Afrika Statistika
Vol. 18 (3), 2023, pages 3521 - 3550.
In: Special issue of the statistical applications of the fep and the *lrfep* to nonparametric estimation
DOI: https://dx.doi.org/10.16929/as/2023.3521.317.sFEP

Afrika Statistika

ISSN 2316-090X

General asymptotic representations of indexes based on the functional empirical process and the residual functional empirical process and applications

Gane Samb Lo [1,3,4,*] , **Tchilabalo Abozou Kpanzou** [2,3], **Gandasor Bonyiri Onesiphore Da** [1,3]

[1] LERSTAD, Gaston Berger University, Saint-Louis, Sénégal
[2] LaMMASD, University of Kara, Togo
[3] Imhotep International Mathematical Center (imh-imc), https://imhotepsciences.org/
[4] Affiliated to : LSTA, Pierre and Marie Curie University, Paris VI, France.
AUST - African University of Sciences and Technology, Abuja, Nigeria
Permanent address : 1178 Evanston Dr NW T3P 0J9,Calgary, Alberta, Canada

Received on January 01, 2023; Accepted on July 1, 2023; Published on September 01, 2024

Abstract The aim of this paper is to state a general asymptotic representation (*GAR*) of a broad class of statistics using two key fundamental processes, namely the functional empirical process (*fep*) and the residual functional empirical process introduced by Lo and Sall (2010a, 2010b) abbreviated *lrfep*. See Page 3522 for a full abstract.

Key words: weak convergence; index asymptotic representation; functional Brownian bridge, functional empirical function; functional empirical process and residual functional empirical process.
AMS 2010 Mathematics Subject Classification Objects : Primary 62E20, 62F05. Secondary 62F12, 62G07

* Correspondence Author: Gane Samb Lo (gslo@aust.edu.ng, ganesamblo@ganesamblo.net, ganesamblo@gmail.com),
Tchilabalo Abozou Kpanzou: kpanzout@gmail.com
Gandasor Bonyiri Onesiphore Da: da.gandasor-bonyiri-onesiphore@ugb.edu.sn

Full Abstract. The objective of this paper is to establish a general asymptotic representation (*GAR*) for a wide range of statistics, employing two fundamental processes: the functional empirical process (*fep*) and the residual functional empirical process introduced by Lo and Sall (2010a, 2010b), denoted as *lrfep*. The functional empirical process (*fep*) is defined as follows:

$$\mathbb{G}_n(h) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \{h(X_j) - \mathbb{E}h(X_j)\},$$

[where X, X_1, $\cdots$, X_n is a sample from a random d-vectors X of size $(n+1)$ with $n \geq 1$ and h is a measurable function defined on $\mathbb{R}^d$ such that $\mathbb{E}h(X)^2 < +\infty$]. It is a powerful tool for deriving asymptotic laws. An earlier and simpler version of this paper focused on the application of the (*fep*) to statistics J_n that can be turned into an asymptotic algebraic expression of empirical functions of the form

$$J_n = \mathbb{E}h(X) + n^{-1/2}\mathbb{G}_n(h) + o_{\mathbb{P}}(n^{-1/2}). \quad \textit{SGAR}$$

However, not all statistics, in particular welfare indexes, conform to this form. In many scenarios, functions of the order statistics $X_{1,n} \leq, \cdots, \leq X_{n,n}$ are involved, resulting in L-statistics. In such cases, the (*fep*) can still be utilized, but in combination with the related residual functional empirical process introduced by Lo and Sall (2010a, 2010b). This combination leads to general asymptotic representations (GAR) for a wide range of statistical indexes

$$J_n = \mathbb{E}h(X) + n^{-1/2}\left(\mathbb{G}_n(h) + \int_0^1 \mathbb{G}_n(\tilde{f}_s)\ell(s)\ ds + o_{\mathbb{P}}(1)\right), \quad \textit{FGAR}$$

where $\tilde{f}_s = 1_{]-\infty, F^{-1}(s)]}$ is the indicator function of the interval $]-\infty, F^{-1}(s)]$ of $\mathbb{R}$, $h \in L^2(\mathbb{P}_X)$ and $\ell(\circ)$ is a measurable function of $s \in (0,1)$. Such representations, when associated with copulas, provide a robust framework for comparing indices over the time or across different areas. The comprehensive theory is presented alongside explicit examples, facilitating the utilization by researchers.

Résumé. (French abstract). L'objectif de cet article est d'établir une représentation asymptotique générale (*RAG*) pour un large éventail de statistiques, en utilisant deux processus : le processus empirique fonctionnel (*fep*) et le processus empirique résiduel introduit par Lo et Sall (2010a, 2010b), abrégé sous le nom de *lrfep*. Le processus empirique fonctionnel (*fep*) est défini comme suit :

$$\mathbb{G}_n(h) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \{h(X_j) - \mathbb{E}h(X_j)\},$$

[où X, X_1, $\cdots$, X_n représentent un échantillon d'un vecteur aléatoire (de longueur d) de taille $(n+1)$ with $n \geq 1$, et h est une fonction mesurable définie sur $\mathbb{R}^d$ telle que $\mathbb{E}h(X)^2 < +\infty$]. Ce processus est un outil puissant pour déduire les propriétés asymptotiques. Une version antérieure et plus simple de cet article se concentrait sur l'application du *fep* aux statistiques J_n qui peuvent être transformées en une expression algébrique asymptotique des fonctions empiriques sous la forme :

$$J_n = \mathbb{E}h(X) + n^{-1/2}\mathbb{G}_n(h) + o_{\mathbb{P}}(n^{-1/2}). \quad \textit{SGAR}$$

Cependant, toutes les statistiques, en particulier les indices de bien-être, ne suivent pas cette forme. Dans de nombreuses situations, des fonctions des statistiques d'ordre $X_{1,n} \leq, \cdots, \leq X_{n,n}$ sont impliquées, ce qui donne naissance aux statistiques L. Dans de tels cas, le *fep* peut toujours être utilisé, mais en combinaison avec le processus empirique résiduel associé introduit par Lo et Sall (2010a, 2010b). Cette combinaison conduit à une représentation asymptotique générale (GAR) pour un large éventail d'indices statistiques :

$$J_n = \mathbb{E}h(X) + n^{-1/2}\biggl(\mathbb{G}_n(h) + \int_0^1 \mathbb{G}_n(\tilde{f}_s)\ell(s)\, ds + o_{\mathbb{P}}(1)\biggr), \quad \textit{FGAR}$$

où $\tilde{f}_s = 1_{]-\infty,F^{-1}(s)]}$ est la fonction indicatrice de l'intervalle $]-\infty, F^{-1}(s)]$ de $\mathbb{R}$, $h \in L^2(\mathbb{P}_X)$, et $\ell(\circ)$ est une fonction mesurable de $s \in (0,1)$. De telles représentations, lorsqu'elles sont combinées avec les copules, fournissent un cadre solide pour la comparaison des indices au fil du temps ou entre différentes zones. La théorie complète est présentée avec des exemples explicites, facilitant son utilisation par les chercheurs.

Presentations of authors.

Gane Samb Lo, Ph.D., is a retired full professor from Université Gaston Berger (2023), Saint-Louis, SENEGAL. He is the founder and The Probability and Statistics Chair holder at: Imhotep International Mathematical Center (imho-imc), https://imhotepsciences.org.

Tchilabalo Abozou Kpanzou, Ph.D., is a professor of Statistics and Mathematics at Universite of Kara, Togo. He is a lead researcher at: Imhotep International Mathematical Center (imh-imc), https://imhotepsciences.org.

Gandasor Bonyiri Onesiphore Da, M.Sc, is preparing a Ph.D. degree at: Doctoral Studies, LERSTAD, Université Gaston Berger, Saint-Louis, SENEGAL, and in the Imhotep International Mathematical Center (imh-imc), https://imhotepsciences.org, under the supervision of the previous two coauthors.

1. Introduction

There is a huge number of statistics and indexes, that are used in Mathematical Statistics and in many other disciplines, whose asymptotic behaviour can be studied in a unified approach based on the functional empirical processes. For instance, indexes of poverty have been extensively used in Ecometrics. A detailed survey of them is available in Zeng (1997), including simple ones as the Foster, Greer and Thorbecke index (See Foster *et al.* (1984)) and most elaborated ones such as the statistics of Kakwani (1980) and Sen (1976) (Economics Nobel prize winner). The first attempt of having a common frame for the asymptotic theory of those statistics was done in Lo (2013). However, the approach in that paper uses the real-valued empirical processes. Later, it was realized that the rightest approach would rely on function-value empirical process, i.e., the functional empirical process (*fep*). The combination of the *fep* with the so-called residual empirical process - that has been later turned into a functional form named *residual functional empirical process* - started a new and global approach that provides general asymptotic representations (*GAR*) in the *fep* for a huge class of indices arising in many fields. (*GAR*)'s have been already provided in a number of papers, in particular in Mergane *et al.* (2018b) for Gini's inequality index, Mergane *et al.* (2018a) for the class of Theil indexes. These *GAR*s have proved to be very effective in addressing the decomposability of statistics topic by tackling the lack of decomposability as in Haidara *et al.* (2018). The main reason of the success of the *fep* is its linearity property that the real-valued empirical process fails to have.

We feel that this method of *GAR*'s in the *fep* is not sufficiently exploited and it is even unknown by the great public of applied researchers. In that sense, the main motivation of this paper is to provide a new and global exposure of this method to researchers in order to show them how to apply it in specific cases and to give explicit examples of usage.

At the same time, a number of applications in a series of papers will follow in the special number of this journal. Another paper on statistical computations and related packages will be published. That sample of cases, more or less numerous, will be the beginning of a general handbook that will be handed to researchers.

What is very amazing is that, despite its power, this technique does not need the heavy weapons of functional weak convergence such as Donsker classes, Vapnik-Chervonenkis theory (see Vapnik and Chervonenkis (1971)) or entropy numbers, etc., nor the contents of books like van der Vaart and Wellner (1996), Gaenssler (1983) or in Pollard (1984) or alike texts. Actually, we only use our tools at the finite-distribution level, where the functional empirical processes can be reasonably easy to use without heavy theories.

Accordingly, we will organize the rest of the paper as follows. In next section (Section 2), we describe the *fep*, justify it and consider a case study on a non-trivial example. Namely, we will entirely show how the method works for the plug-in estimator of the linear correlation index between two random variables. In Section 3, we introduce the *lrfep*, give the conditions ensuring its weak convergence. In Section 4, we use the Gini's index as a study case for applying the recommendations in the previous section. In Section 5, we make a final summary of the whole method and give final remarks and recommendations.

2. The Functional Empirical Process and applications in finding a simple *(GAR)*

2.1. The functional empirical process (fep)

Let us begin by introducing the *fep* and describe its asymptotic representation.

What is the *fep*?.

Let Z_1, Z_2, ... be a sequence of independent copies of a random variable Z defined on the same probability space with values on some metric space (S, d). We define for each $n \geq 1$, the functional empirical process by

$$\mathbb{G}_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(Z_i) - \mathbb{E}f(Z_i)),$$

where f is a real and measurable function defined on $\mathbb{R}$ such that

$$\mathbb{V}_Z(f) = \int (f(x) - \mathbb{P}_Z(f))^2 \, dP_Z(x) < \infty, \tag{1}$$

which entails

$$\mathbb{P}_Z(|f|) = \int |f(x)| \, dP_Z(x) < \infty. \tag{2}$$

We denote by $\mathcal{F}(S)$ - $\mathcal{F}$ for short - the class of real-valued measurable functions that are defined on S such that (1) holds. The space $\mathcal{F}$, when endowed with the addition and the external multiplication by real scalars, is a linear space. Next, it is remarkable that $\mathbb{G}_n$ is linear on $\mathcal{F}$, that is for f and g in $\mathcal{F}$ and for $(a, b) \in \mathbb{R}^2$, we have

$$a\mathbb{G}_n(f) + b\mathbb{G}_n(g) = \mathbb{G}_n(af + bg).$$

We have this result.

Lemma 1. *Given the notation above, then for any finite number of elements $f_1, ..., f_k$ of $\mathcal{F}$, $k \geq 1$, we have*

$${}^t(\mathbb{G}_n(f_1), ..., \mathbb{G}_n(f_k)) \rightsquigarrow \mathcal{N}_k(0, \Gamma(f_i, f_j)_{1 \leq i,j \leq k}),$$

where

$$\Gamma(f_i, f_j) = \int (f_i - \mathbb{P}_Z(f_i)) \, (f_j - \mathbb{P}_Z(f_j)) \, d\mathbb{P}_Z(x), 1 \leq i, j \leq k.$$

Proof. It is enough to use the Cramér-Wold Criterion (see for example Billingsley (1968), page 45), that is to show that for any $a =^t (a_1, ..., a_k) \in \mathbb{R}^k$, by denoting $T_n =^t (\mathbb{G}_n(f_1), ..., \mathbb{G}_n(f_k))$, we have $< a, T_n > \rightsquigarrow < a, T >$, where T follows the $\mathcal{N}_k(0, \Gamma(f_i, f_j)_{1\leq i,j\leq k})$ law and $< \circ, \circ >$ stands for the usual scalar product in $\mathbb{R}^k$. But, by the standard central limit theorem in $\mathbb{R}$, we have

$$< a, T_n >= \mathbb{G}_n \left(\sum_{i=1}^{k} a_i f_i \right) \rightsquigarrow N(0, \sigma_\infty^2),$$

where, for $g = \sum_{1\leq i\leq k} a_i f_i$,

$$\sigma_\infty^2 = \int \left(g(x) - \mathbb{P}_Z(g)\right)^2 dP_Z(x)$$

and this easily gives

$$\sigma_\infty^2 = \sum_{1\leq i,j\leq k} a_i a_j \Gamma(f_i, f_j),$$

so that $N(0, \sigma_\infty^2)$ is the law of $< a, T >$. The proof is finished.

2.2. How to use the tool?

In our daily activities as researchers in asymptotic statistics, we usually need to find the asymptotic laws of statistics on $\mathbb{R}^k$. Once we have our sample Z_1, Z_2, ... as random variables defined on the same probability space with values in $\mathbb{R}^k$, the studied statistic, say H_n, in a significant number of cases, can be expressed in the following form:

$$H_n = \frac{1}{n} \sum_{i=1}^{k} H(Z_i)$$

for $H \in \mathcal{F}$. This gives the simplest asymptotic representation, for $\mu(H) = \mathbb{E}H(Z)$:

$$H_n = \mu(H) + n^{-1/2}\mathbb{G}_n(H). \tag{3}$$

From such exact *GAR*'s, the Delta-method is a common tools to proceed to next steps. For that, we need to address the probability asymototic boundedness of $\mathbb{G}_n(H)$. Since it weakly converges to a random element, say $M(H)$, we get $\|\mathbb{G}_n(H)\| \rightsquigarrow \|M(H)\|$, by the continuous mapping theorem. Next, we get for all $\lambda > 0$, by the assertion of the Portmanteau Theorem concerning open sets,

$$\limsup_{n\to\infty} P(\|\mathbb{G}_n(H)\| > \lambda) \leq P(\|M(H)\| > \lambda)$$

and then

$$\liminf_{\lambda\to\infty} \limsup_{n\to\infty} P(\|\mathbb{G}_n(H)\| > \lambda) \leq \limsup P(\|M(H)\| > \lambda) = 0.$$

From this, we use the big $O_{\mathbb{P}}$ notation, that is $\mathbb{G}_n(H) = O_{\mathbb{P}}(1)$. Formula (3) becomes

$$H_n = \mu(H) + n^{-1/2}\mathbb{G}_n(H) = \mu(H) + O_{\mathbb{P}}(n^{-1/2})$$

and we will be able to use the Delta-method. Indeed, let $g : \mathbb{R} \longmapsto \mathbb{R}$ be continuously differentiable on a neighborhood of $\mu(H)$. The mean value theorem leads to

$$g(H_n) = g(\mu(H)) + g'(\mu_n(H))\ n^{-1/2}\mathbb{G}_n(H), \tag{4}$$

where

$$\mu_n(H) \in [(\mu(H) + n^{-1/2}\mathbb{G}_n(H)) \wedge \mu(H), (\mu(H) + n^{-1/2}\mathbb{G}_n(H)) \vee \mu(H)],$$

so that

$$|\mu_n(H) - \mu(H)| \leq n^{-1/2}\mathbb{G}_n(H) = O_{\mathbb{P}}(n^{-1/2}).$$

Then $\mu_n(H)$ converges to $\mu(H)$ in probability (denoted as: $\mu_n(H) \to_{\mathbb{P}} \mu(H)$). But the convergence in probability to a constant is equivalent to the weak convergence. Then $\mu_n(H) \rightsquigarrow \mu(H)$. Using again the continuous mapping theorem, $g'(\mu_n(H)) \rightsquigarrow g'(\mu(H))$ which in turn yields $g'(\mu_n(H)) \to_{\mathbb{P}} g'(\mu(H))$ by the characterization of the weak convergence to a constant. Now (4) becomes

$$\begin{aligned}
g(H_n) &= g(\mu(H)) + (g'(\mu(H) + o_P(1))\ n^{-1/2}\mathbb{G}_n(H) \\
&= g(\mu(H)) + n^{-1/2}g'(\mu(H))\mathbb{G}_n(H) + o_{\mathbb{P}}(n^{-1/2}) \\
&= g(\mu(H)) + n^{-1/2}\mathbb{G}_n(g'(\mu(H))H) + o_P(n^{-1/2}).
\end{aligned}$$

We arrive at the final expansion

$$g(H_n) = g(\mu(H)) + n^{-1/2}\mathbb{G}_n(g'(\mu(H))H) + o_P(n^{-1/2}). \tag{5}$$

We just get a *GAR* in the *fep*, for some constant C,

$$J_n = C + n^{-1/2}\mathbb{G}_n(h) + o_P(n^{-1/2}). \tag{6}$$

for h^2 being $\mathbb{P}_X$ integrable. A *GAR* like (6), may undergo the Delta-method to give (under the needed conditions)

$$g(J_n) = g(C) + n^{-1/2}\mathbb{G}_n(g'(C)h) + o_P(n^{-1/2}). \quad (7)$$

which is still of the form of (6).

The general methods works as following: from exact *GAR*s or *GAR*s like (6), we get others *GAR*s of the same form. Fortunately, many statistics are algebraic combinations of *GAR*s in the form (6). To get the final *GAR*, we have the tools of handling these algebraic combinations in the next theorem.

Lemma 2. *Let (A_n) and (B_n) be two sequences of statistics expressed in the following GARs: real-valued functions $of\ z \in S$. Suppose that $A_n = A + n^{-1/2}\mathbb{G}_n(L) + o_{\mathbb{P}}(n^{-1/2})$ and $B_n = B + n^{-1/2}\mathbb{G}_n(H) + o_{\mathbb{P}}(n^{-1/2})$, where A, B, L and H are defined accordingly as above. We have the following GARs:*

$$A_n + B_n = A + B + n^{-1/2}\mathbb{G}_n(L+H) + o_{\mathbb{P}}(n^{-1/2}),$$

$$A_nB_n = AB + n^{-1/2}\mathbb{G}_n(BL + AH) + o_{\mathbb{P}}(n^{-1/2}).$$

and if $B \neq 0$,

$$\frac{A_n}{B_n} = \frac{A}{B} + n^{-1/2}\mathbb{G}_n\left(\frac{1}{B}L - \frac{A}{B^2}H\right) + o_{\mathbb{P}}(n^{-1/2}).$$

By putting together all the described steps in a smart way, the methodology will lead us to a final result of the form

$$T_n = t + n^{-1/2}\mathbb{G}_n(h) + o_P(n^{-1/2})$$

which entails the weak convergence

$$\sqrt{n}(T_n - t) = \mathbb{G}_n(h) + o_P(1) \rightsquigarrow N(0, \Gamma(h,h)).$$

2.3. A non-trivial example

We are going to illustrate our *fep* tool on the plug-in estimator of the linear correlation coefficient of the pair of random variables (X, Y), none of the coordinates being degenerate, defined as follows

$$\rho = \frac{\sigma_{xy}}{\sigma_x^2\sigma_y^2},$$

where

$$\mu_x = \int x\ dP_X(x),\ \ \mu_y = \int x\ dP_X(x),\ \ \sigma_{xy} = \int (x-\mu_x)(y-\mu_y)dP_{(X,Y)}(x,y).$$

$$\sigma_x^2 = \int (x-\mu_x)^2 d\mathbb{P}_X(x), \ \sigma_y^2 = \int (x-\mu_x)(y-\mu_y) d\mathbb{P}_X(y).$$

We also dismiss the case the case where $|\rho| = 1$, for which one of X and Y is an affine function of the other, for example $X = aY + b$.

It is clear that centering the variables X and Y and normalizing them by their standard deviations σ_x and σ_y does not change the correlation coefficient ρ. So we consider the coordinates as centered and normalized, i.e.

$$\mu_x = \ \mu_y = 0, \ \sigma_x = \sigma_y = 1.$$

However, we will let these coefficients appear with their names and we only use their particular values at the conclusion stage.

Let us consider the plug-in estimator of ρ. To this end, let (X_1, Y_1), (X_2, Y_2), ... be a sequence independent observations of (X, Y). For each $n \geq 1$, the plug-in estimator is the following

$$\rho_n = \left\{ \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) \right\} \left\{ \frac{1}{n^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 \times \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right\}^{-1/2}.$$

We are going to give the asymptotic theory of ρ_n as an estimator of ρ. Introduce the notation

$$\mu_{(p,x),(q,y)} = E((X-\mu_x)^p(Y-\mu_y)^q), \mu_{4,x} = E(X-\mu_x)^4, \ \mu_{4,x} = E(X-\mu_x)^4).$$

Here is our main Theorem.

Theorem 1. *Suppose that neither of X and Y is degenerated and both have finite fourth moments and that X^3Y and XY^3 have finite expectations. Then, as $n \to \infty$,*

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma^2),$$

where

$$\begin{aligned}\sigma^2 \ &= \ \sigma_x^{-2}\sigma_y^{-2}(1+\rho^2/2)\mu_{(2,x),(2,y)} + \rho^2(\sigma_x^{-4}\mu_{4,x} + \sigma_y^{-4}\mu_{4,y})/4 \\ &\quad -\rho(\sigma_x^{-3}\sigma_y^{-1}\mu_{(3,x),(1,y)} + \sigma_x^{-1}\sigma_y^{-3}\mu_{(1,x),(3,y)}).\end{aligned}$$

This result enables to test independence between X and Y, or to test non linear correlation in the following sense.

Theorem 2. *Suppose that the assumptions of Theorem 1 hold. Then*
***(1)** If X and Y are not linearly correlated, that is $\rho = 0$, we have*

$$\sqrt{n}\rho_n \rightsquigarrow N(0, \sigma_1^2),$$

where

$$\sigma_1^2 = \sigma_x^{-2}\sigma_y^{-2}\mu_{(2,x),(2,y)}.$$

***(2)** If X and Y are independent, then $\rho = 0$, and*

$$\sqrt{n}\rho_n \rightsquigarrow N(0, 1)$$

Proofs. We are going to use the function empirical process based on the observations $(X_i, Y_i), i = 1, 2, \ldots$ that are independent copies of (X, Y). Write

$$\rho_n = \frac{\frac{1}{n}\sum_{i=1}^n X_iY_i - \overline{X}\ \overline{Y}}{\left\{\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2\right\}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^n Y_i^2 - \overline{Y}^2\right\}^{1/2}} = \frac{A_n}{B_n}.$$

Let us say for once that all the functions of (X, Y) that will appear below are measurable and have finite second moments. Let us handle separately the numerator and denominator. To treat A_n, using the empirical process implies that

$$\left\{ \begin{array}{l} \frac{1}{n}\sum_{i=1}^{n} X_iY_i = \mu_{xy} + n^{-1/2}G_n(p), \\ \overline{X} = \mu_x + n^{-1/2}G_n(\pi_1), \\ \overline{Y} = \mu_y + n^{-1/2}G_n(\pi_2), \end{array} \right. \tag{8}$$

where $p(x, y) = xy$, $\pi_1(x, y) = x$, $\pi_2(x, y) = y$. From there we use the fact that $G_n(g) = O_P(1)$ for $\mathbb{E}(g(X, Y)^2) < +\infty$ and get

$$A_n = \mu_{xy} + n^{-1/2}G_n(p) - (\mu_x + n^{-1/2}G_n(\pi_1))(\mu_y + n^{-1/2}G_n(\pi_2)). \tag{9}$$

This leads to

$$A_n = \sigma_{xy} + n^{-1/2}G_n(H_1) + o_P(n^{-1/2})$$

with

$$H_1(x, y) = p(x, y) - \mu_x\pi_2 - \mu_y\pi_{1.}$$

Next, we have to handle B_n. Since the roles of $\left\{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right\}^{1/2}$ and $\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2\right\}^{1/2}$ ar symmetrical, we treat one of them and extend the results to the other. Consider $\left\{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right\}^{1/2}$. From (8), use the Delta-method to get

$$\overline{X}^2 = \left(\mu_x + n^{-1/2}G_n(\pi_1)\right)^2 = \mu_x^2 + 2\mu_x n^{-1/2}G_n(\pi_1) + o_P(n^{-1/2}),$$

that is

$$\overline{X}^2 = \left(\mu_x + n^{-1/2}G_n(\pi_1)\right)^2 = \mu_x^2 + n^{-1/2}G_n(2\mu_x\pi_1) + o_P(n^{-1/2}).$$

From there, we get

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 &= m_{2,x} + n^{-1/2}G_n(\pi_1^2) - \overline{X}^2 \\ &= m_{2,x} - \mu_x^2 + n^{-1/2}G_n(\pi_1^2 - 2\mu_x\pi_1) + o_P(n^{-1/2}) \\ &= \sigma_x^2 + n^{-1/2}G_n(\pi_1^2 - 2\mu_x\pi_1) + o_P(n^{-1/2}). \end{aligned}$$

Using the Delta-method once again leads to

$$\left\{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right\}^{1/2} = \sigma_x + n^{-1/2}G_n\left(\frac{1}{2\sigma_x}\left\{\pi_1^2 - 2\mu_x\pi_1\right\}\right) + o_P(n^{-1/2}).$$

In a similar way, we get

$$\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2\right\}^{1/2} = \sigma_y + n^{-1/2}G_n\left(\frac{1}{2\sigma_y}\left\{\pi_2^2 - 2\mu_y\pi_2\right\}\right) + o_P(n^{-1/2}).$$

We arrive at

$$\begin{aligned} B_n &= \left\{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right\}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2\right\}^{1/2} \\ &= \sigma_x\sigma_y + n^{-1/2}G_n(\frac{\sigma_y}{2\sigma_x}\left\{\pi_1^2 - 2\mu_x\pi_1\right\} + \frac{\sigma_x}{2\sigma_y}\left\{\pi_2^2 - 2\mu_y\pi_2\right\}) + o_P(n^{-1/2}). \end{aligned}$$

By setting

$$H_2(x,y) = \frac{\sigma_y}{2\sigma_x}\left\{\pi_1^2 - 2\mu_x\pi_1\right\} + \frac{\sigma_x}{2\sigma_y}\left\{\pi_2^2 - 2\mu_y\pi_2\right\},$$

we have

$$B_n = \sigma_x\sigma_y + n^{-1/2}G_n(H_2) + n^{-1/2}. \tag{10}$$

Now, combining (9) and (10) and using Lemma 1 yield

$$\sqrt{n}(\rho_n^2 - \rho^2) = n^{-1/2}G_n\left(\frac{1}{\sigma_x\sigma_y}H_1 - \frac{\sigma_{xy}}{\sigma_x^2\sigma_y^2}H_2\right) + o_P(1).$$

Let

$$H = \frac{1}{\sigma_x\sigma_y}(p(x,y) - \mu_x\pi_2 - \mu_y\pi_1) - \frac{\rho}{\sigma_x\sigma_y}\left\{\frac{1}{2\sigma_x^2}\left\{\pi_1^2 - 2\mu_x\pi_1\right\} + \frac{1}{2\sigma_y^2}\left\{\pi_2^2 - 2\mu_y\pi_2\right\}\right\}.$$

Now we continue with the centered and normalized case to get

$$H(x,y) = p(x,y) - \frac{\rho}{2}(\pi_1^2 + \pi_2^2)$$

and

$$H(X,Y) = XY - \frac{\rho}{2}(X^2 + Y^2).$$

Denote

$$\mu_{(p,x),(q,y)} = E((X-\mu_x)^p(Y-\mu_y)^q).$$

We have

$$EH(X,Y) = \sigma_{xy} - \rho = 0$$

and $varH(X,Y)$ is equal to

$$\mu_{(2,x),(2,y)} + \rho^2(\mu_{4,x} + \mu_{4,y})/4 - \rho(\mu_{(3,x),(1,y)} + \mu_{(1,x),(3,y)}) + \rho^2\mu_{(2,x),(2,y)}/2$$

and finally $varH(X,Y) = \sigma_0^2$ with

$$\sigma_0^2 = (1 + \rho^2/2)\mu_{(2,x),(2,y)} + \rho^2(\mu_{4,x} + \mu_{4,y})/4 - \rho(\mu_{(3,x),(1,y)} + \mu_{(1,x),(3,y)}).$$

This gives the conclusion that for centered and normalized X and Y,

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma_0^2).$$

Next, if we use the normalizing coefficients in σ_0, we get

$$\begin{aligned}\sigma^2 &= \sigma_x^2\sigma_y^2(1 + \rho^2/2)\mu_{(2,x),(2,y)} + \rho^2(\sigma_x^4\mu_{4,x} + \sigma_y^4\mu_{4,y})/4 \\ &\quad -\rho(\sigma_x^3\sigma_y\mu_{(3,x),(1,y)} + \sigma_x\sigma_y^3\mu_{(1,x),(3,y)})\end{aligned}$$

and we conclude in the general case that

$$\sqrt{n}(\rho_n - \rho) \rightsquigarrow N(0, \sigma^2).$$

The proof of Theorem 2 follows by easy computations under the particular conditions of ρ and under independence. ■

3. The Lo's residual functional emprical process *(lrfep)* and its use in asymptotic theory

Before we begin introducing the *lrfep*, we need additional notions and notation around the Rényi's representation. Next we will explain how the *lrfep* originated from the study ot the weak convergence of statistics. Finally, we will use it to get the full *GAR* expression of statistics in general.

3.1. Additional notation

In this part, we complete the notations we already gave and specify our probability space.

The notation below concerns the univariate random distributions. We are going to describe the general Gaussian field in which we present our results. Indeed, we use a unified approach when dealing with the asymptotic theories of the welfare statistics. It is based on the Functional Empirical Process (*fep*) and its Functional Brownian Bridge (*fbb*) limit. It is laid out as follows.

When we deal with the asymptotic properties of one statistic or index at a fixed time, we suppose that we have a non-negative random variable of interest which may be the income or the expense X whose probability law on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the Borel measurable space on $\mathbb{R}$, is denoted by $\mathbb{P}_X$. We consider the space $\mathcal{F}_{(1)}$ of measurable real-valued functions f defined on $\mathbb{R}$ such that

$$V_X(f) = \int (f - \mathbb{E}_X(f))^2 d\mathbb{P}_X = \mathbb{E}(f(X) - \mathbb{E}(f(X))^2 < +\infty,$$

where

$$\mathbb{E}_X(f) = \mathbb{E}f(X).$$

On this functional space $\mathcal{F}_{(1)}$, which is endowed with the L_2-norm

$$\|f\|_2 = \left(\int f^2 d\mathbb{P}_X\right)^{1/2},$$

we define the Gaussian process $\{\mathbb{G}_{(1)}(f), f \in \mathcal{F}_{(1)}\}$, which is characterized by its variance-covariance function

$$\Gamma_{(1)}(f,g) = \int^2 (f - \mathbb{E}_X(f))(g - \mathbb{E}_X(g))d\mathbb{P}_X, (f,g) \in \mathcal{F}^2_{(1)}. \tag{11}$$

This Gaussian process is the asymptotic weak limit of the sequence of functional empirical processes (fep) defined as follows. Let $X_1, X_2, ...$ be a sequence of independent copies of X. For each $n \geq 1$, we define the functional empirical process associated with X by

$$\mathbb{G}_{n,(1)}(f) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (f(X_j) - \mathbb{E}f(X_j)), f \in \mathcal{F}_{(1)},$$

and denote the integration with respect to the empirical measure by

$$\mathbb{P}_{n,(1)}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i), \ f \in \mathcal{F}_{(1)},$$

Let us denote by $\ell^{\infty}(T)$ the space of real-valued bounded functions defined on $T = \mathbb{R}$ equipped with its uniform topology. In the terminology of the weak convergence theory, the sequence of objects $\mathbb{G}_{n,(1)}$ weakly converges to $\mathbb{G}_{(1)}$ in $\ell^{\infty}(\mathbb{R})$, as stochastic processes indexed by $\mathcal{F}_{(1)}$, whenever it is a Donsker class. The details of this highly elaborated theory may be found in **?**, Pollard (1984), van der Vaart and Wellner (1996) and similar sources.

Here, we only need the convergence in finite distributions which is a simple consequence of the multivariate central limit theorem, as described in Chapter 3 in Lo *et al.* (2016).

We will use the Rényi's representation of the random variable X_i's of interest by means of the (*cdf*) $F_{(1)}$ of X as follows

$$X =_d F_{(1)}^{-1}(U),$$

where U is a uniform random variable on $(0, 1)$, $=_d$ stands for the equality in distribution and $F_{(1)}^{-1}$ is the generalized inverse of $F_{(1)}$, defined by

$$F_{(1)}^{-1}(s) = \inf\{x, F_{(1)}(x) \geq s\}, \ s \in (0, 1).$$

Based on these representations, we may and do assume that we are on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ holding a sequence of independent $(0, 1)$-uniform random variables U_1, U_2, ..., and the sequence of independent observations of X are given by

$$X_1 = F_{(1)}^{-1}(U_1), \quad X_2 = F_{(1)}^{-1}(U_2), \quad etc. \tag{12}$$

For each $n \geq 1$, the order statistics of $U_1, ..., U_n$ and of $X_1, ..., X_n$ are denoted respectively by $0 \equiv U_{0,n} < U_{1,n} \leq \cdots \leq U_{n,n} < U_{n+1,n} \equiv 1$ and $-\inf \equiv X_{0,n} < X_{1,n} \leq \cdots \leq X_{n,n} < X_{n+1,n}$ ($0 = -\inf \equiv X_{0,n}$ for a non-negative random variable X).

We also associate the sequence of $(U_n)_{n\geq 1}$ with the sequence of real-argument empirical functions

$$\begin{aligned}\mathbb{U}_{n,(1)}(s) &= \frac{1}{n}\#\{j, 1 \leq j \leq n, \ U_j \leq s\}, \ s \in (0, 1) \ n \geq 1 &(13)\\ &= \sum_{j=1}^{n} \frac{j}{n} 1_{(U_{j,n} \leq s < U_{j+1,n})}, &(14)\end{aligned}$$

and the sequence of real-argument uniform quantile function

$$\mathbb{V}_{n,(1)}(s) = U_{1,n}1_{(s=0)} + \sum_{j=1}^{n} U_{j,n}1_{((j-1)/n<s\leq(j/n))}, \ s \in (0,1), \ n \geq 1 \tag{15}$$

and next, the sequence of real-argument uniform empirical processes

$$\alpha_{n,(1)}(s) = \sqrt{n}(\mathbb{U}_{n,(1)} - s), \tag{16}$$

for $s \in (0,1)$ and ≥ 1, and the sequence of real-argument uniform quantile processes

$$\gamma_{n,(1)}(s) = \sqrt{n}(s - \mathbb{V}_{n,(1)}), \ s \in (0,1) \ n \geq 1. \tag{17}$$

The same can be done for the sequence $(X_n)_{n\geq 1}$, and we obtain the associated sequence of real empirical processes as

$$\mathbb{G}_{n,r,(1)}(x) = \sqrt{n}\left(\mathbb{F}_{n,(1)}(x) - F_{(1)}(x)\right), \ x \in \mathbb{R}, \ n \geq 1, \tag{18}$$

where

$$\mathbb{F}_{n,(1)}(x) = \frac{1}{n}\#\{j, 1 \leq j \leq n, \ X_j \leq x\}, \ x \in \mathbb{R} \ n \geq 1 \tag{19}$$

is the associated sequence of empirical functions. We also have the associated sequence of quantile processes

$$\mathbb{Q}_{n,(1)}(x) = \sqrt{n}\left(\mathbb{F}^{-1}_{(n),(1)}(s) - F^{-1}(s)\right), \ s \in (0,1), \ n \geq 1 \tag{20}$$

where, for $n \geq 1$,

$$\mathbb{F}^{-1}_{n,(1)}(s) = X_{1,n}1_{(0\leq s\leq 1/n)} + \sum_{j=1}^{n} X_{j,n}1_{((j-1)/n\leq s\leq(j/n))}, \ s \in (0,1), \tag{21}$$

is the associated sequence of quantile processes.

By passing, we recall that $\mathbb{F}^{-1}_{n,(1)}$ is actually the generalized inverse of $\mathbb{F}_{(n),(1)}$ and for the uniform sequence, we have

$$\mathbb{V}_{n,(1)} = \mathbb{U}^{-1}_{n,(1)}. \tag{22}$$

In virtue of Representation (12), we have the following remarkable relations

$$\mathbb{G}_{n,r,(1)}(x) = \alpha_{n,(1)}(F_{(1)}(x)), \ x \in \mathbb{R} \tag{23}$$

and

$$\mathbb{Q}_{n,(1)}(x) = \sqrt{n}\left(F_{(1)}^{-1}(\mathbb{V}_{n,(1)}(s)) - F_{(1)}^{-1}(s)\right)\ s \in (0,1),\ n \geq 1. \tag{24}$$

We also have the following relations between the empirical functions and quantile functions

$$\mathbb{F}_{n,(1)}(x) = \mathbb{U}_{n,(1)}(F_{(1)}(x)),\ x \in \mathbb{R} \tag{25}$$

and

$$\mathbb{F}_{n,(1)}^{-1}(s) = F_{(1)}^{-1}(\mathbb{V}_{(n),(1)}(s)),\ s \in (0,1),\ n \geq 1. \tag{26}$$

As well, the real-argument and function-argument empirical processes are related as follows: for $n \geq 1$,

$$\mathbb{G}_{n,r,(1)}(x) = \mathbb{G}_{n,(1)}(f_x^*),\ \alpha_{n,(1)}(s) = \mathbb{G}_{n,(1)}(\tilde{f}_s),\ s \in (0,1),\ x \in \mathbb{R}, \tag{27}$$

where for any $x \in \mathbb{R}$, $f_x^* = 1_{]-\infty,x]}$ is the indicator function of $]-\infty, x]$ and for $s \in (0,1)$, $f_s = 1_{[0,s]}$ and $\tilde{f}_s = 1_{]-\infty,F_{(1)}^{-1}(s)]}$.

To finish the description, a result of Kiefer-Bahadur (See Bahadur (1966)) that says that the addition of the sequences of uniform empirical processes and quantiles processes (16) and (17) is asymptotically, and uniformly on $[0,1]$, zero in probability, that is

$$\sup_{s\in[0,1]} \left|\alpha_{n,(1)}(s) + \gamma_{n,(1)}(s)\right| = o_{\mathbb{P}}(1) \text{ as } n \to +\infty. \tag{28}$$

This result is a powerful tool to handle the rank statistics when our studied statistics are L-statistics.

3.2. The origin of the lrfep

The method is developed here for a continuous *cdf* $F_{(1)}$. There is a considerable class of statistics which are combinations of one dimensional statistics of the form

$$L_n = d_n \sum_{1\leq j\leq n} c(j,n) q_0(X_{j,n}),\ n \geq 1,$$

where q_0 is some measurable mapping, $c(\circ, n)$ a function of $j \in \{1, \cdots, n\}$ and $(d_n)_{n\geq 1}$ is a sequence of real numbers. If $F_{(1)}$ is continuous, we may use the rank statistics $(R_{1,n}, \cdots, R_{n,n})$ defined by

$$\forall 1 \leq i \leq n,\ \forall 1 \leq j \leq n,\ R_{j,n} = i \Leftrightarrow X_{i,n} = X_j.$$

But it happens that for any $n \geq 1$, for any $1 \leq j \leq n$,

$$\frac{R_{j,n}}{n} = \mathbb{F}_{n,(1)}(X_j),$$

and this leads to

$$L_n = \frac{1}{n} \sum_{1 \leq j \leq n} \biggl(nd_n c\left(n\mathbb{F}_{n,(1)}(X_j)\right)\biggr) q_0(X_j), \ \ n \geq 1.$$

The question is: how to treat such statistics? Obviously, this is a L-statistic and the theory of Chernoff might help. However, we will use another alternative developed over the years since that uses the so-called *lrfep*. The method is based on the following facts: In many cases, there exists a measurable mapping q_1 such that $\mathbb{E}q_1(X) < +\infty$ and we may write

$$\begin{aligned} L_n &= \frac{1}{n} \sum_{1 \leq j \leq n} q_1(F_{(1)}(X_j)) q_0(X_j) \\ &+ \frac{1}{n} \sum_{1 \leq j \leq n} \biggl(nd_n c\left(n\mathbb{F}_{n,(1)}(X_j)\right) - q_1(F_{(1)}(X_j))\biggr) q_0(X_j) \end{aligned}$$

in a way that, by using the mean value theorem, we arrive at

$$\begin{aligned} &\frac{1}{n} \sum_{1 \leq j \leq n} \biggl(nd_n c\left(n\mathbb{F}_{n,(1)}(X_j)\right) - q_1(F_{(1)}(X_j))\biggr) q_0(X_j) \\ &= \frac{1}{n} \sum_{1 \leq j \leq n} \biggl\{\mathbb{F}_{n,(1)}(X_j) - F_{(1)}(X_j)\biggr\} \ q_1(X_j) q_0(X_j) + o_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

where $q_1(\circ)$ is some measurable function, and further under specific conditions to be found and checked, we finally have the form, for $q = q_0 q_1$,

$$L_n = \frac{1}{n} \sum_{1 \leq j \leq n} h(X_j) + \frac{1}{n} \sum_{1 \leq j \leq n} \biggl\{\mathbb{F}_{n,(1)}(X_j) - F_{(1)}(X_j)\biggr\} \ q(X_j) + o_{\mathbb{P}}(n^{-1/2}).$$

We conclude that, in our effort to asymptotically represent L_n as an application of the empirical measure to some function h, that is $\mathbb{P}_n(h)$, we still have a residual term in the form of

$$Re_n(\ell) = \frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{F}_{n,(1)}(X_j) - F_{(1)}(X_j)\right) q(X_j).$$

This made Lo and Sall (2010a) to name it a residual empirical process and proceeded to its independent study. Now let us describe this stochastic process more deeply.

3.3. The partial GAR of lrfep and conditions of its validity

3.3.1. The main features of the *lrfep*

A residual empirical process is any stochastic process of the form

$$Re_n(\ell) = \frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{F}_{n,(1)}(X_j) - F_{(1)}(X_j)\right) q(X_j).$$

where q is a measurable function from $[0,1]$ to $\mathbb{R}$. Let us define

$$\ell(s) = -q(F_{(1)}^{-1}(s)),\ \ s \in (0,1)$$

and

$$\Delta_n(q,s) = \left\{ q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right) - q\left(F_{(1)}^{-1}(s)\right)\right\},\quad s \in (0,1).$$

We stress that the function ℓ depends of the *cdf* $F_{(1)}$ and should bave been denoted $\ell(\circ) = \ell(F_{(1)}, \circ)$. This warning is important in the situation of spatial analysis, as we will see it.

3.3.2. The general results on the **lrfep**

We have the following results.

Theorem 3. *Let $F_{(1)}$ be continuous. If the following two assertions:*

(1) $\mathbb{E}q(X) < +\infty$ (CRe1)

and,

(2) and, as $n \to +\infty$,

$$\int_0^1 \sqrt{n}\left(s - \mathbb{V}_{n,(1)}(s)\right)\Delta_n(q,s)\, ds \to 0 \ \ (CRe2)$$

hold, we have the representation

$$\sqrt{n}Re_n(\ell) = \int_0^1 \mathbb{G}_{n,(1)}(\tilde{f}_s)\,\ell(s)\,ds + o_p(1).$$

Proof. By using Formulas (13) and (15), we get

$$Re_n = \sum_{j=1}^{n} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \left\{\mathbb{F}_{n,(1)}(\mathbb{F}_{n,(1)}^{-1}(s)) - F_{(1)}(\mathbb{F}_{n,(1)}^{-1}(s))\right\} q\left(\mathbb{F}_{n,(1)}^{-1}(s)\right) ds,$$

and hence

$$Re_n = \int_0^1 \left\{\mathbb{F}_{n,(1)}(F_{n,(1)}^{-1}(s)) - F_{(1)}(\mathbb{F}_{n,(1)}^{-1}(s))\right\} q\left(\mathbb{F}_{n,(1)}^{-1}(s)\right) ds. \tag{29}$$

Before we proceed further, we will get some facts for Formulas (22), (23) and (24). We already have

$$\mathbb{F}_{n,(1)}(\circ) = \mathbb{U}_{n,(1)}\biggl(F_{(1)}(\circ)\biggr), \quad and \;\; \mathbb{F}_{n,(1)}^{-1}(\circ) = F_{(1)}^{-1}\biggl(\mathbb{V}_{n,(1)}(\circ)\biggr),$$

Next, we have (See page 148, Lo *et al.* (2016))

$$F_{(1)}\biggl(F_{(1)}^{-1}(\circ)) - 0\biggr) \leq \circ \leq F_{(1)}\biggl(F_{(1)}^{-1}(\circ)) + 0\biggr).$$

Next, by continuity of $F_{(1)}$, we have

$$\mathbb{F}_{(1)}\biggl(\mathbb{F}_{n,(1)}^{-1}(\circ)\biggr) = \mathbb{V}_{n,(1)}(\circ)$$

and

$$\mathbb{F}_{n,(1)}\biggl(\mathbb{F}_{n,(1)}^{-1}(\circ)\biggr) = \mathbb{U}_{n,(1)}\biggl(\mathbb{V}_{n,(1)}(\circ)\biggr).$$

With this facts, we may handily transform Re_n as follows:

$$\begin{aligned}
\sqrt{n}Re_n &= \int_0^1 \sqrt{n}\left\{\mathbb{U}_{n,(1)}\left(\mathbb{V}_{n,(1)}(s)\right) - \mathbb{V}_{n,(1)}(s)\right\} q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right) ds \\
&= \int_0^1 \sqrt{n}\left(s - \mathbb{V}_{n,(1)}(s)\right) q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right) ds \\
&+ \int_0^1 \sqrt{n}\left(\mathbb{U}_{n,(1)}\left(\mathbb{V}_{n,(1)}(s)\right) - s\right) q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right) ds \\
&=: Re_n(1) + Re_n(2).
\end{aligned}$$

From Shorack and Wellner (1986) (page 585), we have

$$\sup_{0\leq s\leq 1}\left|\mathbb{U}_{n,(1)}\left(\mathbb{V}_{n,(1)}(s)\right)-s\right|\leq \frac{1}{n}.$$

We get

$$\begin{aligned}|Re_n(2)| &\leq \frac{1}{\sqrt{n}}\int_0^1 q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right)\,ds\\ &= \frac{1}{\sqrt{n}}\left(\frac{1}{n}\sum_{j=1}^{n} q(X_j)\right), \qquad (30)\end{aligned}$$

which is an $o_{\mathbb{P}}(n^{-1/2})$ whenever $\mathbb{E}q(X)$ is finite. We obtain

$$\sqrt{n}Re_n=\int_0^1 \sqrt{n}\left(s-\mathbb{V}_{n,(1)}(s)\right)q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right)\,ds+o_p(1). \qquad (31)$$

Next, we try to replace $q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right)$ by $q\left(F_{(1)}^{-1}(s)\right)$. We will be led to

$$\begin{aligned}\sqrt{n}Re_n &= \int_0^1 \sqrt{n}\left(s-\mathbb{V}_{n,(1)}(s)\right)q\left(F_{(1)}^{-1}(s)\right)\,ds \qquad (32)\\ &+ \int_0^1 \sqrt{n}\left(s-\mathbb{V}_{n,(1)}(s)\right)\left\{q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right)-q\left(F_{(1)}^{-1}(s)\right)\right\}ds+o_p(1)\\ &=: \int_0^1 \sqrt{n}\left(s-\mathbb{V}_{n,(1)}(s)\right)q\left(F_{(1)}^{-1}(s)\right)\,ds+\int_0^1 \sqrt{n}\left(s-\mathbb{V}_{n,(1)}(s)\right)\Delta_n(q,s)\,ds+o_p(1).\end{aligned}$$

Finally by the the Bahadur's representation (28) and by Condition (Re2), we finish the proof in the form:

$$\begin{aligned}\sqrt{n}Re_n &= \int_0^1 \gamma_{n,(1)}(s)q\left(F_{(1)}^{-1}(s)\right)\,ds+o_p(1) \qquad (33)\\ &= \int_0^1 \gamma_{n,(1)}(s)+\alpha_{n,(1)}q\left(F_{(1)}^{-1}(s)\right)\,ds\\ &- \int_0^1 \alpha_{n,(1)}(s)q\left(F_{(1)}^{-1}(s)\right)\,ds+o_p(1)\\ &= -\int_0^1 \left(\alpha_{n,(1)}(s)\right)q\left(F_{(1)}^{-1}(s)\right)\,ds+o_p(1)\\ &= \int_0^1 \mathbb{G}_{n,(1)}(\tilde{f}_s)\left(-q\left(F_{(1)}^{-1}(s)\right)\right)\,ds+o_p(1),\end{aligned}$$

i.e,

$$\sqrt{n}Re_n = \int_0^1 \mathbb{G}_{n,(1)}(\tilde{f}_s)\,\ell(s)\,ds + o_p(1),$$

whenever

$$\mathbb{E}(\ell(X)) = -\mathbb{E}q(X) = -\int_0^1 q(F_{(1)}^{-1}(s))\,ds < +\infty.$$

This concludes the proof.

3.4. *The final GAR combining the* ***fep*** *and the* ***lrfep***

Summary of the general method.

We suppose that we study the sequence of statistics J_n. We split the study into steps.

Step 1. Proceed to stochastic calculus and manipulations to try to write as:

$$J_n = \mathbb{E}h(X) + n^{-1/2}\mathbb{G}_n(h) + Re_n(\ell) + o_{\mathbb{P}}(n^{-1/2}), \tag{34}$$

where $\ell(\circ)$ is derived from some $q(x)$ measurable function of $x \in \mathbb{R}$ by $\ell(s) = -q(F_{(1)}^{-1}(s))$, for $(0,1)$, and

$$Re_n(\ell) = \frac{1}{n}\sum_1^n \{\mathbb{F}_{(n),(1)}(X_j) - \mathbb{F}_{(1)}(X_j)\}\; q(X_j), \quad n \geq 1. \tag{35}$$

We recall that

$$\Delta_n(q,s) = \left\{ q\left(F_{(1)}^{-1}\left(\mathbb{V}_{n,(1)}(s)\right)\right) - q\left(F_{(1)}^{-1}(s)\right)\right\}, \quad s \in (0,1). \tag{36}$$

Step 2. Using Theorem 3, we try to get

$$\mathbb{E}q(X) < +\infty \tag{37}$$

and,

(2) as $n \to +\infty$,

$$\int_0^1 \sqrt{n}\left(s - \mathbb{V}_{n,(1)}(s)\right)\Delta_n(s)\,ds \to 0 \tag{38}$$

If so, we get the *GAR* that will be displayed below.

Step 2 (alternative) If we are able to prove that

$$\frac{1}{\sqrt{n}}\left(\frac{1}{n}\sum_{j=1}^{n} q(X_j)\right) = o_P(1),\ as\ n \to +\infty, \tag{39}$$

and

$$\int_0^1 (s(1-s))^{1/2}\Delta_n(s) = o_P(1),\ as\ n \to +\infty, \tag{40}$$

hold, then the general *GAR* is true. That *GAR* is:

$$J_n = \mathbb{E}h(X) + n^{-1/2}\mathbb{G}_n(h) + n^{-1/2}\int_0^1 \mathbb{G}_n(\tilde{f}_s)\ \ell(s)\ ds + o_{\mathbb{P}}(n^{-1/2}), \tag{41}$$

with

$$\tilde{f}_s = 1_{]-\infty, F_{(1)}^{-1}(s)]},\ s\ (0,1). \tag{42}$$

$$\sqrt{n}(J_n - \mathbb{E}h(X)) = \mathbb{G}_n(h) + \int_0^1 \mathbb{G}_n(\tilde{f}_s)\ \ell(s)\ ds + o_{\mathbb{P}}(1). \tag{43}$$

To understand Condition (40), we used the

$$\sup_{s\in(0,1)} \left\| \frac{\sqrt{n}\left(s - \mathbb{V}_{n,(1)}(s)\right)}{(s(1-s))^{1/2}} \right\| = O_{\mathbb{P}}(1).$$

Now, we may conclude. The finite-distribution asymptotic law of the field $\mathbb{G}_n(f),\ f \in L^2(\mathbb{P}_X)$ gives the limit law of $\sqrt{n}(J_n - \mathbb{E}h(X))$, as follows.

Theorem 4. *If $h \in L^2(\mathbb{P}_X)$ and, (37) or (38) and (39) or (40) hold, then the sequence of indexes J_n as in (41) satisfies*

$$\sqrt{n}(J_n - \mathbb{E}h(X)) \rightsquigarrow \mathcal{N}(0,\ \Gamma^{(J)}),\ \ as\ \ n \to +\infty,$$

where

$$\Gamma_{(1)}(h_1, h_2) = \int (h_1 - \mathbb{E}h_1(X))(h_2 - \mathbb{E}h_2(X))\ d\mathbb{P}_X \tag{44}$$

and

$$\begin{aligned}
&\Gamma^{(J)} = \gamma_1 + \gamma_2 + 2\gamma_3, \\
&\gamma_1 = \Gamma_{(1)}(h,h); \\
&\gamma_2 = \int_0^1 \int_0^1 \Gamma_{(1)}(\tilde{f}_s, \tilde{f}_t)\ell(s)\ \ell(t)\ ds\ dt; \\
&\gamma_3 = \int_0^1 \Gamma_{(1)}(h, \tilde{f}_s)\ell(s)\ ds.
\end{aligned} \tag{45}$$

(R) Important Remarks on the existence of the variances and co-variance in (44) and (45).

(A1) Asymptotic law of

$$Re_n(\ell) = \int_0^1 \mathbb{G}_n(\tilde{f}_s)\ \ell(s)\ ds.$$

Since $\tilde{f}_s$ and $\tilde{f}_t$ are in $[0,1]$, that asymptotic variance of $Re_n(\ell)$, is finite

$$\mathbb{E}(q(X)^2) < +\infty. \quad (Th33)$$

(A2) Asymptotic covariance between $\mathbb{G}_n(h)$ and $Re_n(\ell)$. For the same reason, that Asymptotic covariance is finite if

$$\mathbb{E}(h(X)^2 q(X)^2) < +\infty. \quad (Th12)$$

(A3) Asymptotic variance of J_n . Finally, that asymptotic variance is finite if (Th22), (Th12) and

$$\mathbb{E}h(X)^2 < +\infty. \quad (Th11)$$

4. Finding the GAR in Gini's Case study

Before we begin, we want to draw the attention of the reader who wants to work on this topic about the following warning.

Warning. The method we are presenting here uses a lot of stochastic expansions over quantities as $o_{\mathbb{P}}(1)$ and $O_{\mathbb{P}}(1)$. The researcher should have to try to get a good command in such computation skills. He may need to go to the bottom of that by revising Chapter 6 in Lo *et al.* (2016). Here, in the present study, we do not dwell in the details.

Now, let us consider the Gini index of inequality as in Mergane *et al.* (2018b) (where the multiplicative coefficient (2) has been corrected to (1))

$$J_n = \frac{1}{\frac{1}{n}\sum_{j=1}^n X_j}\left\{\frac{1}{n}\sum_{j=1}^{n}\left(\frac{2j-1}{n}-1\right)X_{j,n}\right\},$$

for $n \geq 1$ and the notations and terminology given above are adopted. To apply the method we expose above, we see that we have to treat the two factors

$$A_n = \frac{1}{\frac{1}{n}\sum_{j=1}^n X_j} \quad and \quad B_n = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{2j-1}{n}-1\right)X_{j,n}$$

separately. The first step is easy to handle, for $h_1(x) = x$ and by denoting $\mathbb{E}h_1(X) = \mathbb{E}X = \mu$,

$$\begin{aligned}
A_n &= \frac{1}{\frac{1}{n}\sum_{j=1}^n X_j}\\
&= \frac{1}{\mathbb{E}h_1(X) + n^{-1/2}\mathbb{G}_n(h_1)}\\
&= \frac{1}{\mu}\left(1+\mu^{-1}n^{-1/2}\mathbb{G}_n(h_1)\right)^{-1}\\
&= \frac{1}{\mu}\left(1-\mu^{-1}n^{-1/2}\mathbb{G}_n(h_1)+O_{\mathbb{P}}(n^{-1/2})\right).
\end{aligned}$$

We get a first *GAR*

$$A_n = \frac{1}{\mu} - n^{-1/2}\mathbb{G}_n(\mu^{-2}h_1) + o_{\mathbb{P}}(n^{-1/2}).$$

As to B_n, we remark that

$$\begin{aligned}
B_n &= \left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{2j}{n}-1\right)X_{j,n}\right)-\left(\frac{1}{n^2}\sum_{j=1}^{n}X_j\right)\\
&=: B_{n,1}+B_{n,2}.
\end{aligned}$$

We have

$$\sqrt{n}B_{n,2} = n^{-1}(\mu + o_{\mathbb{P}}(1)) = o_{\mathbb{P}}(n^{-1/2}). \tag{46}$$

Next,

$$\begin{aligned}
B_{n,1} &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{2j}{n}-1\right)X_{j,n}\Bigg)\\
&= \frac{1}{n}\sum_{j=1}^{n}\Bigg(2F_n(X_{j,n})-1\Bigg)X_{j,n}\\
&= \frac{1}{n}\sum_{j=1}^{n}\Bigg(2F_n(X_j)-1\Bigg)X_j\\
&= \frac{1}{n}\sum_{j=1}^{n}\Bigg(2F(X_j)-1\Bigg)X_j\\
&+ \frac{1}{n}\sum_{j=1}^{n}\Bigg(F_n(X_j)-F(X_j)\Bigg)2X_j\\
&= \mathbb{P}_n(H) + Re_n(\ell_1)
\end{aligned}$$

with

$$H(x) = (2F(x)-1)x \quad \textit{and} \quad \ell_1(x) = -2F^{-1}(s), \quad x\in\mathbb{R}, \quad s\in(0,1).$$

So, upon checking Conditions (37) or (38) and (39) or (40), we would get

$$B_n = \mathbb{E}H(X) + n^{-1/2}\mathbb{G}_n(H) + n^{-1/2}\int_0^1 \mathbb{G}_n(\tilde{f}_s)\ \ell_1(s)\ ds + o_{\mathbb{P}}(n^{-1/2}) \tag{47}$$

Finally, algebraic manipulations of the product of the expansions of A_n and B_n, based on $\mathbb{G}_n(\circ) = o_{\mathbb{P}}(1)$ anywhere here, lead to

$$J_n = \frac{\mathbb{E}H(X)}{\mu} + n^{-1/2}\mathbb{G}_n\left(\frac{H}{\mu} - \frac{\mathbb{E}H(X)h_1}{\mu^2}\right) + n^{-1/2}\int_0^1 (1/\mu)\mathbb{G}_n(\tilde{f}_s)\ \ell_1(s)\ ds + o_{\mathbb{P}}(n^{-1/2}),$$

which gives us our final *GAR*

$$J_n = \frac{\mathbb{E}H(X)}{\mu} + n^{-1/2}\mathbb{G}_n(h) + n^{-1/2}\int_0^1 \mathbb{G}_n(\tilde{f}_s)\ell(s)\ ds + o_{\mathbb{P}}(n^{-1/2}), \tag{48}$$

where

$$h(x) = \frac{\mu H(x) - \mathbb{E}H(X)h_1(x)}{\mu^2}, \tag{49}$$

$$h_1(x) = x, \tag{50}$$

$$H(x) = (2F(x)-1)x,$$

$$\ell(s) = \frac{-2F^{-1}(s)}{\mu}, \quad x \in \mathbb{R}, \quad s \in (0,1).$$

5. Conclusion, perspectives and applications

The theory reaches two kinds of results.

(R1) A definitive result of asymptotic normality as in Theorem 4 that can be applied in statistical estimation by confidence intervals and in statistical tests.

(R2) A progressive result in the form of the asymptotic representation as in (41) (page 3544) can then be used for comparison with other statistics or with values of the same statistic for other areas, or for following the evolution of the values of the statistics other times in a way to ensure convergence of joint laws.

The following remarks are important to be noticed.

(R3) Let us be aware that the final result may come after algebraic manipulations of a number of *GAR* founded separately.

(R4) The computations of variances and co-variances of the kind of statistics used here are well-handled with current computers and numerical methods with weak execution times.

(R5) To be completely useful to researchers and practitioners, a package should be available and ready-to-be used. Such a work is underway and will be reported on soon.

References

Bahadur R.R. (1966). A note on quantiles in large samples, *Ann. Math. Stat.* **37**, pp. 577–580. MR :32:6522. ZL: 0147.18805. doi:10.1214/aoms/1177699450. euclid.aoms/1177699450.

Billingsley P. (1968). *Convergence of probability measures*. Wiley, new-York.

Csőrgő, H., Csőrő, M., Mason D.M. and Horvàth, L. (1986) Weighted empirical process and quantile process. *Ann. Probab.* 14 (1), 31-85.

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

Gaenssler, P. (1983). *Empirical processes*. IMS Lecture Notes-Monograph Series, Hayward, California.

Haidara M.C, Kpanzou T.A, Mergane P.D. and Lo G.S. (2018). Asymptotic Theory and Statistical Decomposability gap Estimation for Takayamas Index. In A Collection of Papers in Mathematics and Related Sciences, a festschrift in honour of the late Galaye Dia (Editors : Seydi H., Lo G.S. and Diakhaby A.). Spas Editions, Euclid Series Book, pp. 195 –220. Doi : 10.16929/sbs/2018.100-02-08.

Jarque C.M. and Bera A.K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*. 55(5): 163-172.

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

Lo G.S. (2013). On the General Poverty Index. *Far East Journal of Theoretical Statistics*. Volume 42. (1), 1-22.

Lo, G.S., Ngom M. and Kpanzou T.A. (2016). Weak Convergence (IA). Sequences of random vectors. SPAS Books Series. (2016). Doi : 10.16929/sbs/2016.0001. Arxiv : 1610.05415.

Lo G.S. and Sall S.T. (2010a). Asymptotic Representation Theorems for Poverty Indices. *Afrika Statistika, Special Volume (5) : Proceedings of the International Workshop on Multiple Risks and Copula, Biskra 2010, pp. 238-244. Ed. Abdelhakim Necir.*

Lo G.S. and Sall S.T. (2010b). Uniform weak convergence of the time-dependent poverty measures for continuous longitudinal data, *Brazilian Journal of Probability and Statistics*, Vol. 24, Issue 3, pp 457-467.

Lo G.S., Thiam, O. and Haidara C.M. (2015) High Moments Jarque-Bera Tests for Arbitrary Distribution Functions. *Applied Mathematics*. Vol. 6 (5), 707-716. http://dx.doi.org/10.4236/am.2015.64066.

Mergane P.D., Kpanzou T.A., Ba D. and Lo G.S. (2018a). A Theil-like class of inequality measures, its asymptotic normality theory and applications. *Afrika Statistika* Vol. 13 (3), pp 1699-1715.

Mergane P.D., Lo G.S. and Kpanzou T.A. (2018b). On the joint distribution of variations of the Gini index and Welfare indices. In A Collection of Papers in Mathematics and Related Sciences, a festschrift in honour of the late Galaye Dia. Spas Editions, Euclid Series Book, pp. 163– 194. Doi : 10.16929/sbs/2018.100-02-07.

Pollard D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, Berlin.

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

Shorack G.R. and Wellner J.A. (1986). Empirical Processes with Applications to Statistics, wiley-Interscience, New-York.

van der Vaart A.W. and Wellner J.A. (1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer, New-York.

Vapnik V.N. and Chervonenkis A.Y. (1971). On the uniform convergence of relative frequencies of event to their probabilities. *Theory of Probability and Its applications*. 16, 264-280.

Zheng B. (1997). Aggregate poverty measures. *Journal of Economic Surveys*, 11 123–162.