

# Statistical Estimation of Gap of Decomposability of the General Poverty Index

Mohamed Cheikh Haidara<sup>1</sup> & Gane Samb Lo<sup>1,2</sup>

<sup>1</sup> LERSTAD, Université Gaston Berger, Senegal

<sup>2</sup> (LSTA) Université Pierre et Marie Curie (UPMC), France

Correspondence: Gane Samb Lo, Université Gaston Berger, Saint-Louis BP 234, Senegal. Tel: 221-33-961-2340.  
E-mail: gane-samb.lo@ugb.edu.sn, ganesamblo@ufrsat.org

Received: July 25, 2012 Accepted: August 20, 2012 Online Published: October 29, 2012

doi:10.5539/ijsp.v1n2p211 URL: <http://dx.doi.org/10.5539/ijsp.v1n2p211>

## Abstract

For the decomposability property is very a practical one in Welfare analysis, most researchers and users favor decomposable poverty indices such as the Foster-Greer-Thorbeck poverty index. This may lead to neglect the so important weighted indices like the Kakwani and Shorrocks ones which have interesting other properties in Welfare analysis. To face up to this problem, we give in this paper, statistical estimations of the gap of decomposability of a large class of such indices using the General Poverty Index (GPI) and of a new asymptotic representation Theorem for it, in terms of functional empirical processes theory. The results then enable independent handling of targeted groups and next global reporting with significant confidence intervals. Data-driven examples are given with real data.

**Keywords:** functional empirical processes, asymptotic normality, statistical estimation, poverty indices, decomposability, default of decomposability

## 1. 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. Suppose that we have some statistic of the functional form  $J_n = J(Y_1, \dots, Y_n)$  where  $\mathcal{E} = \{Y_1, \dots, Y_n\}$  is a 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 divided into  $K$  subgroups  $S_1, \dots, S_K$  and let us, for each  $i \in \{1, \dots, K\}$ , denote the subset of the random sample  $\{Y_1, \dots, Y_n\}$  coming from  $S_i$  by  $\mathcal{E}_i = \{Y_{i,1}, \dots, Y_{i,n_i}\}$  and then put  $J_{n_i}(i) = J(Y_{i,1}, \dots, Y_{i,n_i})$ . The statistic  $J_n$  is said to be decomposable whenever one always has

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

whatever may be the way in which  $\mathcal{E}$  is partitioned into the  $\mathcal{E}_i$ 's ( $i = 1, \dots, K$ ). This property is a very practical one when dealing with the poverty measures or welfare measures in general for the following reason. If we are willing to monitor the poverty situation, it may be very useful to target some sensitive areas or subgroups. By dividing the population into targeted groups, and estimating the poverty intensity by  $J_{n_i}(i)$  (resp. variation of poverty by  $\Delta J_{n_i}(i)$ ) in each group, one would be able to report the poverty intensity (resp. global poverty variation) by (1.1) (resp.  $\Delta J_n = \frac{1}{n} \sum_{i=1}^K n_i \Delta J_{n_i}(i)$ ), provided that the samples are the same as it is the case in longitudinal data. Thus, decomposability allows an independent handling of poverty for different areas and next an easy reconstruction of the global situation.

Now in the specific case of poverty indices, we mainly have the non-weighted ones and the weighted ones. The statistics in the first case are automatically decomposable and then are mostly preferred by users. However, the weighted measures, which in general are not decomposable, have very interesting properties in poverty analysis. Dismissing them only for non-decomposability would result in a disaster. We tackle this problem in this paper. Indeed, by estimating the following gap of decomposability

$$gd_n = J_n - \frac{1}{n} \sum_{i=1}^K n_i J_{n_i}(i) \quad (1.1)$$

with significant confidence intervals, we would be able to handle separated analyses in the subgroups and report the global case and, at the same time, make benefit of the other properties of such statistics.

The remainder of the paper is organized as follows. In Section 2, we give a brief introduction of the poverty measures and to the General Poverty Index (GPI). In Section 3, we return back to the decomposability problem by describing the drawing scheme under which the results are given. In Section 4, we state the results which are applied to the Senegalese and Mauritanian data in Section 5. The concluding remarks are in Section 7. In Section 8, we show the reader how to apply the methods given here to his own data from scripts posted in our website. Finally, the proofs are given in Section 6 after the bibliography. Although very important, the proofs are not necessary for nonmathematicians who would be able to skip their reading and only focus on methodology questions.

## 2. 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 although 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 (the poverty line).

Consider also a random sample  $Y_1, Y_2, \dots, Y_n$  of size  $n$  of incomes, with empirical distribution function  $G_n(y) = n^{-1} \#\{Y_i \leq y : 1 \leq i \leq n\}$ . The number of poor individuals within the sample is then equal to  $Q_n = nG_n(Z)$ . And, from now on, all the random elements used in the paper are defined on the same probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

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

$$GPI_n = \delta \left( \frac{A(Q_n, n, Z)}{nB(Q_n, n)} \sum_{j=1}^{Q_n} w(\mu_1 n + \mu_2 Q_n - \mu_3 j + \mu_4) d\left(\frac{Z - Y_{j,n}}{Z}\right) \right), \quad (2.1)$$

where  $\mu_1, \mu_2, \mu_3, \mu_4$  are constants. This global form of poverty indices was introduced in Lo, Sall and Seck (2006) (see also Lo, 2009; Sall & Lo, 2010) 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 Sen, 1976) from an axiomatic point of view. A survey of these indices is to be found in Zheng (1997), who also discussed their introduction, from an axiomatic point of view. We will cite a few number of them here just to make clear the minds and prepare the data-driven applications in Section 5.

One may divide the poverty indices into two classes. The first includes the nonweighted ones. The most popular of them is the Foster, Greer and Shorrocks (1984) class which is defined for  $\alpha \geq 0$ , by

$$FGT(\alpha) = \frac{1}{n} \sum_{j=1}^{Q_n} \left( \frac{Z - Y_{j,n}}{Z} \right)^\alpha. \quad (2.2)$$

For  $\alpha = 0$ , (2.2) reduces to  $Q_n/n$ , the headcount of poor individuals. For  $\alpha = 1$  and  $\alpha = 2$ , it is respectively interpreted as the severity of poverty and the depth in poverty. (2.2) is obtained from (2.1) by taking

$$\delta = I_d, \quad w \equiv 1, \quad d(u) = u^\alpha, \quad B(Q_n, n) = Q_n \text{ and } A(Q_n, n, Z) = Q_n.$$

Next, we have for  $\alpha \geq 0$ ,

$$C(\alpha) = \frac{1}{n} \sum_{j=1}^{Q_n} \left( 1 - \left( \frac{Y_{j,n}}{Z} \right)^\alpha \right),$$

the Chakravarty family class of poverty measures is obtained from (2.1) by taking  $Y^\alpha$  and  $Z^\alpha$  as respectively transformed income  $Y$  and threshold  $Z$  and

$$\delta = I_d, \quad w \equiv 1, \quad d(u) = u, \quad B(Q_n, n) = Q_n \text{ and } A(Q_n, n, Z) = Q_n.$$

The statistics in this class are decomposable and are not concerned by the present work.

The second class consists of the weighted indices. We mention here two of its famous members. The Sen index (see Sen, 1976)

$$P_{SE,n} = \frac{2}{n(Q_n + 1)} \sum_{j=1}^{Q_n} (Q_n - j + 1) \left( \frac{Z - Y_{j,n}}{Z} \right) \quad (2.3)$$

(2.3) is obtained from (2.1), by taking

$$d(u) = u, w(u) = u, A(Q_n, n, Z) = Q_n, \\ B(Q_n) = Q_n(Q_n + 1)/2, \mu_1 = 0 \text{ and } \mu_3 = \mu_2 = \mu_4 = 1.$$

The Shorrocks index (see Shorrocks, 1995)

$$P_{SH,n} = \frac{1}{n^2} \sum_{j=1}^{Q_n} (2n - 2j + 1) \left( \frac{Z - Y_{j,n}}{Z} \right) \quad (2.4)$$

is obtained from (2.1) by taking

$$B(Q_n, n) = Q_n(Q_n + 1)/2, A(n, Q_n, Z) = Q_n(Q_n + 1)/2n, \\ \delta = I_d, w(u) \equiv (u), d(u) = u, \mu_1 = 2, \mu_2 = 0, \mu_3 = 2, \mu_4 = 1.$$

Measures (2.3) and (2.4) evaluate the poverty intensity by giving a more important weight on the poorest individuals. This means that a small decrease of the intensity on the poorest household indicates significant improvement in the population.

In the applications, we mainly deal with these two specific measures because of their importance in poverty analysis. Notice that the Thon measure Thon (1979) is different from the Shorrocks one only by their normalization coefficients which are respectively  $n(n + 1)$  and  $n^2$ , so that they have the same asymptotic behavior. Finally, we have the following generalization of the Sen measure given by Kakwani (1980).

$$J_n(k) = \frac{Q}{n \sum_{j=1}^Q j^k} \sum_{j=1}^Q (Q - j + 1)^k d \left( \frac{Z - Y_{j,n}}{Z} \right),$$

where  $k$  is a positive parameter. Notice that  $J_n(1)$  is the Sen measure. Notice also that, under mild conditions,  $J_n$  converges in probability to the Exact General Poverty Index (EGPI) (see Barrett & Donald, 2009; Bishop, Chow, & Zheng, 1995; Bishop, Formby, & Zheng, 1997; Lo, 2009),

$$J(G) = GPI = \int_0^Z L_1(u, G) d \left( \frac{Z - u}{Z} \right) dG(u), \quad (2.5)$$

where  $L_1$  is some weight function depending on the distribution function. This result will be proved again in Theorem 1 below.

### 3. Statistical Decomposability

From now, we suppose that our studied population of households is divided into  $K$  subgroup such that, for each  $i \in \{1, \dots, K\}$ , the probability that a randomly drawn household comes from the  $i^{\text{th}}$  subgroup is  $p_i > 0$ , with  $p_1 + \dots + p_K = 1$ . Let us suppose that we draw a sample of size  $n$  from the population:  $Y_1, \dots, Y_n$  and let us denote those of the  $n_i^*$  observations coming from the  $i^{\text{th}}$  subgroup, ( $1 \leq i \leq K$ ) by  $Y_{i,j}$ ,  $j = 1, \dots, n_i^*$ . Let  $J_{n_i^*}(G_i) = J_{n_i^*}(Y_{i,1}, \dots, Y_{i,n_i^*})$  the empirical index measured on the  $i^{\text{th}}$  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,  $n^* = (n_1^*, \dots, n_K^*)$  follows a multinomial law with parameters  $n$  and  $p = (p_1, \dots, p_K)$ . Since each  $p_i > 0$ , we have that for each  $1 \leq i \leq K$ ,  $n_i^* \rightarrow \infty$  a.s., as  $n \rightarrow \infty$ . We will have by (1.1) and by (2.5),

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

The right member of this equation is 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. As a first result, we get that the decomposability does not, asymptotically at least, matter for a more or less homogeneous population. That is, the decomposability is not only a functional form matter (of the index), but it is also a statistical one since whatever might be the index, decomposability is asymptotically obtained when the subgroups have the same distribution. For example, it has been pointed out in Haidara and Lo (2009), for the Senegalese poverty databases from 1996 to 2001, that the gaps of decomposability were very low for various stratifications (in regions, gender, ethnic groups, etc.). The apparent reason was the homogeneity of the income. Such results are confirmed in Section 5.

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  $i$ th subgroup ( $\mathcal{E}_i$ ) having the occurring probability  $p_i$ . And we put

$$\pi_{i,j}(\omega_1) = \mathbb{I}_{(\text{the } i^{\text{th}} \text{ subgroup is drawn at the } j^{\text{th}} \text{ trial})}(\omega_1),$$

$1 \leq i \leq K, 1 \leq j \leq n$ . Now, given that the  $i^{\text{th}}$  subgroup is drawn at the  $j^{\text{th}}$  trial, we pick one individual in this subgroup and observe its income  $Y_j(\omega_1, \omega_2)$ . We then have the observations

$$\{Y_j(\omega_1, \omega_2), 1 \leq j \leq n\}.$$

We have these simple facts. First, for  $1 \leq i \leq K$ ,

$$n_i^* = \sum_{j=1}^n \pi_{i,j}. \quad (3.1)$$

Secondly, the distribution of  $Y_j$  given  $(\pi_{i,j} = 1)$ , is  $G_i$ , that is

$$\mathbb{P}(Y_j \leq y / \pi_{i,j} = 1) = G_i(y).$$

Next

$$\forall (y \in \mathbb{R}), \mathbb{P}(Y_j \leq y) = \sum_{i=1}^K \mathbb{P}(\pi_{i,j} = 1) \mathbb{P}(Y_j \leq y / \pi_{i,j} = 1) = \sum_{i=1}^K p_i G_i(y).$$

We conclude that  $\{Y_1, \dots, Y_n\}$  is an independent sample drawn from  $G(y) = \sum_{i=1}^K p_i G_i(y)$ , the mixture of the distribution functions of the subgroups incomes. Finally, we readily see that conditionally on  $n^* \equiv (n_1^*, n_2^*, \dots, n_K^*) = (n_1, n_2, \dots, n_K) \equiv \bar{n}$  with  $n_1 + n_2 + \dots + n_K = n$ ,  $\{Y_{i,j}, 1 \leq j \leq n_i^*\}$  are independent random variables with distribution function  $G_i$ .

#### 4. Our Results

The results stated here hold for a very large class of poverty measures summarized in the GPI. This is why we need the representation Theorem of the GPI in Lo and Sall (2010). In fact, we do not need here the complete form of Lo and Sall (2010), but a special case of it, based on the assumptions described below. For that, suppose that  $G_i$  ( $1 \leq i \leq K$ ), is the distribution function of the income for the  $i$ th subgroup, and  $G$  is the distribution function of the income for the global population. Let also  $\gamma(x) = d\left(\frac{Z-x}{Z}\right) \mathbb{I}_{(x \leq Z)}$  and  $e(x) = \mathbb{I}_{(x \leq Z)}$ . The following assumptions are required.

(HD0)  $G_0(Z) \in ]0, 1[$  for  $G_0 \in \{G, G_1, \dots, G_K\}$ .

(HD1) There exist 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  $n \rightarrow +\infty$ ,

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

(HD2) For the function  $h$  found in (HD1), there exists a function  $\pi(s, t)$  of  $(s, t) \in \mathbb{R}^2$  such that as  $n \rightarrow +\infty$ ,

$$\max_{1 \leq j \leq Q} \left| w(j) h^{-1}(n, Q) - \frac{1}{n} \pi(Q/n, j/n) \right| = o_{\mathbb{P}}(n^{-3/2}).$$

(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  $G_0 \in \{G, G_1, \dots, G_K\}$ .

(HD6) We have for any  $G_0 \in \{G, G_1, \dots, G_K\}$

$$0 < H_c(G_0) = \int_{\mathbb{R}} c(G_0(Z), G_0(y))\gamma(y)dG_0(y) < +\infty$$

and

$$0 < H_\pi(G_0) = \int_{\mathbb{R}} \pi(G_0(Z), G_0(y))e(y)dG_0(y) < +\infty$$

We also need the following definitions, for  $G_0 \in \{G, G_1, \dots, G_K\}$ ,

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

$$g_0(\cdot) = H_\pi^{-1}(G_0)g_{c,0}(\cdot) - H_c(G_0)H_\pi^{-2}(G_0)g_{\pi,0}(\cdot) + K(G_0)e(\cdot), \quad (4.1)$$

with

$$g_{c,0}(\cdot) = c(G_0(Z), G_0(\cdot))\gamma(\cdot), g_{\pi,0}(\cdot) = \pi(G_0(Z), G_0(\cdot))e(\cdot), \quad (4.2)$$

$$K(G_0) = H_\pi^{-1}(G_0)K_c(G_0) - H_c(G_0)H_\pi^{-2}(G_0)K_\pi(G_0) \quad (4.3)$$

with

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

$$K_\pi(G_0) = \int_0^1 \frac{\partial \pi}{\partial x}(G_0(Z), s)e(G_0^{-1}(s))ds, \quad (4.4)$$

$$v_0(\cdot) = H_\pi^{-1}(G_0)v_{c,0}(\cdot) - H_c(G_0)H_\pi^{-2}(G_0)v_{\pi,0}(\cdot), \quad (4.5)$$

where

$$v_{c,0}(\cdot) = \frac{\partial c}{\partial y}(G_0(Z), G_0(\cdot))\gamma(\cdot), v_{\pi,0}(\cdot) = \frac{\partial \pi}{\partial y}(G_0(Z), G_0(\cdot))e(\cdot).$$

With the conventions that for  $G_0 = G$ , we denote  $g_0 = g$  and  $v_0 = v$ . For  $G_0 = G_i, 1 \leq i \leq K$ , we put  $g_0 = g_i$  and  $v_0 = v_i$ . Finally define

$$\ell_i(t) = (g - g_i)(G_i^{-1}(t)), c_i(t) = (p_i v - v_i)(G_i^{-1}(t)), 0 \leq t \leq 1. \quad (4.6)$$

We are now able to briefly describe the approximation of Lo and Sall (2010): if  $G_0$  fulfills (HD1),..., (HD6), then as  $n \rightarrow +\infty$ , we have

$$\sqrt{n}(J_n(G_0) - J(G_0)) = \alpha_n(g_0) + \beta_n(v_0) + o_{\mathbb{P}}(1),$$

where

$$\alpha_n(g_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^n g_0(G_0(V_j)) - \mathbb{E}g_0(G_0(V_j))$$

is the functional empirical process and

$$\beta_n(v_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{G_n(V_j) - G_0(V_j)\} v_0(V_j) \quad (4.7)$$

is a *residual* stochastic process introduced in Lo and Sall (2010) and widely studied in Lo (2010), where  $G_n$  is the empirical distribution function associated with  $\{V_1, \dots, V_n\}$  sampled from  $G_0$ .

Finally, we introduce these constants of whom the variances of our theorem are based on:

$$A_1 = \sum_{i=1}^K p_i \left\{ \int_0^{G_i(Z)} (\bar{g} - \bar{g}_i)^2 (G_i^{-1}(t)) dt - \left( \int_0^{G_i(Z)} (\bar{g} - \bar{g}_i) (G_i^{-1}(t)) dt \right)^2 \right\},$$

$$\begin{aligned}
 A_2 &= \sum_i^K p_i \int_0^{G_i(Z)} \int_0^{G_i(Z)} (s \wedge t - st)(p_i \bar{v} - \bar{v}_i)(G_i^{-1}(s))(p_i \bar{v} - \bar{v}_i)(G_i^{-1}(t)) ds dt, \\
 A_{31} &= \sum_{i=1}^K p_i^2 \sum_{h \neq i}^K p_h \int_0^{G_i(Z)} \int_0^{G_i(Z)} [G_h(G_i^{-1}(s)) \wedge G_h(G_i^{-1}(t)) - G_h(G_i^{-1}(s))G_h(G_i^{-1}(t))] \bar{v}(G_i^{-1}(s)) \bar{v}(G_i^{-1}(t)) ds dt, \\
 A_{32} &= \sum_{i=1}^K p_i \sum_{j \neq i}^K p_j \sum_{h \notin \{i,j\}}^K p_h \int_0^{G_i(Z)} \int_0^{G_j(Z)} [G_h(G_i^{-1}(s)) \wedge G_h(G_j^{-1}(t)) - G_h(G_i^{-1}(s))G_h(G_j^{-1}(t))] \bar{v}(G_i^{-1}(s)) \bar{v}(G_j^{-1}(t)) ds dt, \\
 B_1 &= \sum_{i=1}^K p_i \int_0^{G_i(Z)} \left\{ \int_0^{s \wedge G_i(Z)} (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt - s \int_0^{G_i(Z)} (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt \right\} (p_i \bar{v} - \bar{v}_i)(G_i^{-1}(s)) ds, \\
 B_2 &= \sum_{j=1}^K p_j \sum_{i \neq j}^K p_i \int_0^{G_i(Z)} \int_0^{G_j(Z)} [s \wedge G_i(G_j^{-1}(t)) - s G_i(G_j^{-1}(t))] \times (p_i \bar{v} - \bar{v}_i)(G_i^{-1}(s)) \bar{v}(G_j^{-1}(t)) ds dt,
 \end{aligned}$$

and

$$B_3 = \sum_{j=1}^K p_j \sum_{i \neq j}^K p_i \int_0^{G_j(Z)} \left\{ \int_0^{G_i(G_j^{-1}(s)) \wedge G_i(Z)} (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt - G_i(G_j^{-1}(s)) \times \int_0^{G_i(Z)} (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt \right\} \bar{v}(G_j^{-1}(s)) ds,$$

where

$$g_0(\cdot) = \bar{g}_0(\cdot) \times e(\cdot) \text{ and } v_0(\cdot) = \bar{v}_0(\cdot) \times e(\cdot),$$

and

$$(g_0, v_0) \in (g, g_1, \dots, g_K) \times (v, v_1, \dots, v_K) \text{ and } i = 1, \dots, K.$$

We are now able to state our main result.

**Theorem 1** Let (HD0)-(HD6) hold. Then  $gd_{n,0}^* = \sqrt{n}(gd_n - gd_0) \rightsquigarrow \mathcal{N}(0, \vartheta_1^2 + \vartheta_2^2)$ , and  $gd_n^* = \sqrt{n}(gd_n - gd) \rightsquigarrow \mathcal{N}(0, \vartheta_1^2 + \vartheta_2^2)$  with

$$\begin{aligned}
 \vartheta_1^2 &= A_1 + A_2 + A_3 + 2(B_1 + B_2 + B_3) \\
 \vartheta_2^2 &= \sum_{h=1}^K F_h^2 p_h - \left( \sum_{h=1}^K F_h p_h \right)^2
 \end{aligned}$$

for  $F_h = \mathbb{E}g(Y^h) - J(G_h) + \sum_{i=1}^K p_i \mathbb{E}G_h(Y^i)v(Y^i)$ , 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  $M_h = \mathbb{E}g(Y^h) + \sum_{i=1}^K p_i \mathbb{E}G_h(Y^i)v(Y^i)$ .

**Remark 1** This clearly makes the so important decomposability requirement less crucial since the default of decomposability may be estimated by confidence intervals based on this theorem, as we showed it in the next section.

### 5. Examples and Applications

#### 5.1 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 1 apply with

$$\begin{aligned}
 J(G_0) &= 2 \int_0^{G_0(Z)} \left( 1 - \frac{s}{G_0(Z)} \right) \left( \frac{Z - G_0^{-1}(s)}{Z} \right) ds, \\
 K(G_0) &= 2 \left( 1 - \frac{1}{Z G_0(Z)} \int_0^{G_0(Z)} G_0^{-1}(s) ds \right) + \frac{J(G_0)}{G_0(Z)}, \\
 g_0(y) &= \left\{ 2 \left[ \left( 1 - \frac{G_0(y)}{G_0(Z)} \right) \left( \frac{Z - y}{Z} \right) - \left( \frac{G_0(y)}{G_0(Z)} \right) \left( \frac{J(G_0)}{G_0(Z)} \right) \right] + K(G_0) \right\} \mathbb{I}_{(y \leq Z)},
 \end{aligned}$$

and

$$v_0(y) = -\frac{2}{G_0(Z)} \left[ \left( \frac{Z-y}{Z} \right) + \frac{J(G_0)}{G_0(Z)} \right] \mathbb{I}_{(y \leq Z)}.$$

### 5.2 Shorrocks' Case

We have the same conclusion of the previous case with  $c(x, y) = 2(1 - y)$ ,  $K(G_0) = 0$ ,

$$J(G_0) = 2 \int_0^{G_0(Z)} (1 - G_0(Z)) \left( \frac{Z - G_0^{-1}(s)}{Z} \right) ds, \quad (5.1)$$

$$g_0(y) = 2(1 - G_0(y)) \left( \frac{Z-y}{Z} \right) \mathbb{I}_{(y \leq Z)},$$

and

$$v_0(y) = -2 \left( \frac{Z-y}{Z} \right) \mathbb{I}_{(y \leq Z)}.$$

### 5.3 Kakwani Case

We also have the same conclusion for the Kakwani measure of parameter  $k \geq 1$  with  $c(x, y) = (x - y)^k$  and  $\pi(x, y) = y^k/x$ ,

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

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

$$g_0(y) = \left\{ (k+1) \left[ \left( 1 - \frac{G_0(y)}{G_0(Z)} \right)^k \left( \frac{Z-y}{Z} \right) - \frac{J(G_0)}{G_0(Z)} \left( \frac{G_0(y)}{G_0(Z)} \right)^k \right] + K(G_0) \right\} \mathbb{I}_{(y \leq Z)},$$

and

$$v_0(y) = -\frac{k(k+1)}{G_0(Z)} \left[ \left( 1 - \frac{G_0(y)}{G_0(Z)} \right)^{k-1} \left( \frac{Z-y}{Z} \right) + \frac{J(G_0)}{G_0(Z)} \left( \frac{G_0(y)}{G_0(Z)} \right)^{k-1} \right] \mathbb{I}_{(y \leq Z)}.$$

### 5.4 Data-driven Applications

In this note, let us focus on the Sen case, which is more tricky than the Shorrocks one. We consider the Senegalese database ESAM 1 of 1996 which includes 3278 households. We first consider the geographical decomposition into the areas, Dakar is the Capital. We have the Sen measure values for the whole Senegal and for its ten sub-areas in Table 1 below.

Let us compute the different variances  $\vartheta_1^2$ ,  $\vartheta_2^2$  and  $\vartheta_3^2$  of Theorem 1 with the empirical estimations  $p_i \approx n_i/n$ . We obtain for the geographical decomposability in Senegal:  $\vartheta_1^2 + \vartheta_2^2 = 0,093195$ ;  $\vartheta_1^2 + \vartheta_3^2 = 0,093224$  and  $gd_n = 1,25450 \cdot 10^{-3}$ . This gives the 95%-confidence:

$$dg \in [-0.00919\%; 0.00117\%],$$

that is

$$J(G) \in [34.7\%; 34.71\%],$$

We remark the very accurate estimation of the Sen index for the whole country of Senegal which makes us tell that this index is practically decomposable in this empirical case. We have already explained that decomposability does not matter when the distribution is uniform in the population. It happens that earlier works show that the senegalese data are well fitted by the lognormal or the Singh-Maddala model for each area with very similar parameters. Now for a decomposition with respect to the household chief gender, we get the sen measure values in Table 2.

We see here that  $\vartheta_1^2 + \vartheta_2^2 = 1,87$ ,  $\vartheta_1^2 + \vartheta_3^2 = 1,78$ ,  $gd_n = 1,496 \times 10^{-4}$  and this 95%-confidence:

$$dg \in [-0.00437\%; 0.0016\%],$$

that is

$$J(G) \in [34.696\%; 34.704\%],$$

We get the same conclusion that the gap of decomposability is significantly very low.

We have for the Mauritanian data (EPCV 2004) the following geographical and gender decomposability estimates. From the Sen measures values given in Table 3 for the whole country and its thirteen sub-areas, computation of the variances gives:

$\vartheta_1^2 + \vartheta_2^2 = 2,31 \times 10^{-2}$ ,  $\vartheta_1^2 + \vartheta_3^2 = 2,31 \times 10^{-2}$  and  $gd_n = 3,25 \times 10^{-4}$ . This gives the 95%-confidence:

$$dg \in [-0.00276\%; 0.00341\%]$$

For a stratification with respect to the gender of the chief household, from the values of Table 4, we have:

$\vartheta_1^2 + \vartheta_2^2 = 5,12 \times 10^{-2}$ ,  $\vartheta_1^2 + \vartheta_3^2 = 5,12 \times 10^{-2}$ ,  $gd_n = 3,99 \times 10^{-5}$  and the 95%-confidence:

$$dg \in [-0.00454\%; 0.00462\%],$$

Our general conclusion is that for all these cases, the sen measure is almost decomposable. But, this does not really matter. The important result is that we are able to have an accurate estimation of the gap of decomposability.

### 5.5 Tables

Table 1. Sen measures values for the ten sub-areas for SENEGAL

Area	Senegal	Kolda	Dakar	Diourbel	Saint-Louis	Louga
Sen Index	34.71%	51.66%	22.73%	40.16%	37.51%	34.53%
Size	3278	198	1122	231	314	174

  

Area	Tambacounda	Kaolack	Thies	Fatick	Ziguinchor
Sen Index	47.47%	37.91%	41.31%	42.22%	39.13%
Size	126	316	401	180	216

Table 2. Sen measures values for the genders in SENEGAL

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

Table 3. Sen measures values for the thirteen sub-areas for Mauritania

Area	Mauritanie	Hodh Charghy	Hodh Gharby	Guidimagha
Sen Index	7.5%	6.73%	7.59%	10.89%
Size	9360	1211	469	234

  

Area	Adrar	Nouadhibou	Tagant	Tiris Zemmour	Assaba
Sen Index	5.5%	0.83%	13.34%	2.78%	6.49%
Size	568	585	490	284	514

  

Area	Brakna	Trarza	Inchiri	Gorgol	Nouakchott
Sen Index	11.57%	9.12%	4.89%	12.43%	3.49%
Size	1190	1217	205	796	1597

Table 4. Sen measures values for the genders in Mauritania

Gender	Mauritania	Male	female
Sen Index	7.5 %	7.46 %	7.64 %
size	9360	7513	1847



## 6. Appendix

We would like to provide indications to the reader for using the techniques developed here. We have a zipped file at:

[http : //www/ufrsat.org/lerstad/sen – decomposabilite.rar](http://www/ufrsat.org/lerstad/sen-decomposabilite.rar)

It includes the executable *sendecomp.exe* file which performs the computation of *dg*. Here is how to proceed:

- (i) Download the zipped file and unzip him in a folder named *sen-decomposabilite* directly placed C folder.
- (ii) Upload in the *sen-decomposabilite* folder the following user files: The income file *dep.txt* of size  $n$  at most equal to 10000, the equivalent-adult file *eq.txt* of the same size  $n$  and finally the labels file *labels.txt* including the names of the different strates. If the income file is already scaled for individuals, use an *eq.txt* file of size  $n$  having unity at each line. The number of labels is at most equal to 15. They must be enumerated from 1 to  $KK < 16$ .
- (iii) Execute *sendecomp.exe* by clicking on it. The user is prompted to provide the income file name, the equivalent-adult file name and the labels file name **without** the suffixs *.txt*.
- (v) The package provides the *sen* measures value for the differents strates and report the gap of decomposability value.
- (vi) or the user's practice we provided in the zipped folder the following income variables (*depm.txt*), equivalent-adult variable (*eom.txt*) and labels (here areas) file named after *regm.txt*. Practice with a poverty line equal to 94300.
- (vi) If the data size exceeds  $n = 10000$  or the strates number exceeds  $KK = 15$ , the user is free to write to the authors and adapted packages will be provided.

Finally for those who want to set their own packages in some langage, we provide a Visual Basic module including the main program and the subroutines.

## 7. Conclusion

We just illustrated how apply our results for the Sen Measure and the Senegalese database ESAM I and the Mauritanian EPCV 2004 data. But It would be more interesting and instructive to conduct large scale data-driven 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. This study is underway.

## 8. Proofs

To begin, we need more notations to describe the representation result of Lo and Sall (2010), in an appropriate way to our proof. Let  $G_0 \in \{G, G_1, \dots, G_K\}$  and let a sample of incomes  $\{V_1, \dots, V_m\}$  from  $G_0$ . Let  $\alpha_{G_0,m}$  the uniform empirical functional process based on

$$\{G_0(V_1), \dots, G_0(V_m)\},$$

defined by

$$\alpha_{G_0,m}(g_0) = \frac{1}{\sqrt{m}} \sum_{j=1}^m g_0(G_0(V_j) - \mathbb{E}g_0(G_0(V_j))),$$

and define an other empirical process, called here residual empirical process,

$$\beta_{G_0,m}(v_0) = \frac{1}{\sqrt{m}} \sum_{j=1}^m \{G_{G_0,m}(V_j) - G_0(V_j)\} v_0(V_j), \quad (8.1)$$

where  $G_{G_0,m}$  is the empirical distribution function associated with  $\{V_1, \dots, V_m\}$ . The representation Theorem of Sall and Lo (2010) establishes under the hypotheses (HD0)-(HD6), for  $J(G_0) = H_c(G_0)/H_\pi(G_0)$ ,

$$\sqrt{m}(J_m(G_0) - J(G_0)) = \alpha_{G_0,m}(g_0) + \beta_{G_0,m}(v_0) + o_{\mathbb{P}}(1)$$

as  $m \rightarrow \infty$ , where  $g_0$  and  $v_0$  are described in (4.1) and (4.5).

Before going any further, we should precise the notations for the global population and the subgroups. For  $G = G_0$ , we drop the subscript  $G_0$  so that  $\alpha_n, \beta_n, G_n, J_n$  are respectively the empirical, the residual empirical process (8.1), the empirical distribution function and the GPI based on the sample  $Y_1, \dots, Y_n$ , and  $J = J(G) = H_c(G)/H_\pi(G)$ . As well the functions  $g_0$  and  $v_0$  are denoted as  $g$  and  $v$  for  $G = G_0$ . For  $G = G_i, 1 \leq i \leq K$ , we use the subscript  $i$  so

that  $\alpha_{i,n_i^*}, \beta_{i,n_i^*}, G_{i,n_i^*}, J_{i,n_i^*}$  will respectively denote the empirical, the residual empirical process (8.1), the empirical distribution function and the GPI based on the sample  $Y_{i,1}, \dots, Y_{i,n_i^*}$ , and  $J_i(G_i) = H_c(G_i)/H_\pi(G_i)$ , accordingly to the notations of Section 4, and the functions  $g_0$  and  $v_0$  are denoted as  $g_i$  and  $v_i$  in this case. But sometimes we may feel the notations so heavy and then lessen them. For example, we only put  $J_i(G_i) = J(G_i)$  and  $J_{i,n_i^*}(G_i) = J_{n_i^*}(G_i)$ ,  $i \in \{1, \dots, K\}$ .

To begin the proof, we remark that  $n^*(\omega_1) = (n_1^*(\omega_1), \dots, n_K^*(\omega_1)) \rightarrow_{\mathbb{P}_1} \{+\infty\}^K$  as  $n = n_1^*(\omega_1) + \dots + n_K^*(\omega_1) \rightarrow \infty$ . We then get

$$\sqrt{n}(J_n(G) - J(G)) = \alpha_n(g) + \beta_n(v) + o_{\mathbb{P}}(1) := \gamma_n + o_{\mathbb{P}}(1) \quad (8.2)$$

and for any  $1 \leq i \leq K$ ,

$$\sqrt{n_i^*}(J_{n_i^*}(G_i) - J(G_i)) = \alpha_{i,n_i^*}(g_i) + \beta_{i,n_i^*}(v_i) + o_{\mathbb{P}}(1) := \gamma_{i,n_i^*} + o_{\mathbb{P}}(1) \quad (8.3)$$

Now we use the intermediate centering coefficient

$$gd_{0,n} = J(G) - \sum_{i=1}^K \frac{n_i^*}{n} J(G_i).$$

to get from (8.2) and (8.3)

$$\left| \sqrt{n}(gd_n - gd_{0,n}) - \left\{ \gamma_n - \sum_{j=1}^K \left( \frac{n_j^*}{n} \right)^{1/2} \gamma_{j,n_j^*} \right\} \right| (\omega_1, \omega_2) \rightarrow_{\mathbb{P}_1 \otimes \mathbb{P}_2} 0, \quad (8.4)$$

as  $n \rightarrow \infty$ . Then, we have

$$S_n^* = \gamma_n - \sum_{j=1}^K \left( \frac{n_j^*}{n} \right)^{1/2} \gamma_{j,n_j^*} = \alpha_n(g) - \sum_{j=1}^K \left( \frac{n_j^*}{n} \right)^{1/2} \alpha_{j,n_j^*}(g_j) + \beta_n(v) - \sum_{j=1}^K \left( \frac{n_j^*}{n} \right)^{1/2} \beta_{j,n_j^*}(v_j).$$

Remark that

$$\begin{aligned} \alpha_n(g) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n (g(Y_j) - \mathbb{E}g(Y)) = \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n g(Y_j) - \mathbb{E}g(Y) \right) \\ &=: \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n g(Y_j) - \sum_{i=1}^K \frac{n_i}{n} \mathbb{E}g(Y^i) \right) + D^*(n, 1) \end{aligned}$$

with

$$D(n, 1) = \sum_{i=1}^K \frac{n_i - np_i}{\sqrt{np_i}} \sqrt{p_i} \mathbb{E}g(Y^i),$$

and

$$D^*(n, 1) = \sum_{i=1}^K \frac{n_i^* - np_i}{\sqrt{np_i}} \mathbb{E}g(Y^i) \sqrt{p_i}.$$

This leads to

$$S_n^* = \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n g(Y_j) - \sum_{i=1}^K \frac{n_i}{n} \mathbb{E}g(Y^i) \right) - \sum_{j=1}^K \left( \frac{n_j^*}{n} \right)^{1/2} \alpha_{j,n_j^*}(g_j) + \beta_n(v) - \sum_{j=1}^K \left( \frac{n_j^*}{n} \right)^{1/2} \beta_{j,n_j^*}(v_j) + D^*(n, 1).$$

Now, by denoting

$$C^*(n, 1) = \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n g(Y_j) - \sum_{i=1}^K \frac{n_i}{n} \mathbb{E}g(Y^i) \right) - \sum_{i=1}^K \left( \frac{n_i^*}{n} \right)^{1/2} \alpha_{i,n_i^*}(g_i),$$

one has

$$C^*(n, 1) = \sum_{i=1}^K \left( \frac{n_i^*}{n} \right)^{1/2} \left\{ \frac{1}{\sqrt{n_i^*}} \sum_{j=1}^{n_i^*} \left\{ (g - g_i)(Y_{ij}) - \mathbb{E}(g - g_i)(Y^i) \right\} \right\}. \quad (8.5)$$

We get

$$S_n^* = C^*(n, 1) + D^*(n, 1) + \beta_n(v) - \sum_{j=1}^K \left( \frac{n_j^*}{n} \right)^{1/2} \beta_{j,n_j^*}(v_j). \quad (8.6)$$

Further one has

$$\sum_{j=1}^K \binom{n_i^*}{n} \beta_{i,n_i^*}^*(v_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^K \sum_{j=1}^{n_i^*} [G_{i,n_i^*}(Y_{ij}) - G_i(Y_{ij})] v_i(Y_{ij}). \quad (8.7)$$

But

$$G(Y_{ij}) = \sum_{h=1}^K p_h G_h(Y_{ij}),$$

and for  $x \in \mathbb{R}$ ,

$$G_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(Y_j \leq x)} = \frac{1}{n} \sum_{i=1}^K \sum_{j=1}^{n_i^*} \mathbb{I}_{(Y_{ij} \leq x)} = \sum_{i=1}^K \left( \frac{n_i^*}{n} \right) \frac{1}{n_i^*} \sum_{j=1}^{n_i^*} \mathbb{I}_{(Y_{ij} \leq x)} = \sum_{i=1}^K \frac{n_i^*}{n} G_{i,n_i^*}(x).$$

Thus

$$\beta_n(v) = \frac{1}{\sqrt{n}} \sum_{i=1}^K \sum_{j=1}^{n_i^*} \left[ \sum_{h=1}^K \binom{n_h^*}{n} G_{h,n_h^*}(Y_{ij}) - p_h G_h(Y_{ij}) \right] v(Y_{ij}).$$

From this, we put and subtract  $\sum_{h=1}^K \left( \frac{n_h^*}{n} \right) G_h(Y_{ij})$  to have

$$\begin{aligned} \beta_n(v) &= \frac{1}{\sqrt{n}} \sum_{i=1}^K \sum_{j=1}^{n_i^*} \left[ \sum_{h=1}^K \binom{n_h^*}{n} G_{h,n_h^*}(Y_{ij}) - \sum_{h=1}^K \binom{n_h^*}{n} G_h(Y_{ij}) \right] v(Y_{ij}) + \frac{1}{\sqrt{n}} \sum_{i=1}^K \sum_{j=1}^{n_i^*} \left[ \sum_{h=1}^K \left( \frac{n_h^*}{n} - p_h \right) G_h(Y_{ij}) \right] v(Y_{ij}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^K \sum_{j=1}^{n_i^*} \sum_{h=1}^K \binom{n_h^*}{n} \{G_{h,n_h^*}(Y_{ij}) - G_h(Y_{ij})\} v(Y_{ij}) + \frac{1}{\sqrt{n}} \sum_{i=1}^K \sum_{j=1}^{n_i^*} \left[ \sum_{h=1}^K \left( \frac{n_h^*}{n} - p_h \right) G_h(Y_{ij}) \right] v(Y_{ij}). \end{aligned} \quad (8.8)$$

Now we put together (8.7) and (8.8), while separating the two cases  $h = i$  and  $h \neq i$  in (8.8) to get

$$\begin{aligned} &\beta_n(v) - \sum_{j=1}^K \binom{n_i^*}{n}^{1/2} \beta_{i,n_i^*}^*(v_i) \\ &= \sum_{i=1}^K \binom{n_i^*}{n}^{1/2} \left\{ \frac{1}{\sqrt{n_i^*}} \sum_{j=1}^{n_i^*} \{G_{i,n_i^*}(Y_{ij}) - G_i(Y_{ij})\} \left( \frac{n_i^*}{n} v - v_i \right) (Y_{ij}) \right\} \\ &\quad + \sum_{i=1}^K \binom{n_i^*}{n}^{1/2} \sum_{h \neq i} \frac{n_h^*}{n} \left\{ \frac{1}{\sqrt{n_i^*}} \sum_{j=1}^{n_i^*} \{G_{h,n_h^*}(Y_{ij}) - G_h(Y_{ij})\} v(Y_{ij}) \right\} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^K \sum_{j=1}^{n_i^*} \left[ \sum_{h=1}^K \left( \frac{n_h^*}{n} - p_h \right) G_h(Y_{ij}) \right] v(Y_{ij}) \\ &=: C^*(n, 2) + C^*(n, 3) + D^*(n, 2), \end{aligned} \quad (8.9)$$

with

$$C^*(n, 2) = \sum_{i=1}^K \binom{n_i^*}{n}^{1/2} \left\{ \frac{1}{\sqrt{n_i^*}} \sum_{j=1}^{n_i^*} \{G_{i,n_i^*}(Y_{ij}) - G_i(Y_{ij})\} \left( \frac{n_i^*}{n} v - v_i \right) (Y_{ij}) \right\}, \quad (8.10)$$

and

$$C^*(n, 3) = \sum_{i=1}^K \binom{n_i^*}{n}^{1/2} \sum_{h \neq i} \frac{n_h^*}{n} \left\{ \frac{1}{\sqrt{n_i^*}} \sum_{j=1}^{n_i^*} \{G_{h,n_h^*}(Y_{ij}) - G_h(Y_{ij})\} v(Y_{ij}) \right\}. \quad (8.11)$$

We arrive, by comparing (8.6) and (8.9), at

$$S_n^* = C^*(n, 1) + C^*(n, 2) + C^*(n, 3) + D^*(n, 1) + D^{**}(n, 2). \quad (8.12)$$

Let us have a look at

$$D^{**}(n, 2) = \sqrt{n} \sum_{h=1}^K \left( \frac{n_h^*}{n} - p_h \right) \left\{ \sum_{i=1}^K \left( \frac{n_i^*}{n} \right) \frac{1}{n_i^*} \sum_{j=1}^{n_i^*} G_h(Y_{ij}) \nu(Y_{ij}) \right\}.$$

By the weak law of large numbers

$$\left\{ \sum_{i=1}^K \left( \frac{n_i^*}{n} \right) \frac{1}{n_i^*} \sum_{j=1}^{n_i^*} G_h(Y_{ij}) \nu(Y_{ij}) \right\} \rightarrow_{\mathbb{P}} \sum_{i=1}^K p_i \mathbb{E} G_h(Y^i) \nu(Y^i) = H_h.$$

That is

$$D^{**}(n, 2) = \sum_{h=1}^K \left( \frac{n_h^* - np_h}{\sqrt{np_h}} \right) H_h \sqrt{p_h} + o_{\mathbb{P}}(1) =: D^*(n, 2) + o_{\mathbb{P}}(1).$$

Finally

$$gd_n^* = S_n^* + \sqrt{n}(gd_{0,n} - gd). \tag{8.13}$$

Hence

$$\begin{aligned} gd_n^* &= C^*(n, 1) + C^*(n, 2) + C^*(n, 3) + D^*(n, 1) + D^*(n, 2) - \sum_{i=1}^K \left( \frac{n_i^* - np_i}{\sqrt{np_i}} \right) J_i(G_i) \sqrt{p_i} + o_{\mathbb{P}}(1), \\ &=: C^*(n) + D^*(n) + o_{\mathbb{P}}(1), \end{aligned} \tag{8.14}$$

with

$$C^*(n) = C^*(n, 1) + C^*(n, 2) + C^*(n, 3) \tag{8.15}$$

and

$$\begin{aligned} D^*(n) &= D^*(n, 1) + D^*(n, 2) - \sum_{i=1}^K \left( \frac{n_i^* - np_i}{\sqrt{np_i}} \right) J_i(G_i) \sqrt{p_i} \\ &= \sum_{i=1}^K \left( \frac{n_i^* - np_i}{\sqrt{np_i}} \right) (H_i + \mathbb{E}g(Y^i) - J_i(G_i)) \sqrt{p_i} =: \sum_{i=1}^K \left( \frac{n_i^* - np_i}{\sqrt{np_i}} \right) F_i \sqrt{p_i}. \end{aligned}$$

We have now to prove that  $gd_n^* = \sqrt{n}(gd_n - gd)$  weakly converges to a  $\mathcal{N}(0, \vartheta_1^2 + \vartheta_2^2)$  random variable. For this it suffices, based on (8.14), to prove that  $S_n^{**} = C^*(n) + D^*(n)$  converges to  $\mathcal{N}(0, \vartheta_1^2 + \vartheta_2^2)$ . Now put

$$\mathbb{N}(K) = \{\bar{n} = (n_1, \dots, n_K), n_i \geq 0, n_1 + \dots, n_K = n\}.$$

Since  $n^* = (n_1^*, \dots, n_K^*) \rightarrow_{P_1} \{\infty\}^K$ , we find for a fixed  $\varepsilon > 0$ ,  $K$  positive numbers  $N_i$  ( $1 \leq i \leq K$ ) such that for  $n_i \geq N_i$  ( $1 \leq i \leq K$ ), which implies that  $n \geq N = N_1 + \dots + N_K$ ,

$$\mathbb{P}(\exists(1 \leq i \leq K), n_i^* < N_i) < \varepsilon.$$

Let

$$\mathbb{N}(K, 1) = \mathbb{N}(K) \cap \{\bar{n} = (n_1, \dots, n_K), \exists(1 \leq i \leq K), n_i < N_i\}$$

and  $\mathbb{N}(K, 2) = \mathbb{N}(K) \setminus \mathbb{N}(K, 1)$ . We remark that conditionally on  $(n^* = \bar{n})$ ,  $C^*(n)$  becomes  $C(n)$ , does not depend on  $\omega_1$  and only include the independent random variables  $\{Y_{i,j}, 1 \leq j \leq n_i, 1 \leq i \leq K\}$ . From Lemma 1 below, we have

$$C(n) \rightarrow \mathcal{N}(0, \vartheta_1^2).$$

Also conditionally on  $(n^* = \bar{n})$ ,  $D^*(n)$  becomes  $D(n)$  and we denote it  $D(n)$ . Now for  $h^2 = -1$ ,

$$\begin{aligned} \psi_{S_n^{**}}(t) &= \mathbb{E}(\exp(htS_n^{**})) = \sum_{\bar{n} \in \mathbb{N}(K)} \mathbb{P}(n^* = \bar{n}) \mathbb{E}(\exp(htC^*(n) + htD^*(n)) / (n^* = \bar{n})) \\ &= \sum_{\bar{n} \in \mathbb{N}(K)} \mathbb{P}(n^* = \bar{n}) \mathbb{E}(\exp(htD(n)) \mathbb{E}(\exp(htC^*(n)) / (n^* = \bar{n})). \end{aligned}$$

Recall that, by the classical limiting law of the multinomial  $K$ -vector,

$$D^*(n) \rightarrow D = \sum_{i=1}^K Z_i F_i \sqrt{p_i},$$

where  $(Z_1, \dots, Z_K)^t$  is a Gaussian vector with  $\mathbb{V} ar(Z_i) = 1 - p_i$  and  $Cov(Z_i, Z_j) = -\sqrt{p_i p_j}$ , for  $i \neq j$ . Then

$$D^*(n) \rightarrow \mathcal{N}(0, \vartheta_2^2),$$

with

$$\vartheta_2^2 = \sum_{h=1}^K F_h^2 p_h (1 - p_h) - \sum_{1 \leq h \neq k \leq K} F_h F_k p_h p_k = \sum_{h=1}^K F_h^2 p_h - \left( \sum_{h=1}^K F_h p_h \right)^2.$$

We remark that this is the variance of the function  $F_h$  of  $h \in [1, K]$  with respect to the probability measure  $\sum_{1 \leq h \leq K} p_h \delta_h$ .

Put now

$$\mathbb{N}(K, 1) = \mathbb{N}(K) \cap \{\bar{n} = (n_1, \dots, n_K), \exists(1 \leq i \leq K), n_i < N_i\}$$

and  $\mathbb{N}(K, 2) = \mathbb{N}(K) \setminus \mathbb{N}(K, 1)$ . Then

$$\sum_{\bar{n} \in \mathbb{N}(K)} \exp(htD(n)) \mathbb{P}(n^* = \bar{n}) \mathbb{E}(\exp(htC(n))) = B(n, 1) + B(n, 2)$$

with

$$\begin{aligned} |B(n, 1)| &= \left| \sum_{\bar{n} \in \mathbb{N}(K, 1)} \exp(htD(n)) \mathbb{P}(n^* = \bar{n}) \mathbb{E}(\exp(htC(n))) \right| \\ &\leq \mathbb{P}(\exists(1 \leq i \leq K), n_i^* < N_i) \rightarrow 0, \end{aligned} \quad (8.16)$$

and

$$\begin{aligned} \left| B(n, 2) - \sum_{\bar{n} \in \mathbb{N}(K, 2)} \exp(-(\vartheta_1 t)^2 / 2) \exp(htD(n)) \mathbb{P}(n^* = \bar{n}) \right| \\ \leq \varepsilon \sum_{\bar{n} \in \mathbb{N}(K, 2)} \mathbb{P}(n^* = \bar{n}) \leq \varepsilon. \end{aligned} \quad (8.17)$$

Finally, for

$$B^*(n, 2) = \sum_{\bar{n} \in \mathbb{N}(K, 2)} \exp(-(\vartheta_1 t)^2 / 2) \exp(htD(n)) \mathbb{P}(n^* = \bar{n}), \quad (8.18)$$

we are able to use (8.18) and to get

$$\limsup_{n \rightarrow \infty} \left| B^*(n, 2) - \sum_{\bar{n} \in \mathbb{N}(K)} \exp(htD(n)) \mathbb{P}(n^* = \bar{n}) \mathbb{E}(\exp(-(\vartheta_1 t)^2 / 2)) \right| = 0. \quad (8.19)$$

But

$$\begin{aligned} \mathbb{E} \exp(thD^*(n)) &= \sum_{\bar{n} \in \mathbb{N}(K)} \exp(htD^*(n) / (n^* = \bar{n})) \mathbb{P}(n^* = \bar{n}) \\ &= \sum_{\bar{n} \in \mathbb{N}(K)} \exp(htD(n)) \mathbb{P}(n^* = \bar{n}) \rightarrow \exp(-(\vartheta_2 t)^2 / 2) \end{aligned} \quad (8.20)$$

By putting together the previous formulas, and by letting  $\varepsilon \downarrow 0$ , we arrive at

$$\psi_{d_n^{**}}(t) \rightarrow \exp(-(\vartheta_1^2 + \vartheta_2^2)t^2 / 2).$$

This proves the asymptotic normality of  $dg_n^*$  of the theorem corresponding to  $S_n^{**}$ . That of  $dg_{n,0}^*$  corresponds to  $S_n^*$ . This latter is achieved by omitting the term  $\sqrt{n} \sum_{i=1}^K (\frac{n_i^*}{n} - p_i) J_i(G_i)$  in (8.13). This leads to  $M_h$  obtained from  $F_h$  by dropping  $J_i(G_i)$ . This completes the proofs.

We now prove this lemma used in the proof.

**Lemma 1** Let  $C(n) = C(n, 1) + C(n, 2) + C(n, 3)$ , where the  $C(n, i)$  are respectively defined in (8.5), (8.10) and (8.11) for  $i = 1, 2, 3$ . Then, as  $n \rightarrow +\infty$ ,

$$C(n) \rightsquigarrow \mathcal{N}(0, \vartheta_1^2).$$

*Proof.* Recall that

$$C(n) = C(n, 1) + C(n, 2) + C(n, 3). \tag{8.21}$$

Let for each  $i \in [1, K]$ ,  $\mathbb{G}_{n_i}(i, f)$  be the functional empirical process based on  $\{G_i(Y_{i,j}), 1 \leq j \leq n_i, 1 \leq i \leq K\}$ . We consider the three terms in (8.21), that is the  $C(n, i)$ ,  $1 \leq i \leq 3$ , defined in (8.5), (8.10) and in (8.11), and prove that each of them converges to a random variable  $C(i)$  depending on the limiting Gaussian processes  $\mathbb{G}(i, \cdot)$  of  $\mathbb{G}_{n_i}(i, \cdot)$ . This is enough to prove the asymptotic normality. The variance  $\vartheta_1^2$  will be nothing else but that of  $C(1)+C(2)+C(3)$ . Firstly, we treat  $C(n, 1)$ . Remark that conditionally on  $(n^* = \bar{n})$ , the random sequences  $\{Y_{i,j}, 1 \leq j \leq n_i, 1 \leq i \leq K\}$  are independent and only depend on the  $\omega_2 \in \Omega_2$ . We have

$$\sum_{i=1}^K \left(\frac{n_i}{n}\right)^{1/2} \alpha_{n_i}(g_i) = \frac{1}{\sqrt{n}} \left[ \sum_{i=1}^K \sum_{j=1}^{n_i} g_i(Y_{ij}) - \sum_{i=1}^K n_i \mathbb{E}(g_i(Y^i)) \right] = \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^K \sum_{j=1}^{n_i} g_i(Y_{ij}) - \sum_{i=1}^K \left(\frac{n_i}{n}\right) \mathbb{E}(g_i(Y^i)) \right],$$

and

$$\alpha_n(g, 1) = \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n g(Y_j) - \sum_{i=1}^K \left(\frac{n_i}{n}\right) \mathbb{E}(g(Y^i)) \right) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^K \sum_{j=1}^{n_i} g(Y_{ij}) - \sum_{i=1}^K \left(\frac{n_i}{n}\right) \mathbb{E}(g(Y^i)) \right).$$

Then, by (8.5) and replacing  $n_i^*$  by  $n_i$ ,  $i = 1, \dots, K$ , we get

$$\begin{aligned} C(n, 1) &= \alpha_n(g, 1) - \sum_{i=1}^K \left(\frac{n_i}{n}\right) \alpha_{n_i}(g_i) \\ &= \sum_{i=1}^K \left(\frac{n_i}{n}\right)^{1/2} \left\{ \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \{(g - g_i)(Y_{ij}) - \mathbb{E}(g - g_i)(Y^i)\} \right\}. \end{aligned} \tag{8.22}$$

This implies that

$$C(n, 1) = \sum_{i=1}^K \left(\frac{n_i}{n}\right)^{1/2} \mathbb{G}_{n_i}(i, (g - g_i)G_i^{-1}).$$

We finally have that

$$C(n, 1) \rightarrow C(1) = \sum_{i=1}^K p_i^{1/2} \mathbb{G}(i, (g - g_i)G_i^{-1}).$$

Since the  $\mathbb{G}(i, (g - g_i)G_i^{-1})$  are independent, centered and Gaussian, we get that

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

In the sequel we take

$$g_0(x) = \bar{g}_0(x) \times e(x) \text{ and } v_0(x) = \bar{v}_0(x) \times e(x),$$

and

$$(g_0, v_0) \in (g, g_1, \dots, g_K) \times (v, v_1, \dots, v_K) \text{ and } i = 1, \dots, K.$$

Then we arrive

$$A_1 = \sum_{i=1}^K p_i \left\{ \int_0^{G_i(Z)} (\bar{g} - \bar{g}_i)^2(G_i^{-1}(t)) dt - \left( \int_0^{G_i(Z)} (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt \right)^2 \right\}.$$

Secondly, one has

$$C(n, 2) = \sum_{i=1}^K \left(\frac{n_i}{n}\right)^{1/2} \left\{ \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \{G_{i,n_i}(Y_{ij}) - G_i(Y_{ij})\} \left(\frac{n_i}{n} v - v_i\right)(Y_{ij}) \right\}.$$

We have

$$\frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \{G_{n_i}(Y_{ij}) - G_i(Y_{ij})\} \left(\frac{n_i}{n} v - v_i\right)(Y_{ij}) = \int_0^1 -\varepsilon_{n_i}(i, s)(p_i v - v_i)(G_i^{-1}(s)) ds + o_{\mathbb{P}}(1)$$

$$= \int_0^1 \mathbb{G}_{n_i}(i, s)(p_i v - v_i)(G_i^{-1}(s))ds + o_{\mathbb{P}}(1) \rightarrow \int_0^1 \mathbb{G}(i, s)(p_i v - v_i)(G_i^{-1}(s))ds,$$

and thus

$$C(n, 2) \rightarrow C(2) = \sum_{i=1}^K p_i^{1/2} \int_0^1 \mathbb{G}(i, s)(p_i v - v_i)(G_i^{-1}(s))ds. \quad (8.23)$$

Finally, one has

$$C(n, 3) = \sum_{i=1}^K \left(\frac{n_i}{n}\right)^{1/2} \sum_{h \neq i}^K \frac{n_h}{n} \left\{ \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \{G_{h,n_h}(Y_{ij}) - G_h(Y_{ij})\} v(Y_{ij}) \right\}.$$

But, for each fixed  $i \in \{1, \dots, K\}$ ,

$$\frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \{G_{h,n_h}(Y_{ij}) - G_h(Y_{ij})\} v(Y_{ij}) = \int_0^1 \sqrt{n_i} \{G_{h,n_h}(G_i^{-1}(V_{n_i}(i, s))) - G_h(G_i^{-1}(V_{n_i}(i, s)))\} \times v(G_i^{-1}(V_{n_i}(i, s)))ds.$$

We remember that  $v$  is of the form

$$v(y) = \bar{v}(y)\mathbb{1}_{(y \leq Z)}$$

where  $v_a$  is continuous on compact sets  $[0, L]$ ,  $L > 0$ . Since, as  $n \rightarrow \infty$ ,

$$\sup_{s \in (0,1)} |V_{n_i}(i, s) - s| \rightarrow 0, \text{ a.s.},$$

we see that, for large values of  $n$ , these integrals are performed at most on some interval  $[0, G_i(Z) + \varepsilon]$ , which includes those  $s$  satisfying  $V_{n_i}(i, s) \leq G_i(Z)$ . By the assumptions, the functions  $\bar{v}$  and  $G$  are continuous on such compact sets. Thus

$$\begin{aligned} & \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} [G_{h,n_h}(Y_{ij}) - G_h(Y_{ij})]v(Y_{ij}) \\ &= \sqrt{\frac{n_i}{n_h}} \int_0^1 \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(V_{n_i}(i, s)))) \times v(G_i^{-1}(V_{n_i}(i, s)))ds \\ &= \sqrt{\frac{n_i}{n_h}} \int_0^1 \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(V_{n_i}(i, s)))) \times v(G_i^{-1}(s))ds + o_{\mathbb{P}}(1). \end{aligned}$$

Next

$$= \sqrt{\frac{n_i}{n_h}} \int_0^1 \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(s))) \times v(G_i^{-1}(s))ds + R_n + o_{\mathbb{P}}(1),$$

with

$$R_n = \int_0^1 \left\{ \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(V_{n_i}(i, s)))) - \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(s))) \right\} \times v(G_i^{-1}(s))ds$$

and

$$|R_n| \leq \int_0^{G_i(Z)+\varepsilon} \left| \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(V_{n_i}(i, s)))) - \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(s))) \right| \times v(G_i^{-1}(s))ds.$$

We surely have, by continuity of  $G_h$  on  $(0, G_i^{-1}(G(Z) + \varepsilon))$ ,

$$\sup_{s \leq G_i(Z)+\varepsilon} |G_h(G_i^{-1}(V_{n_i}(i, s))) - G_h(G_i^{-1}(s))| = a_n \rightarrow 0.$$

We obtain here a continuous modulus of the uniform empirical process (see Shorack & wellner, 1986, p. 531) and then

$$\sup_{s \leq G_i(Z)+\varepsilon} \left| \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(V_{n_i}(i, s)))) - \mathbb{G}_{n_h}(h, G_h(G_i^{-1}(s))) \right| = O(\sqrt{-a_n \log a_n}).$$

We finally get

$$R_n = O(\sqrt{-a_n \log a_n}) \int_0^1 v(G_i^{-1}(s))ds \rightarrow 0$$

and we arrive at

$$C(n, 3) \rightarrow C(3) = \sum_{i=1}^K p_i \sum_{h \neq i}^K \sqrt{p_h} \int_0^1 \mathbb{G}(h, G_h(G_i^{-1}(s))) \times v(G_i^{-1}(s)) ds. \tag{8.24}$$

We are now going to compute the variance  $\vartheta_1^2$  based on the independent functional Brownian bridges  $\mathbb{G}(i, \cdot)$  which are limits of the functional empirical process  $\mathbb{G}_n(i, \cdot)$  respectively associated with  $\{G_i(Y_{i,j}), 1 \leq j \leq n_i\}, i = 1, \dots, K$ . Straightforward calculations give what comes. First

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

We denote  $l_i = (g - g_i)G_i^{-1}$  in the sequel for sake of simplicity. Next for

$$C(2) = \sum_{i=1}^K p_i^{1/2} \int_0^1 \mathbb{G}(i, s)(p_i v - v_i)(G_i^{-1}(s)) ds$$

we have

$$\begin{aligned} A_2 &= \mathbb{E}(C^2(2)) = \sum_{i=1}^K p_i \int_0^1 \int_0^1 (s \wedge t - st) c_i(t) c_i(s) ds dt \\ &= \sum_i^K p_i \int_0^{G_i(Z)} \int_0^{G_i(Z)} (s \wedge t - st) (p_i \bar{v} - \bar{v}_i)(G_i^{-1}(s)) (p_i \bar{v} - \bar{v}_i)(G_i^{-1}(t)) ds dt, \end{aligned}$$

where  $c_i(t) = (p_i v - v_i)(G_i^{-1}(t))$ . Now for

$$C(3) = \sum_{i=1}^K p_i \sum_{h \neq i}^K \sqrt{p_h} \int_0^1 \mathbb{G}(h, G_h(G_i^{-1}(s))) \times v(G_i^{-1}(s)) ds,$$

we have

$$A_3 = \mathbb{E}(C^2(3)) = \mathbb{E} \left\{ \sum_{i=1}^K p_i^2 \left( \sum_{h \neq i}^K K_{i,h} \right)^2 + \sum_{i=1}^K \sum_{j \neq i}^K p_i p_j \left( \sum_{h \neq i}^K K_{i,h} \right) \left( \sum_{h' \neq j}^K K_{j,h'} \right) \right\}.$$

Put

$$K_{i,h} = \sqrt{p_h} \int_0^1 \mathbb{G}(h, G_h(G_i^{-1}(s))) \times v(G_i^{-1}(s)) ds,$$

split  $A_3$  into

$$A_{31} = \mathbb{E} \left( \sum_{i=1}^K p_i^2 \left( \sum_{h \neq i}^K K_{i,h} \right)^2 \right)$$

and

$$A_{32} = \mathbb{E} \left( \sum_{i=1}^K \sum_{j \neq i}^K p_i p_j \left( \sum_{h \neq i}^K K_{i,h} \right) \left( \sum_{h' \neq j}^K K_{j,h'} \right) \right).$$

Now by using the independence of the centered stochastic process  $\mathbb{G}(h, \cdot)$  for different values of  $h \in \{1, \dots, K\}$ , one gets

$$A_{31} = \mathbb{E} \left( \sum_{i=1}^K p_i^2 \left( \sum_{h \neq i}^K K_{i,h} \right)^2 \right)$$

and then

$$A_{31} = \sum_{i=1}^K p_i^2 \sum_{h \neq i}^K p_h \int_0^{G_i(Z)} \int_0^{G_i(Z)} [G_h(G_i^{-1}(s)) \wedge G_h(G_i^{-1}(t)) - G_h(G_i^{-1}(s)) G_h(G_i^{-1}(t))] \bar{v}(G_i^{-1}(s)) \bar{v}(G_i^{-1}(t)) ds dt.$$



Next, one has

$$\begin{aligned} A_{32} &= \mathbb{E} \sum_{i=1}^K p_i \sum_{j \neq i}^K p_j \sum_{h \neq i}^K p_h^{1/2} \sum_{h' \neq j}^K p_{h'}^{1/2} \int_0^1 \int_0^1 \mathbb{G}(h, G_h(G_i^{-1}(s))) \mathbb{G}(h', G_{h'}(G_j^{-1}(t))) \nu(G_i^{-1}(s)) \nu(G_j^{-1}(t)) dt ds \\ &= \sum_{i=1}^K p_i \sum_{j \neq i}^K p_j \sum_{h \notin \{i, j\}}^K p_h \int_0^{G_i(Z)} \int_0^{G_j(Z)} [G_h(G_i^{-1}(s)) \wedge G_h(G_j^{-1}(t)) - G_h(G_i^{-1}(s)) G_h(G_j^{-1}(t))] \bar{\nu}(G_i^{-1}(s)) \bar{\nu}(G_j^{-1}(t)) ds dt. \end{aligned}$$

Now we have

$$C(1)C(2) = \left( \sum_{i=1}^K p_i^{1/2} \mathbb{G}(i, \ell_i) \right) \left( \sum_{i=1}^K p_i^{1/2} \int_0^1 \mathbb{G}(i, s) c_i(s) ds \right) = \sum_{i=1}^K p_i^{1/2} \sum_{j=1}^K p_j^{1/2} \int_0^1 \mathbb{G}(i, s) c(s) \mathbb{G}(j, \ell_j) c_i(s) ds.$$

And get

$$\begin{aligned} B_1 &= \mathbb{E} C(1)C(2) = \sum_{i=1}^K p_i \int_0^1 \mathbb{E}(\mathbb{G}(i, s) \mathbb{G}(i, \ell_i) c_i(s)) ds \\ &= \sum_{i=1}^K p_i \int_0^1 \left\{ \int_{-\infty}^{G_i^{-1}(s)} (g - g_i)(y) dG_i(y) - s \mathbb{E}(g - g_i)(Y^i) \right\} c_i(s) ds \\ &= \sum_{i=1}^K p_i \int_0^{G_i(Z)} \left\{ \int_0^{s \wedge G_i(Z)} (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt - s \int_0^1 (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt \right\} (p_i \bar{\nu} - \bar{\nu}_i)(G_i^{-1}(s)) ds. \end{aligned}$$

We have next

$$\begin{aligned} C(2)C(3) &= \left( \sum_{i=1}^K p_i^{1/2} \int_0^1 \mathbb{G}(i, s) c_i(s) ds \right) \times \left( \sum_{i=1}^K p_i \sum_{h \neq i}^K p_h^{1/2} \int_0^1 \mathbb{G}(h, G_h(G_i^{-1}(s))) \times \nu(G_i^{-1}(s)) ds \right) \\ &= \sum_{i=1}^K p_i^{1/2} \sum_{j=1}^K p_j \sum_{h \neq j}^K p_h^{1/2} \int_0^1 \int_0^1 \mathbb{G}(i, s) \mathbb{G}(h, G_h(G_j^{-1}(t))) c_i(s) \nu(G_j^{-1}(t)) ds dt. \end{aligned}$$

It comes that

$$B_2 = \mathbb{E} C(2)C(3) = \sum_{j=1}^K p_j \sum_{i \neq j}^K p_i \int_0^{G_i(Z)} \int_0^{G_j(Z)} [s \wedge G_i(G_j^{-1}(t)) - s G_i(G_j^{-1}(t))] \times (p_i \bar{\nu} - \bar{\nu}_i)(G_i^{-1}(s)) \bar{\nu}(G_j^{-1}(t)) ds dt.$$

Now finally for

$$\begin{aligned} C(1)C(3) &= \left( \sum_{i=1}^K p_i^{1/2} \mathbb{G}(i, \ell_i) \right) \times \left( \sum_{i=1}^K p_i \sum_{h \neq i}^K p_h^{1/2} \int_0^1 \mathbb{G}(h, G_h(G_i^{-1}(s))) \times \nu(G_i^{-1}(s)) ds \right) \\ &= \sum_{i=1}^K p_i^{1/2} \sum_{j=1}^K p_j \sum_{h \neq j}^K p_h^{1/2} \int_0^1 \mathbb{G}(h, G_h(G_j^{-1}(s))) \mathbb{G}(i, \ell_i) \times \nu(G_j^{-1}(s)) ds, \end{aligned}$$

where the  $\ell'_i$ 's are defined in (4.6), we have

$$\begin{aligned} B_3 &= \mathbb{E} C(1)C(3) = \sum_{j=1}^K p_j \sum_{i \neq j}^K p_i \int_0^1 \mathbb{E} \left\{ \mathbb{G}(i, \ell_i) \mathbb{G}(i, G_i(G_j^{-1}(s))) \right\} \times \nu(G_j^{-1}(s)) ds \\ &= \sum_{j=1}^K p_j \sum_{i \neq j}^K p_i \int_0^{G_j(Z)} \left\{ \int_0^{G_i(G_j^{-1}(s)) \wedge G_i(Z)} (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt - G_i(G_j^{-1}(s)) \int_0^{G_i(Z)} (\bar{g} - \bar{g}_i)(G_i^{-1}(t)) dt \right\} \bar{\nu}(G_j^{-1}(s)) ds. \end{aligned}$$

We have now finished the variance computation, that is

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

## References

- Barrett, G., & Donald, S. (2009). Statistical inference with generalized Gini indices of inequality, poverty, and welfare. *J. Bus. Econom. Statist.*, 27(1), 1-17. <http://dx.doi.org/10.1198/jbes.2009.0001>
- Bishop, J. A., Chow, K. V., & Zheng, B. (1995). Statistical Inference and Decomposable Poverty Measures. *Bulletin of Economic Research*, 47, 329-340. <http://dx.doi.org/10.1111/j.1467-8586.1995.tb00619.x>
- Bishop, J. A., Formby, J. P., & Zheng, B. (1997). Statistical Inference and the Sen Index of Poverty. *International Economic Review*, 38(2), 381-387. <http://dx.doi.org/10.2307/2527379>
- Foster, J., Greer, J., & Shorrocks, A. (1984). A class of Decomposable Poverty Measures. *Econometrica*, 52, 761-766. <http://dx.doi.org/10.2307/1913475>
- Haidara, M. C., & Lo, G. S. (2009). Sur la décomposabilité empirique des indicateurs de pauvreté. LERSTAD. [http://www.ganesamblo.net/haidara\\_lo\\_edpi.pdf](http://www.ganesamblo.net/haidara_lo_edpi.pdf)
- Kakwani, N. (1980). On a Class of Poverty Measures. *Econometrica*, 48, 437-446. <http://dx.doi.org/10.2307/1911106>
- Lo, G. S. (2009). Estimation asymptotique des indices de pauvreté: modélisation continue et analyse spatio-temporelle de la pauvreté au Sénégal. *J. Afric. Commun. Sci. Techn.*, 4, 341-377.
- Lo, G. S. (2010). A simple note on some empirical stochastic process as a tool in uniform L-statistics weak laws. *Afrika Statistika*, 5, 437-446.
- Lo, G. S., Sall, S., & Seck, C. T. (2006). Une Théorie asymptotique des indicateurs de pauvreté. *C. R. Math. Acad. Sci. Soc. R. Can.*, 31(2), 45-52.
- Lo, G. S., & Sall, S. T. (2010). Asymptotic Representation Theorems for Poverty Indices. *Afrika Statistika*, 5, 238-244.
- Sall, S. T., & 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. <http://dx.doi.org/10.1214/08-BJPS101>
- Sen, A. K. (1976). Poverty: An Ordinal Approach to Measurement. *Econometrica*, 44, 219-231. <http://dx.doi.org/10.2307/1912718>
- Shorack, G. R., & Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. *wiley-Interscience*, New-York. <http://dx.doi.org/10.1137/1.9780898719017>
- Shorrocks, A. (1995). Revisiting the Sen Poverty Index. *Econometrica*, 63, 1225-1230. <http://dx.doi.org/10.2307/2171728>
- Thon, D. (1979). On Measuring Poverty. *Review of Income and Wealth*, 25, 429-440. <http://dx.doi.org/10.1111/j.1475-4991.1979.tb00117.x>
- Zheng, B. (1997). Aggregate Poverty Measures. *Journal of Economic Surveys*, 11(2), 123-162. <http://dx.doi.org/10.1111/1467-6419.00028>