^*, ..., n_K^*) \equiv \pi \) 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 \(i\)th subgroup, and \( G \) the distribution function of the income for the global population. Let also

\[
\gamma(x) = d\left(\frac{Z - x}{Z}\right) 1_{x \leq Z}
\]

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 \( n \rightarrow +\infty, \)

\[
\max_{1 \leq j \leq Q} |A(n, Q)h^{-1}(n, Q)w(\mu_{1}n + \mu_{2}Q - \mu_{3}j + \mu_{4}) - c(Q/n, j/n)| = o_{P}(n^{-1/2}).
\]

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

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

(HD3) The bivariate functions \( c \) and \( \pi \) have continuous partial differentials. (HD4) For a fixed \( x, \) the functions \( y \rightarrow \frac{\partial c}{\partial y}(x, y) \) and \( y \rightarrow \frac{\partial \pi}{\partial y}(x, y) \) 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
\[ \nu_0 = H^{-1}_\pi(G_0)\nu_{c,0} - H_c(G_0)H^{-2}_\pi(G_0)\nu_{\pi,0}, \]
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,
\[ l_i = (g - g_i)G^{-1}_i, \quad c_i(t) = (p_i\nu - \nu_i)G^{-1}(t), \]
and let
\[ A_1 = \sum_{i=1}^{K} p_i\mathbb{E} G^2 (i, (g - g_i)G^{-1}_i), \]
\[ A_2 = \sum_{i=1}^{K} \int_0^1 \int_0^1 (s \wedge t - st)c_i(t)c_i(s)dsdt, \]
\[ A_3 = \sum_{i=1}^{K} \sum_{h \neq i}^{K} p_h^2 \int_0^1 \int_0^1 [G_h(G^{-1}_i(s)) \wedge G_h(G^{-1}_i(t)) - G_h(G^{-1}_i(s))G_h(G^{-1}_i(t))] \times \nu(G^{-1}_i(s))\nu(G^{-1}_i(t))dtds + \]
\[ \sum_{i \neq j} \sum_{h \notin \{i, j\}}^{K} p_h \int_0^1 \int_0^1 [G_h(G^{-1}_i(s)) \wedge G_h(G^{-1}_j(t)) - G_h(G^{-1}_i(s))G_h(G^{-1}_j(t))] \times \nu(G^{-1}_i(s))\nu(G^{-1}_j(t))dtds, \]
\[ B_1 = \sum_{i=1}^{K} p_i^K \int_0^1 \left\{ \int_{-\infty}^{G^{-1}_i(s)} (g - g_i)(y)dG_i(y) - s\mathbb{E}(g - g_i)(Y^i) \right\} c_i(s)ds, \]
\[ B_2 = \sum_{i \neq j} \sum_{h \notin \{i, j\}}^{K} p_i p_j \int_0^1 \int_0^1 [s \vee G_i(G^{-1}_j(t)) - sG_i(G^{-1}_j(t))]c_i(s)\nu(G^{-1}_j(t))dtds, \]
\[ B_3 = \sum_{i \neq j} \sum_{h \notin \{i, j\}}^{K} p_i p_j \int_0^1 \left\{ \int_{-\infty}^{G^{-1}_j(s)} (g - g_i)(y)dG_i(y) - G_i(G^{-1}_j(s)) \times \mathbb{E}(g - g_i)(Y^i) \right\} \times \nu(G^{-1}_j(s))ds. \]

**Theorem:** Let (HD0)-(HD6) hold. Then
\[ gd_{n,0}^* = \sqrt{n}(gd_n - gd_0) \sim N(0, \vartheta_1^2 + \vartheta_3^2), \]
and
\[ gd_n^* = \sqrt{n}(gd_n - gd) \sim N(0, \vartheta_1^2 + \vartheta_2^2), \]
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
\[ F_h = J_h(G_h)/\sqrt{p_h} + E_g(Y^h) + \sum_{i=1}^{K} p_i E G_h(Y^i) \nu(Y^i), \]
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
\[ M_h = E_g(Y^h) + \sum_{i=1}^{K} p_i E G_h(Y^i) \nu(Y^i). \]

**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
\[ J(G_0) = 2 \int_{0}^{G_0(Z)} \left( 1 - \frac{s}{G_0(Z)} \right) \left( \frac{Z - G_0^{-1}(s)}{Z} \right) ds, \]
\[ K(G_0) = 2 \left( 1 - \frac{1}{Z G_0(Z)} \int_{0}^{G_0(Z)} G_0^{-1}(s) ds \right) + \frac{J(G_0)}{G_0(Z)}, \]
\[ g_0(y) = 2 \left\{ \left[ \left( 1 - \frac{G_0(y)}{G_0(Z)} \right) \left( \frac{Z - y}{Z} \right) - 2 \left( \frac{G_0(y)}{G_0(Z)} \right) \left( \frac{J(G_0)}{G_0(Z)} \right) \right] + K(G_0) \right\} 1_{(y \leq Z)}, \]
and
\[ \nu_0(y) = \left[ -\frac{2}{G_0(Z)} \left( \frac{Z - y}{Z} \right) - 4 \frac{J(G_0)}{G_0(Z)^2} \right] 1_{(y \leq Z)}. \]

**Datadriven applications**

In this note, we particularize our results to the Sen [5] as a example. The same can be done for the Shorrock's 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) :
Computing $\vartheta_1^2$, $\vartheta_2^2$ and $\vartheta_3^2$ 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 $gd_n = 1.25450 \times 10^{-3}$. 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:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Sénégé</th>
<th>Male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sen Index</td>
<td>34.7%</td>
<td>35.27%</td>
<td>32.62%</td>
</tr>
<tr>
<td>size</td>
<td>3278</td>
<td>2559</td>
<td>919</td>
</tr>
</tbody>
</table>

Computing $\vartheta^2$ 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 decomposability is been significantly very low. The Sen measure is then practically 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**


