REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

A multinomial asymptotic representation of Zenga's discrete index, its influence function and data-driven applications

Authors: Tchilabalo Abozou Kpanzou ✉
– Department of Mathematics, University of Kara, Togo
t.kpanzou@univkara.net

Diam Ba
– Department of Applied Mathematics, Gaston Berger University, Sénégal
diamba79@gmail.com

Gandasor Bonyiri Onesiphore Da
– Department of Applied Mathematics, Gaston Berger University, Sénégal
gandasor@gmail.com

Gane Samb Lo
– Department of Applied Mathematics, Gaston Berger University, Sénégal
gane-samb.lo@ugb.edu.sn

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- In this paper, we consider the Zenga index, one of the most recent inequality indices. We keep the finite-valued original form and address the asymptotic theory. The asymptotic normality is established through a multinomial representation. The influence function is also given. The results are simulated and applied to Senegalese income data.

Keywords:

- *Inequality measures; asymptotic behaviour; asymptotic representations; functional empirical process.*

AMS Subject Classification:

✉ Corresponding author

- 62G05; 62G20; 62G07; 91B82; 62P20.

1. INTRODUCTION

Over the years, a number of measures of inequality have been developed. Examples include the generalized entropy, the Atkinson, the Gini, the quintile share ratio and the Zenga measures (see, e.g., Zenga [20]; Zenga [21]; Cowell and Flachaire [3]; Cowell et al. [4]; Kpanzou [11]; Kpanzou [10]; Hulliger and Schoch [9]). Recently, Mergane and Lo [15] gathered a significant number of inequality measures under the name of Theil-like family. Such inequality measures are very important in capturing inequality in income distributions. They also have applications in many other branches of Science, e.g., in ecology (see, e.g., Magurran [14]), sociology (see, e.g., Allison [2]), demography (see, e.g., White [19]) and information science (see, e.g., Rousseau [16]).

The inequality measure of Zenga [22] is one of the most recent ones. It is receiving a considerable attention from researchers for its novelty indeed, but for its interesting properties. Papers dealing with that measure cover theoretical aspects including asymptotic theory and statistical inference (Greselin et al. [7]; Eldin and Marilou [5]) and applied works to income data (Greselin et al. [6], Greselin et al. [8]), etc.

In this paper, we focus on the discrete form of the inequality measure as introduced by Zenga [22]. We justify the asymptotic study of the discrete and finite form by a number of reasons. In some situations, only aggregated data exist. Although this is hardly conceivable today, it is still possible and it is highly probable that the researcher does not have access to the original data and has in hand only data in form of frequency tables. Some other times, frequency tables may be available while the full data is destroyed or lost. Right now, in Gambia, health data collected from the health centers are stored in daily books and the national health direction extracts frequency tables from those books and this type of data is the only one available in their computerized system. So one of the main reasons to work on the finite discrete data is the lack of accessibility to the full data for one reason or another. The second main reason is that an asymptotic theory on such kind of data will give the structure of the limit results with also no severe conditions. By replacing the discrete finite probability law of the aggregated data by a general probability law, we get the precise general asymptotic case. From that simplified study, we see what might be expected in general theory before we proceed with it.

Here, we suppose that the full data have been summarized into a frequencies table as given in Table 1, where each class (c_{i-1}, c_i) is represented by a single point x_i^*, usually taken as the middle of the class, $x_i^* = (c_{i-1}+c_i)/2$ (other possible choices are the mean or the median of the observations falling in the class).

Table 1: Frequencies Tables

classes (c_{i-1}, c_i)	Representatives x_i*	frequencies n_i
(c_0, c_1)	x_1^*	frequencies n_1
(c_1, c_2)	x_2^*	frequencies n_2
$\vdots$	$\vdots$	$\vdots$
(c_{m-1}, c_m)	x_m^*	frequencies n_m
Total	x_i^*	n

We may thus adopt to approximately reconstitute the $n \geq 1$ data as

$$\underbrace{x_1^*}_{n_1 \ times} \cdots \underbrace{x_j^*}_{n_j \ times} \cdots \underbrace{x_m^*}_{n_m \ times}$$

In the sequel, we suppose that the data themselves are discrete and take a predetermined number of m values. First, we will give an asymptotic theory which will be in the form of representation in multinomial laws, instead of a representation in Brownian Bridges as in general. Next, the influence function (IF) will be derived by direct computations and this usually allows to again find the asymptotic variance and sometimes, as in our case, to find a different but equivalent expression of that variance.

The work presented here will be applied to income data available in an aggregated form. At the same time, it serves as a paving way to a more general approach.

Let us suppose that the income variable X is discrete and takes the m $(m > 1)$ ordered values $-\infty = x_0 < x_1 < ... < x_m < x_{m+1} = +\infty$ with the probabilities $p_j > 0$, $j \in \{1, ..., m\}$ with $p_1 + p_2 + ... + p_m = 1$. If the income is continuously observed, we have a sequence of random replications $X_1, X_2, ...$ defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For each $n \geq 1$, the empirical distribution of X on the sample is characterized by the empirical frequencies

$$n_0 = 0, \quad n_j = \#\{h \in \{1, ..., n\}, \quad X_h = x_j\}, \quad j \in \{1, ..., m\},$$

and their normalized and cumulative forms given respectively by

$$f_0 = 0, \ f_j = \frac{n_j}{n}, \ j \in \{1, ..., m\}$$

and

$$n_0^* = f_0^* = 0, n_j^* = \sum_{h=1}^{j} n_h, \ f_j^* = \sum_{h=1}^{j} f_h, \quad j \in \{1, ..., m\},$$

with

$$\sum_{j=1}^{m} n_j = n, \quad \sum_{j=1}^{m} f_j = 1, \quad n_m^* = n, \quad f_m^* = 1.$$

We also define

$$p_0^* = 0, \quad p_j^* = \sum_{h=1}^{j} p_h, \; p_m^* = 1.$$

The empirical and discrete Zenga [22]'s index is given by

$$Z_{d,n} = 1 - \sum_{j=1}^{m-1} f_j \frac{(n_j^*/n)^{-1} \sum_{1 \leq h \leq j} n_h x_h}{(1-(n_j^*/n))^{-1} \sum_{j+1 \leq h \leq m} n_h x_h},$$

which is obtained by summing Formula (3.1) in [22] over $j \in \{1, ..., m\}$ and presented as a synthetic measure of inequality. The empirical cumulative distribution function (cdf) based on the sample of size $n \geq 1$ is

$$F_n(x) = \frac{1}{n} \sum_{h=1}^{m} n_h 1_{[x_h, x_{h+1}[}(x), \quad x \in \mathbb{R}$$

and is the non-parametric estimator of the true cdf

$$F(x) = \sum_{h=1}^{m} p_j 1_{[x_h, x_{h+1}[}(x), \quad x \in \mathbb{R}.$$

We also have the empirical probability generated by the sample, given by

$$P_{X,n}(A) = \frac{1}{n} \sum_{j=1}^{m} 1_A(x_j).$$

We may express $Z_{d,n}$ in terms of the empirical probability measure by

$$Z_{d,n} = 1 - \sum_{j=1}^{m-1} \mathbb{P}_{X,n}(x_j) \frac{\left(\int 1_{]0,x_j]}(t) d\mathbb{P}_{X,n}(t)\right)^{-1} \left(\int t 1_{]0,x_j]}(t) d\mathbb{P}_{X,n}(t)\right)}{\left(\int 1_{]x_j,+\infty[}(t) d\mathbb{P}_{X,n}(t)\right)^{-1} \left(\int t 1_{]x_j,+\infty[}(t) d\mathbb{P}_{X,n}(t)\right)}.$$

Finally by considering the discrete measure $\nu = \sum_{1 \leq j \leq n} \delta_{x_j}$, where δ_{x_j} is the Dirac measure concentrated at x_j with mass one, we may also write

$$Z_{d,n} = 1 - \int \frac{\left(\int 1_{]0,s]}(t) d\mathbb{P}_{X,n}(t)\right)^{-1} \left(\int t 1_{]0,s]}(t) d\mathbb{P}_{X,n}(t)\right)}{\left(\int 1_{]s,+\infty[}(t) d\mathbb{P}_{X,n}(t)\right)^{-1} \left(\int t_{]s,+\infty[}(t) d\mathbb{P}_{X,n}(t)\right)} \mathbb{P}_{X,n}(s) d\nu(s).$$

It is clear, by the convergence in law of a sequence of probability measures $\mathbb{P}_{X,n}$ to $\mathbb{P}_X = \mathbb{P}X^{-1}$ (the probability law of X), that $Z_{d,n}$ converges to

$$Z_d = 1 - \int \frac{\left(\int 1_{]0,s]}(t) d\mathbb{P}_X(t)\right)^{-1} \left(\int t 1_{]0,s]}(t) d\mathbb{P}_X(t)\right)}{\left(\int 1_{]s,+\infty[}(t) d\mathbb{P}_X(t)\right)^{-1} \left(\int t 1_{]s,+\infty[}(t) d\mathbb{P}_X(t)\right)} \mathbb{P}_X(s) d\nu(s).$$

In this simple setting, the convergence is easily justified because of the finiteness of the summations and of the functions. In terms of cdf and mathematical expectation, we have

$$Z_d = 1 - \int_{x_1}^{x_m} \frac{\frac{1}{F(s)} \int_0^s t d\mathbb{P}_X(t)}{\frac{1}{1-F(s)} \int_s^\infty t d\mathbb{P}_X(t)} \mathbb{P}_X(s) d\nu(s).$$

The integral in the last expression should be read as

$$\int_{x_1}^{x_m -} \frac{\frac{1}{F(s)} \int_0^s t d\mathbb{P}_X(t)}{\frac{1}{1-F(s)} \int_s^\infty t d\mathbb{P}_X(t)} \mathbb{P}_X(s) d\nu(s) = \int 1_{[x_1,x_m[}(s) \frac{\frac{1}{F(s)} \int_0^s t d\mathbb{P}_X(t)}{\frac{1}{1-F(s)} \int_s^\infty t d\mathbb{P}_X(t)} \mathbb{P}_X(s) d\nu(s),$$

so that neither $1 - F(s)$ nor $F(s)$ never vanishes on the integration domain.

On one hand, we are going to draw an asymptotic normality theory of $Z_{d,n}$ using the m-multivariate binomial laws. On the other hand, the sensitivity of a statistic $T(F)$ and the impact of extreme observations on it are also two recurrent questions in the research in the field (see Cowell and Flachaire [3]).

In that context, the asymptotic variance of the plug-in estimator $T(F_n)$ of the statistic $T(F)$ is of the form $\sigma^2 = \int L(x, T(F))^2 dF(x)$. From this, we may say that the influence function behaves in nonparametric estimation as the score function does in the parametric setting (See Wasserman [17], Page 19). To define the notion of *IF*, let us consider the contaminated probability law $\mathbb{P}_X^{-(\varepsilon)}$ of $\mathbb{P}_X$ at x with mass $\varepsilon > 0$, defined by

$$\mathbb{P}_X^{(\varepsilon)} = (1-\varepsilon)\mathbb{P}_X + \varepsilon\delta_x \tag{1.1}$$

and a functional of $\mathbb{P}_X$, namely $T(\mathbb{P}_X)$. The influence function of the functional T at x, if it exists, is given by

$$IF(T, x) = \lim_{\varepsilon \to 0} \frac{T(\mathbb{P}_X^{(\varepsilon)}) - T(\mathbb{P}_X)}{\varepsilon}. \tag{1.2}$$

The previous remarks motivate us to derive the *IF* of $Z_d(\mathbb{P}_X)$ and to compare it with the asymptotic variance the Zenga's plug-in estimator.

Before we proceed to our task, we point out that asymptotic normality results for Zenga's index are available in the literature, among them those of Greselin et al. [7] and Eldin and Marilou [5].

Here is how the paper is organized, we give our asymptotic results as described above in Section 2 in Theorems 2.1 and 2.2, and the proofs of these theorems are given in the Appendixes A and B. Section 3 is devoted to simulation studies and data-driven application to Senegalese Data. We conclude in Section 4.

2. ASYMPTOTIC THEORY FOR THE DISCRETE ZENGA MEASURE

2.1. Asymptotic normality

Let us begin with the following reminder. For each $m \geq 1$, the random vector $(n_1, ..., n_m)$ follows a m-dimensional multimonial law of parameters $n \geq 1$

and $p = (p_1, ..., p_m)^t$. In such a case, a classical result of weak convergence (see, e.g., Lo et al. [13]), as $n \to +\infty$, is the following

$$\left(\frac{n_1 - np_1}{\sqrt{np_1}}, \cdots, \frac{n_m - np_m}{\sqrt{np_m}}\right)^t \equiv (N_{1,n}, \cdots, N_{m,n})^t$$
$$\rightsquigarrow Z = (Z_1, \cdots, Z_m)^t \sim \mathcal{N}_m(0, \Sigma),$$

the variance-covariance matrix $\Sigma = (\sigma_{h,k})_{1 \leq h,k \leq m}$ of Z is defined, for $(h,k) \in \{1, ..., m\}^2$, $h \neq k$, by

$$\sigma_{hh} = \mathbb{E}(Z_h^2) = 1 - p_h \textit{ and } \sigma_{hk} = \mathbb{E}(Z_h Z_k) = -\sqrt{p_h p_k}.$$

We invoke the Skorohod-Wichura Theorem (See Wichura [18]) to suppose that Z is defined on the same probability space and that

$$(N_{1,n}, \cdots, N_{m,n})^t \to_{\mathbb{P}} Z, \textit{ as } n \to +\infty.$$

Let us give some notation. Define vectors $C = (c_1, ..., c_m)^t$ such that

$$c_j = \sqrt{p_j} \frac{(1/p_j^*)\mu_{(j)}}{(1/(1 - p_j^*))\mu^{(j)}} 1_{(j \neq m)}, j \in \{1, ..., m\},$$

for $j \in \{1, ..., m-1\}$, $i \in \{1, 2\}$, $D_{j,i} = (d_{j,i,1}, ..., d_{j,i,m})^t$ such that

$$d_{j,1,h} = (x_h \sqrt{p_h})\, 1_{(h \leq j)}, \quad d_{j,2,h} = -\,(x_h \sqrt{p_h})\, 1_{(h \geq j+1)},$$

$$\gamma_{j,1} = p_j \frac{(1/p_j^*)}{(1/(1 - p_j^*))\mu^{(j)}}, \quad \gamma_{j,2} = p_j \frac{(1/p_j^*)}{(1/(1 - p_j^*))} \frac{\mu_{(j)}}{\left(\mu^{(j)}\right)^2}$$

and let $E_j = (e_{j,1}, ..., e_{j,m})^t$ be the vector defined by its components as follows

$$e_{j,h} = -\,(\sqrt{p_h})\, 1_{(h \leq j)}.$$

Finally, let us defined

$$-H = C + \sum_{j=1}^{m-1} \left(\gamma_{j,1} D_{j,1} + \gamma_{j,2} D_{j,2} + \left(p_j^*\right)^{-2} E_j \right).$$

Theorem 2.1. *Under the notation given above, we have, as $n \to +\infty$,*

$$\sqrt{n}(Z_{d,n} - Z_d) \rightsquigarrow \mathcal{N}_m \left(0, H^t \Sigma H\right). \diamond$$

Proof. The proof is given in Appendix A.

2.2. Influence function of Z_d

Theorem 2.2. *Under the notations given above, the influence function of Z_d is given, for $x_1 \leq x \leq x_m$, by*

$$\begin{aligned} IF(Z_d, x) \;=\; & \int \mathbb{P}_X(s) \left(\frac{R_1(s)}{R_2(s)^2(1 - F(s)}) 1_{]s,+\infty]}(x) - \frac{1}{R_2(s)F(s)} 1_{]0,s]}(x) \right) x d\nu \\ & + \int \mathbb{P}_X(s) \left(\frac{R_1(s)}{R_2(s)F(s)} 1_{]0,s]}(x) - \frac{R_1(s)}{R_2(s)(1 - F(s))} 1_{]s,+\infty]}(x) \right) d\nu \\ & - \int \delta_x(s) \frac{R_1(s)}{R_2(s)} d\nu + \int \mathbb{P}_X(s) \frac{R_1(s)}{R_2(s)} d\nu. \end{aligned}$$

Proof. The proof is given in Appendix B.

3. SIMULATION AND DATA-DRIVEN APPLICATIONS

3.1. Simulation study

Quality of the convergence: We choose a Probability distribution of yearly income supported by $m = 10$ points with lower endpoint $x_1 = 4,515,000$ XOF (9,030 nearly) and upper endpoint $x_m = 9,000,000$ XOF (170,490 nearly), characterized as in Table 2.

Table 2: Underlying Probability Law (to be continued)

Values	x_1	x_2	x_3	x_4	...
	4.515×10^6	13.485×10^6	22.455×10^6	31.425×10^6	...
$\mathbb{P}(X = x_i)$	0.05	0.05	0.05	0.05	...

Table 3: Continuation of Table 2

Values	...	x_5	x_6	x_7	...
	...	40.395×10^6	49.365×10^6	58.335×10^6	...
$\mathbb{P}(X = x_i)$	...	0.1	0.1	0.2	...

Table 4: End of Table 2

Values	...	x_8	x_9	x_{10}
	...	67.305×10^6	76.275×10^6	85.245×10^6
$\mathbb{P}(X = x_i)$	...	0.2	0.1	0.1

Table 2 shows the good performance of the nonparametric estimation of the Zenga index for sample size from $n = 100$ to $n = 1500$. Such sizes are comparable with those of sample survey from populations counted in dozen of millions.

Table 5: Mean errors (ERM), Mean Square Errors (MSE), to be continued

Size	100	200	500	...
ERM	3.6×10^{-3}	-5.36×10^{-3}	10^{-3}	...
MSE	6.4×10^{-2}	3.35×10^{-2}	2.49×10^{-2}	...

Figure 2 shows the pretty good asymptotic normality approximation of the centered and normalized empirical Zenga's estimator.

3.2. Data-driven applications

We use the income Data in Senegal (2001-2002) from the database related to ANSD [1]. The incomes are given by households. We should use an adult-equivalence scale to consider to be able to compare households. The notion of adult-equivalence has already been described in Lo [12] and implemented on different sets of data, among them the data just described above. The data are available for the whole country (Senegal) and for the 10 areas given in the following order:

Table 6: End of Table 5

Size	...	750	1000	1500
ERM	...	-8.41×10^{-4}	4.56×10^{-5}	-1.44×10^{-3}
MSE	...	2.16×10^{-2}	1.9×10^{-2}	1.64×10^{-2}

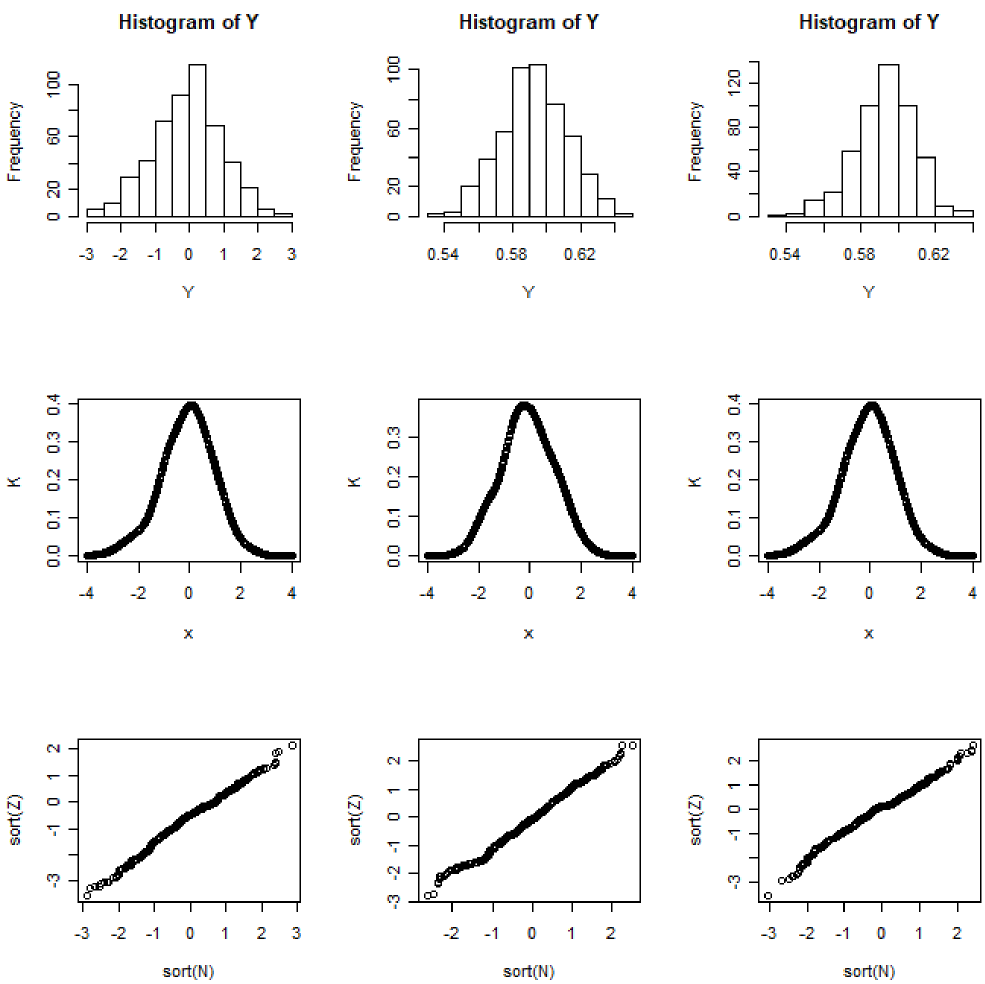

Figure 1: Histograms, Parzen Estimators and QQ-plots for sample sizes 500, 1000 and 1500 from left to right

(OA): Dakar, Diourbel, Fatik, Kaolack, Louga, Saint-Louis, Tamba, Thies, Ziguinchor, Kolda.

Dakar is the most urbanized area of Senegal and includes the capital of the country, also named Dakar. It concentrates almost 23.1% of the population.

Table 7: Zenga and Gini index measures for Senegal's administrative areas (2000), to be continued

Index	Senegal	Dakar	Diourbel	Fatick	Kaolack	Louga	..
Zenga	80.65	93.33	81.34	92.54	81.11	84.00	..
Gini	75.00	80.90	75.26	80.39	75.16	16.25	...

Table 8: End of Table 7

Index	...	Saint-Louis	Tamba	Thies	Ziguinchor	Kolda
Zenga	...	87.69	86.64	82.61	82.11	80.24
Gini	...	78.83	77.26	75.72	75.52	47.86

The Zenga and the Gini indices have been computed for the 11 areas from the aggregate data, and are display in Table 7 (continued in Table 8). Note that these values are given in percentage (%).

Through the values in theses tables, the 11 areas are ordered from the least inequality index to the greatest as follows:

Ordering by Zenga's index: Kolda (1), Senegal (2), Kaolack (3), Diourbel (4), Ziguinchor (5), Thies (6), Louga (7), Tamba (8), Saint-Louis (9), Fatick (10), Dakar (11).

Ordering by Gini's index: Louga (1), Kolda (2), Senegal (3), Kaolack (4), Diourbel (5), Ziguinchor (6), Thies (7), Tamba (8), Saint-Louis (9), Fatick (10), Dakar (11).

These orderings are illustrated in Figure 2.

The most striking fact is that both indices do not order the areas in an exact similar way. The most unfair areas (with the greatest values of the inequality index) are the same with the same ordering, from areas 8 to 11. From areas 1 to 7, the ordering has slightly changed but the case of Louga is remarkable. It is ranked first by Gini and seventh by Zenga.

One may think that the inequality should be greater in urban areas than in rural zones. Indeed we see that with the areas of Thies, Saint-Louis, Dakar. But Factik and Tamba are so urbanized areas. Investigating why the inequality indices (both Zenga and Gini) are high should be investigated in accordance with local realities.

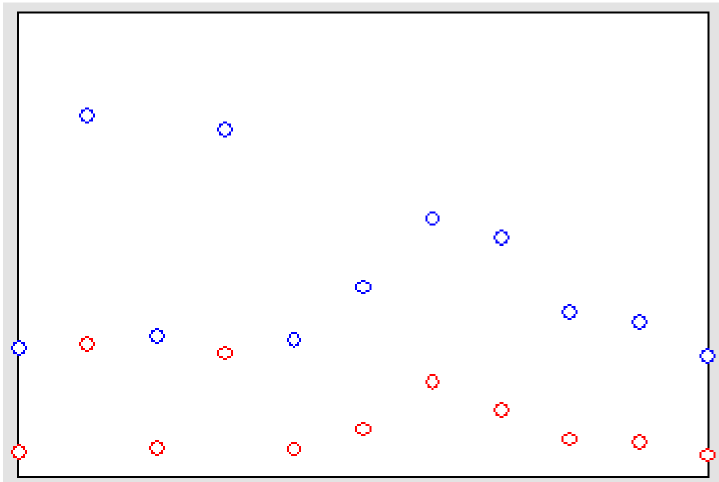

Figure 2: The areas are given in the horizontal line and are ordered according to the ranking **(AO)** above. Blue: Zenga's index. Red: Gini's index

4. CONCLUSION

In this paper, we have considered the discrete Zenga index for which we derived the influence function and studied the asymptotic theory. The asymptotic normality is established through a multinomial representation. Through simulation, we confirmed the asymptotic normality result obtained theoretically. The results are also applied to Senegalese income data.

APPENDIXES

Appendix A. Proof of Theorem 2.1

Let us fix $n \geq 1$. We have

$$Z_{n,d} = 1 - \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{n}{n_j^*} - 1 \right) \frac{\sum_{1 \leq h \leq j} n_h x_h}{\sum_{j+1 \leq h \leq m} n_h x_h}.$$

We define

$$Z_{d,n}^* = \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{n}{n_j^*} - 1 \right) \frac{\sum_{1 \leq h \leq j} n_h x_h}{\sum_{j+1 \leq h \leq m} n_h x_h}.$$

and for $1 \leq j \leq m-1$,

$$\mu_{(j)} = \sum_{h=1}^{j} p_h x_h \ \ and \ \mu^{(j)} = \sum_{h=j+1}^{m} p_h x_h.$$

We have

$$\begin{aligned}
& \frac{\sum_{1\leq h \leq j} n_h x_h}{\sum_{j+1\leq h \leq m} n_h x_h} - \frac{\mu_{(j)}}{\mu^{(j)}} \\
= & \frac{\sum_{1\leq h \leq j} n_h x_h}{\sum_{j+1\leq h \leq m} n_h x_h} - \frac{n\mu_{(j)}}{\sum_{j+1\leq h \leq m} n_h x_h} \\
+ & \frac{n\mu_{(j)}}{\sum_{j+1\leq h \leq m} n_h x_h} - \frac{\mu_{(j)}}{\mu^{(j)}} \\
= & \frac{\sum_{h=1}^{j} x_h N_{h,n}\sqrt{p_h}}{\sqrt{n}\sum_{j+1\leq h \leq m} n_h x_h/n} - \frac{\mu_{(j)} \sum_{h=j+1}^{m} x_h N_{h,n}\sqrt{p_h}}{\sqrt{n}\mu^{(j)} \left(\sum_{j+1\leq h \leq m} n_h x_h/n\right)}.
\end{aligned}$$

Then

$$\begin{aligned}
& Z^*_{d,n} \\
= & \sum_{j=1}^{m-1} \frac{n_j}{n}\left(\frac{n}{n^*_j} - 1\right)\frac{\mu_{(j)}}{\mu^{(j)}} \\
+ & \frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} \frac{n_j}{n}\left(\frac{n}{n^*_j} - 1\right)\left(\frac{\sum_{h=1}^{j} x_h N_{h,n}\sqrt{p_h}}{\sum_{j+1\leq h \leq m} n_h x_h/n} - \frac{\mu_{(j)} \sum_{h=j+1}^{m} x_h N_{h,n}\sqrt{p_h}}{\mu^{(j)} \left(\sum_{j+1\leq h \leq m} n_h x_h/n\right)}\right) \\
= & : Z^*_{d,n}(1) + R_n(1,1).
\end{aligned}$$

We also have

$$\begin{aligned}
\left(\frac{n}{n^*_j} - 1\right) - \left(\frac{1}{p^*_j} - 1\right) &= \left(\frac{n}{n^*_j} - 1\right) - \left(\frac{n}{\sum_{h=1}^{j} np_h} - 1\right) \\
&= -\frac{\sum_{h=1}^{j} n_h - \sum_{h=1}^{j} p_h}{\left(\sum_{h=1}^{j} p_h\right)\left(\sum_{h=1}^{j} n_h\right)} \\
&= -\frac{1}{\sqrt{n}} \frac{\sum_{h=1}^{j} \sqrt{p_h} N_{h,n}}{\left(\sum_{h=1}^{j} p_h\right)\left(\sum_{h=1}^{j} n_h/n\right)}.
\end{aligned}$$

This leads to

$$Z_{d,n}^*(1) = \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{1}{p_j^*} - 1 \right) \frac{\mu_{(j)}}{\mu^{(j)}} - \sum_{j=1}^{m-1} \frac{n_j}{n} \frac{n\sqrt{n} \sum_{h=1}^{j} \sqrt{p_h} N_{h,n}}{\left(\sum_{h=1}^{j} n_h\right) \left(\sum_{h=1}^{j} np_h\right)} \frac{\mu_{(j)}}{\mu^{(j)}}$$
$$=: Z_{d,n}^*(2) + R_n(1,2).$$

Finally, we have

$$\begin{aligned} Z_{d,n}^*(2) &= \sum_{j=1}^{m-1} p_j \left(\frac{1}{p_j^*} - 1 \right) \frac{\mu_{(j)}}{\mu^{(j)}} + \frac{1}{n} \sum_{j=1}^{m-1} \frac{\sqrt{np_j} N_{j,n}}{} \left(\frac{1}{p_j^*} - 1 \right) \frac{\mu_{(j)}}{\mu^{(j)}} \\ &= \sum_{j=1}^{m-1} p_j \left(\frac{1}{p_j^*} - 1 \right) \frac{\mu_{(j)}}{\mu^{(j)}} + \frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} \sqrt{p_j} N_{j,n} \left(\frac{1}{p_j^*} - 1 \right) \frac{\mu_{(j)}}{\mu^{(j)}} \qquad \text{(L2)} \\ &= \sum_{j=1}^{m-1} \frac{(1/p_j^*)\mu_{(j)}}{(1/(1-p_j^*))\mu^{(j)}} + \frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} \sqrt{p_j} N_{j,n} \left(\frac{1}{p_j^*} - 1 \right) \frac{\mu_{(j)}}{\mu^{(j)}} \\ &=: Z_d^* + R_n(3). \end{aligned}$$

It is clear that

$$Z_d = 1 - Z_d^*.$$

We finally get

$$\sqrt{n}(Z_{d,n}^* - Z_d^*) = \sqrt{n} R_n(1) + \sqrt{n} R_n(2) + \sqrt{n} R_n(3).$$

By using the convergence (strong and weak) on binomial probabilities, we get

$$\begin{aligned} &\sqrt{n} R_n(1,1) \\ &= \sum_{j=1}^{m-1} \frac{n_j}{n} \left(\frac{n}{n_j^*} - 1 \right) \left(\frac{\sum_{h=1}^{j} \left(x_h \sqrt{p_h}\right) N_{h,n}}{\sum_{j+1 \le h \le m} n_h x_h / n} - \frac{\mu_{(j)} \sum_{h=j+1}^{m} \left(x_h \sqrt{p_h}\right) N_{h,n}}{\mu^{(j)} \left(\sum_{j+1 \le h \le m} n_h x_h / n\right)} \right) \\ &\to_{\mathbb{P}} \sum_{j=1}^{m-1} p_j \frac{(1/p_j^*)}{(1/(1-p_j^*))} \left(\frac{\sum_{h=1}^{j} \left(x_h \sqrt{p_h}\right) Z_h}{\mu^{(j)}} - \frac{\mu_{(j)} \sum_{h=j+1}^{m} \left(x_h \sqrt{p_h}\right) Z_h}{\left(\mu^{(j)}\right)^2} \right). \qquad (A1) \end{aligned}$$

Next

$$\begin{aligned}\sqrt{n}R_n(1,2) &= -\frac{\sum_{h=1}^{j}\sqrt{p_h}N_{h,n}}{\left(\sum_{h=1}^{j}p_h\right)\left(\sum_{h=1}^{j}n_h/n\right)}\\ &\rightarrow_{\mathbb{P}} -\frac{\sum_{h=1}^{j}\sqrt{p_h}Z_h}{(p_j^*)^2}. \quad (A2)\end{aligned}$$

Finally,

$$\begin{aligned}\sqrt{n}R_n(3) &= \sum_{j=1}^{m-1}\sqrt{p_j}\left(\frac{1}{p_j^*}-1\right)\frac{\mu_{(j)}}{\mu^{(j)}}N_{j,n}\\ &\rightarrow_{\mathbb{P}} \sum_{j=1}^{m-1}\sqrt{p_j}\frac{(1/p_j^*)\mu_{(j)}}{(1/(1-p_j^*))\mu^{(j)}}Z_j. \quad (A3)\end{aligned}$$

By combining Developments (A1), (A2) and (A3), we get

$$\begin{aligned}&\sqrt{n}(Z_{d,n}^*-Z_d^*)\\ &\rightarrow \sum_{j=1}^{m-1}p_j\frac{(1/p_j^*)}{(1/(1-p_j^*))}\left(\frac{\sum_{h=1}^{j}\left(x_h\sqrt{p_h}\right)Z_h}{\mu^{(j)}}-\frac{\mu_{(j)}\sum_{h=j+1}^{m}\left(x_h\sqrt{p_h}\right)Z_h}{\left(\mu^{(j)}\right)^2}\right)\\ &- \frac{\sum_{h=1}^{j}\sqrt{p_h}Z_h}{(p_j^*)^2}\\ &+ \sum_{j=1}^{m-1}\sqrt{p_j}\frac{(1/p_j^*)\mu_{(j)}}{(1/(1-p_j^*))\mu^{(j)}}Z_j\\ &= \left(\sum_{j=1}^{m-1}\langle\gamma_{j,1}D_{j,1},Z\rangle+\langle\gamma_{j,2}D_{j,2},Z\rangle+\langle(p_j^*)^{-2}E_j,Z\rangle\right)+\langle C,Z\rangle.\end{aligned}$$

We conclude that

$$\sqrt{n}(Z_{d,n}^*-Z_d^*)\rightarrow_{\mathbb{P}} H^tZ. \square$$

Appendix B. Proof of Theorem 2.2

Let us write, for $s\in\mathcal{R}$,

$$R_1(s)=R_1(s,\mathbb{P}_X)=\frac{\int t1_{]0,s]}(t)d\mathbb{P}_X(t)}{\int 1_{]0,s]}(t)d\mathbb{P}_X(t)},$$

and

$$R_2(s) = R_2(s, \mathbb{P}_X) = \frac{\int t 1_{]s,+\infty[}(t) d\mathbb{P}_X(t)}{\int]s, +\infty[d\mathbb{P}_X(t)}.$$

We have

$$Z_d(\mathbb{P}_X) = Z_d = 1 - \int \frac{R_1(s)}{R_2(s)} \mathbb{P}_X(s) d\nu(s).$$

By using Formula (1.1), we have

$$\frac{d(\mathbb{P}_X^{(\varepsilon)} - \mathbb{P}_X)}{\varepsilon} = -d\mathbb{P}_X + d\delta_x.$$

For short, we write

$$R_i(s, \mathbb{P}_X) = R_i(s) \text{ and } R_i(s, \mathbb{P}_X^{(\varepsilon)}) = R_i(s, \varepsilon), i \in \{1, 2\}.$$

We have

$$\begin{aligned} Z_d(\mathbb{P}_X^{(\varepsilon)}) - Z_d(\mathbb{P}_X) &= -(1-\varepsilon) \int \mathbb{P}_X(s) \frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} d\nu - \varepsilon \int \delta_x(s) \frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} d\nu \\ &+ \int \mathbb{P}_X(s) \frac{R_1(s)}{R_2(s)} d\nu \\ &= - \int \mathbb{P}_X(s) \left(\frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} - \frac{R_1(s)}{R_2(s)} \right) d\nu \\ &+ \varepsilon \int \mathbb{P}_X(s) \frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} d\nu - \varepsilon \int \delta_x(s) \frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)} d\nu. \end{aligned}$$

Let us apply the definition of the *IF* as in Formula (1.2). Since $\mathbb{P}_X^{(\varepsilon)} \to \mathbb{P}_X$ as $\varepsilon \to 0$ (The convergence being meant as a convergence in law), we have no problem to see that

$$
\begin{aligned}
\lim_{\varepsilon\to 0}\frac{Z_d(\mathbb{P}_X^{(\varepsilon)})-Z_d(\mathbb{P}_X)}{\varepsilon} &= \int \mathbb{P}_X(s)\frac{R_1(s)}{R_2(s)}d\nu - \int \delta_x(s)\frac{R_1(s)}{R_2(s)}d\nu \\
&- \int \mathbb{P}_X(s)\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\frac{R_1(s,\varepsilon)}{R_2(s,\varepsilon)}-\frac{R_1(s)}{R_2(s)}\right)d\nu. \qquad (0.1)
\end{aligned}
$$

So we have to find the influence function of $R_1(s)/R_2(s)$. By formally representing the differentiation of a functional $T(\mathbb{P}_X)$ by

$$\frac{\partial T(\mathbb{P}_X)}{\partial \lambda},$$

we have that the influence function of $R_1(s)/R_2(s)$ is given by

$$IF(R_1(s)/R_2(s),x)=\frac{R_2(s)\frac{\partial R_1(s)}{\partial\lambda}-R_1(s)\frac{\partial R_2(s)}{\partial\lambda}}{R_2(s)^2}.$$

But

$$
\begin{aligned}
R_1(s,\varepsilon)-R_1(s) &= \frac{\int t1_{]0,s]}(t)d\mathbb{P}_X(t)}{\int 1_{]0,s]}(t)d\mathbb{P}_X^{(\varepsilon)}(t)} \\
&- \frac{\varepsilon\int t1_{]0,s]}(t)d\mathbb{P}_X(t)}{\int 1_{]0,s]}(t)d\mathbb{P}_X^{(\varepsilon)}(t)} \\
&+ \frac{\varepsilon\int t1_{]0,s]}(t)d\delta_x(t)}{\int 1_{]0,s]}(t)d\mathbb{P}_X^{(\varepsilon)}(t)} - \frac{\int t1_{]0,s]}(t)d\mathbb{P}_X(t)}{\int 1_{]0,s]}(t)d\mathbb{P}_X(t)} \\
&= \frac{\int t1_{]0,x_j]}(t)d(\mathbb{P}_X^{(\varepsilon)}(t)-\mathbb{P}_X(t))}{\int 1_{]0,s]}(t)d\mathbb{P}_X^{(\varepsilon)}(t)} \\
&- \frac{\int 1_{]0,s]}(t)d(\mathbb{P}_X^{(\varepsilon)}(t)-\mathbb{P}_X(t))}{\left(\int 1_{]0,s]}(t)d\mathbb{P}_X^{(\varepsilon)}(t)\right)\left(\int 1_{]0,s]}(t)d\mathbb{P}_X(t)\right)}\int t1_{]0,s]}(t)d\mathbb{P}_X(t).
\end{aligned}
$$

We get

$$\begin{aligned}
\lim_{\varepsilon\to 0}\frac{R_1(s,\varepsilon)-R_1(s)}{\varepsilon} &= \frac{\int t1_{]0,s]}(t)d(-\mathbb{P}_X(t)+\delta_x)}{\int 1_{]0,s]}(t)d\mathbb{P}_X(t)} \\
&- \frac{\int 1_{]0,s]}(t)d(-\mathbb{P}_X(t)+\delta_x)}{\left(\int 1_{]0,s]}(t)d\mathbb{P}_X(t)\right)^2}\int t1_{]0,s]}(t)d\mathbb{P}_X(t) \\
&= \frac{-\left(\int t1_{]0,s]}(t)d\mathbb{P}_X(t)\right)+x1_{]0,s]}(x)}{\int 1_{]0,s]}(t)d\mathbb{P}_X(t)} \\
&- \frac{-\left(\int 1_{]0,s]}(t)d\mathbb{P}_X(t)\right)+1_{]0,s]}(x)}{\left(\int 1_{]0,s]}(t)d\mathbb{P}_X(t)\right)^2}\int t1_{]0,s]}(t)d\mathbb{P}_X(t)
\end{aligned}$$

and so

$$\frac{\partial R_1(s)}{\partial\lambda} = -R_1(s)+\frac{x1_{]0,s]}(x)}{F(s)}+R_1(s)-\frac{R_1(s)}{F(s)}1_{]0,s]}(x).$$

By treating $R_2(s)$ in the same manner, we have (we should not forget that we differentiate in the probability)

$$\begin{aligned}
\frac{\partial R_1(s)}{\partial\lambda} &= \frac{x1_{]0,s]}(x)}{F(s)}-\frac{R_1(s)}{F(s)}1_{]0,s]}(x) \\
\frac{\partial R_2(s)}{\partial\lambda} &= \frac{x1_{]s,+\infty]}(x)}{1-F(s)}-\frac{R_2(s)}{1-F(s)}1_{]s,+\infty]}(x).
\end{aligned}$$

Thus

$$\lim_{\varepsilon\to 0}\frac{R_1(s,\varepsilon)-R_1(s)}{\varepsilon} = \left(\frac{1_{]0,s]}(x)}{R_2(s)F(s)}-\frac{R_1(s)1_{]s,+\infty]}(x)}{R_2^2(s)(1-F(s))}\right)x$$
$$+\left(\frac{R_1(s)}{R_2(s)(1-F(s))}1_{]s,+\infty]}(x)-\frac{R_1(s)}{R_2(s)F(s)}1_{]0,s]}(x)\right).$$

By replacing this limit with its expression in Equation (0.1) we get.

$$\lim_{\varepsilon\to 0}\frac{Z_d(\mathbb{P}_X^{(\varepsilon)})-Z_d(\mathbb{P}_X)}{\varepsilon} = \int\mathbb{P}_X(s)\frac{R_1(s)}{R_2(s)}d\nu-\int\delta_x(s)\frac{R_1(s)}{R_2(s)}d\nu$$
$$+\int\mathbb{P}_X(s)\left(\frac{R_1(s)}{R_2(s)^2(1-F(s)})1_{]s,+\infty]}(x)-\frac{1}{R_2(s)F(s)}1_{]0,s]}(x)\right)xd\nu$$

$$+\int \mathbb{P}_X(s)\left(\frac{R_1(s)}{R_2(s)F(s)}1_{]0,s]}(x)-\frac{R_1(s)}{R_2(s)(1-F(s))}1_{]s,+\infty]}(x)\right)d\nu.$$

From this, the proof is concluded. ■

ACKNOWLEDGMENTS

The authors thank the Editor-in-Chief and the anonymous reviewers for their very valuable comments and suggestions that helped to improve the quality of the paper.

REFERENCES

[1] Agence Nationale de la Statistique et de la Démographie (ANSD) (2019). Enquête Sénégalaise Auprès des Ménages 2 (ESAM II, 2001-2002). Report Website: http://anads.ansd.sn/index.php/catalog/47/study-description.

[2] Allison, P.D. (1978). Measures of inequality, *American Sociological Review*, **43**, 478–484.

[3] Cowell, F.A. and Flachaire, E. (2007). Income Distribution and Inequality Measurement: The Problem of Extreme Values, *Journal of Econometrics*, **141**, 1044–1072.

[4] Cowell, F.A., Flachaire, E. and Bandyopadhyay, S. (2009). Goodness-of-Fit: An Economic Approach, *Distributional Analysis Research Programme (DARP 101).* Discussion Paper. Department of Economics, University of Oxford.

[5] Emad-Eldin, A.A.A. and Marilou, O.H. (1999). Nonparametric inference for Zenga's measure of income inequality, *Metron - International Journal of Statistics*, **LVII**, 1-2, 69–84.

[6] Greselin, F., Pasquazzi, L. and Zitikis, R. (2010a). Zenga's new index of economic inequality, its estimation, and analysis of incomes in Italy, *Journal of Probability and Statistics*, **2010**, 1–26.

[7] Greselin, F., Pasquazzi, L. and Zitikis, R. (2010b). Asymptotic Theory for Zenga's New Index of Economic Inequality. In "THE PROCEEDINGS OF THE 45TH SCIENTIFIC MEETING OF THE ITALIAN STATISTICAL SOCIETY" (CLEUP Eds.), Padova.

[8] Greselin, F., Pasquazzi, L. and Zitikis, R. (2014). Heavy tailed capital incomes: Zenga index, statistical inference, and ECHP data analysis, *Extremes*, **17**, 127–155.

[9] Hulliger, B. and Schoch, T. (2009). Robust estimation of the quintile share ratio with bias reduction, Presented at the Swiss Statistics Meeting, October 30, 2009, Geneva.

[10] Kpanzou, T.A. (2014). Asymptotic distribution of the quintile share ratio estimator, *Afrika Statistika*, **9**, 1, 659–670.

[11] Kpanzou, T. A., de Wet, T. and Neethling, A. (2013). Semi-parametric Estimation of Inequality Measures, *South African Statistical Journal*, **47**, 33–48.

[12] Lo, G.S. (2016). VB and R codes using Households databases available in the NSI's: A prelude to statistical applied studies, *African Journal to Applied Statistics*, **3**, 121–156, DOI: http://dx.doi.org/10.16929/ajas/206

[13] Lo, G.S., Ngom, M. and Kpanzou, T.A. (2016). *WEAK CONVERGENCE (IA). SEQUENCES OF RANDOM VECTORS*, SPAS Books Series, Saint-Louis, Senegal - Calgary, Canada.

[14] Magurran, A.E. (1991). *ECOLOGICAL DIVERSITY AND ITS MEASUREMENT*, Chapman and Hall.

[15] Mergane, P.D. and Lo, G.S. (2013). On the Functional Empirical Process and Its Application to the Mutual Influence of the Theil-Like Inequality Measure and the Growth, *Applied Mathematics*. **4**, 986–1000, http://dx.doi.org/10.4236/am.2013.47136.

[16] Rousseau, R. (1993). Measuring concentration: sampling design issues as illustrated by the case of perfectly stratified samples, *Scientometrics*, **28**, 3–14.

[17] Wasserman, L. (2006). *ALL OF NONPARAMETRIC STATISTICS*, Springer.

[18] Wichura, M.J. (1970). On the Construction Of Almost sure Uniformly Convergent Random Variable with Given Weakly Convergent Images laws, *Annals of Mathematical Statistics*, **41**, 1, 284–291.

[19] White, M.J. (1986). Segregation and diversity measures in population distribution, *Population Index*, **52**, 198–221.

[20] Zenga, M. (1984). Proposta per un Indice di Concentrazione Basato sui Rapporti tra Quantili di Popolazione e Quantili di Reddito, *Giornale degli Economisti e Annali di Economia*, **43**, 301–326.

[21] Zenga, M. (1990). Concentration Curves And Concentration Indexes Derived From Them. In "INCOME AND WEALTH DISTRIBUTION, INEQUALITY AND POVERTY. STUDIES IN CONTEMPORARY ECONOMICS" (C. Dagum, M. Zenga, Eds.), Springer, Berlin, Heidelberg, 94–110, https://doi.org/10.1007/978-3-642-84250-4_7.

[22] Zenga, M. (2006). Inequality Curve and Inequality Index Based on the Ratios Between Lower and Upper Arithmetic Means, (provisional work). Rapporto di Ricerca 115, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Universita degli Studi du Milano-Bicocca.