C. R. Math. Rep. Acad. Sci. Canada Vol. **31** (2) 2009, pp. 45–52

UNE THÉORIE GÉNÉRALE ASYMPTOTIQUE DES MESURES DE PAUVRETÉ

SERIGNE TOUBA SALL, CHEIKH TIDIANE SECK, ET GANE SAMB LO,

Présenté par M. Csörgő, FRSC

RÉSUMÉ. Dans cette note, nous proposons une forme très générale de l'indicateur de pauvreté de façon à inclure les indices disponibles dans la littérature. La théorie de la normalité asymptotique est ensuite établie dans sa globalité avec des conditions relativement douces sur la distribution des revenus ou de la dépense. Les résultats simulés avec satisfaction se révèlent efficaces pour l'évaluation et le suivi spatio-temporel de la pauvreté au moyen d'intervalles de confiance précis.

ABSTRACT. In this note, we introduce a Generalized form of the Poverty Index (GPI) including almost all the available ones in the literature as well as the general Exact asymptotic Poverty Index. The asymptotic normality theory is then established for the GPI when mild conditions on the distribution of the income or the expenditure variable are assumed. The results are conclusively simulated and turn out to be efficient for poverty monitoring (in time) and poverty comparison (in space) with the help of accurate confidence intervals.

1. **Introduction.** Cette étude a pour objet principal l'établissement des lois limites, ou asymptotiques, des indicateurs de pauvreté. Ceux-ci sont, dans leur grande majorité, recensés et étudiés du point de vue axiomatique [16]. Bien que nombreux et variés, nous pouvons les regrouper dans la forme générale,

$$J_N = \frac{1}{a(Q)b(N)} \sum_{j=1}^{Q} w(N,Q,j) d\Big(\frac{Z - Y_{j,N}}{Z}\Big), \tag{1.1}$$

où $a(\cdot), b(\cdot), d(\cdot), w(\cdot)$ sont des fonctions measurables, Q est le nombre total (inconnu) de pauvres dans la population P des N individus ou ménages étudiés, Z est la ligne de pauvreté et enfin $Y_{1,N} \leq Y_{2,N} \leq \cdots \leq Y_{N,N}$ sont les revenus ou dépenses ordonnés des membres de la population classés du plus pauvre au moins pauvre. *Il convient de noter dès le début que les termes indicateur, mesure ou indice de pauvreté sont couramment utilisés de manière équivalente dans ce texte et ailleurs.*

Reçu le 17 avril, 2009.

Classification (de l'AMS): Primary: 60F05; secondary: 60F15, 91C05.

Mots clés: mesures et indices de pauvreté, normalité asymptotique, estimation et simulation, processus empiriques, approximations hongroises de processus empiriques, théorie des valeurs extrêmes.

La ligne de pauvreté Z est fixée par les autorités ou les analystes économistes eux-mêmes de sorte que tout individu ou ménage ayant un revenu ou une dépense Y (disons revenu annuel ou dépense annuelle) en deça de Z est considéré comme pauvre. Le lecteur intéressé par cette question est renvoyé à [8]. Elle ne sera pas abordée dans cette étude méthodologique. Elle le sera forcément dans les applications appropriées de nos méthodes, par exemple [9].

A l'entame de cette introduction, il faut aussi signaler que la variable Y peut être le revenu ou la dépense. De même, l'individu de la population étudiée est en fait un ménage, ramené à un individu par une échelle d'équivalence-adulte. La variable Y considérée dans ce rapport, est bien l'équivalent adulte.

Avant de repréciser le champs de cette investigation, nous allons décrire la classe des indicateurs de pauvreté, qui peut se diviser en deux grandes sous-classes. La première concerne les indicateurs non pondérés, ceux pour lesquels le poids est l'unité, c'est-à-dire $w(N, Q, j) \equiv 1$. Dans cette sous-classe, l'indicateur le plus populaire et certainement le plus utilisé est celui de Foster, Greer, et Thorbecke (FGT) [5] défini pour $\alpha \geq 0$, par

$$J_{FGT}(\alpha) = \frac{1}{N} \sum_{j=1}^{Q} \Big(\frac{Z - Y_{j,N}}{Z} \Big)^{\alpha}. \tag{1.2}$$

Quand $\alpha = 0$, (1.2) se réduit à la prévalence de pauvreté Q/N, tandis que pour $\alpha = 1$ et $\alpha = 2$, il est respectivement interprété comme l'intensité et la sévérité de pauvreté. La deuxième catégorie d'indices concerne ceux qui sont pondérés. Mentionnons deux parmi eux. D'abord, la mesure de Sen [13]

$$J_{\mathrm{SEN},N} = \frac{2}{N(Q+1)} \sum_{j=1}^{Q} (Q - j + 1) \Big(\frac{Z - Y_{j,N}}{Z} \Big), \tag{1.3}$$

qui a été généralisée par Kakwani [7] à travers

$$J_{\mathrm{KAK},N}(k) = \frac{Q}{N \Phi_k(Q)} \sum_{j=1}^{Q} (Q - j + 1)^k \Big(\frac{Z - Y_{j,N}}{Z} \Big), \tag{1.4}$$

pour $k \geq 1$ et $\Phi_k(Q) = \sum_{j=1}^{j=Q} j^k$. La généralisation veut dire ici que la mesure de Sen est obtenue en faisant $k = 1$ dans (1.4), c'est-à-dire que $P_{\mathrm{SEN},N} = P_{\mathrm{KAK},N}(1)$. Ensuite Shorrocks [14] introduisit

$$P_{\mathrm{SH},N} = \frac{1}{N^2} \sum_{j=1}^{Q} (2N - 2j + 1) \Big(\frac{Z - Y_{j,N}}{Z} \Big). \tag{1.5}$$

Plusieurs autres indices existent dans la littérature, recensés en général dans [16]. L'étude de ces indicateurs, dans leur formes discrètes, suivant l'approche pionnière et axiomatique de Sen, est largement menée, dans le cadre de la réduction de la pauvreté. Les lecteurs intéressés sont orientés vers [13], [16], et [10]. Nous ne sommes pas directement concernés par cette approche dans notre présente investigation.

Par contre, notre préoccupation précise est l'établissement de résultats globaux de propriétés asymptotiques des indices de pauvreté, en vue d'une estimation efficace par intervalles de confiance de (1.1). En effet, le paramètre inconnu (1.1) doit être estimé par sondage. Il est alors important d'établir la normalité limite en vue d'obtenir des tests statistiques. Pourtant, il n'existe pas encore une théorie aymptotique normale cohérente afférente à ces indices. Certains travaux ont été consacrés à l'inférence statistique d'indices particuliers, par exemple [1]. Dans un autre registre, l'index FGT a fait l'objet de travaux dans ce sens, par Dia [4] en utilisant les processus ponctuels et par les auteurs [11], dans le cas particulier $\alpha = 1$ avec des fonctions de distribution du domaine extrêmal. Notre ambition est d'établir d'un seul coup une théorie asymptotique des indicateurs de pauvreté, en particulier définis par (1.2)–(1.5). Les méthodes utilisées contiennent au delà des résultats obtenus, une méthodologie applicable aux indices encore inconnus.

En résumé, nous offrons une approche unifiée, en ce sens que, tous les indicateurs sont à la fois étudiés et la forme considérée sous sa forme mathématique, englobe des indicateurs que les chercheurs sur le bien-être social pourraient déduire et proposer plus tard. Il faut noter cependant que l'indice de Kakwani n'est pas couvert par ces résultats. Mais le travail supplémentaire à faire pour le couvrir repose sur les preuves ci-bas exposées. Il sera complet dans un papier à venir. Dans une autre étude d'applications, une investigation à large échelle sera menée sur ESAM I et II (Enquêtes Sénégalaises Auprès de Ménages). Dans la présente étude, les résultats seront simulés et appliqués aux bases ESAM à titre illustratif.

Formulons maintenant notre problème. Nous tirons un échantillon de n individus totalement aléatoire, avec remise, de la population P, et observons la variable Y, si bien que nous obtenons $Y_1, Y_2, \ldots, Y_n$, n variables aléatoires indépendantes et identiquement distribuées selon la fonction de répartition G de Y, dont la borne inférieure de son support est $y_0 = \inf\{x, \mathrm{G}(x) > 0\} \geq 0$. Etant donné la ligne de pauvreté Z, le nombre aléatoire de pauvres dans l'échantillon est $q_n = q$. L'index échantillonné et aléatoire, considéré comme estimateur de (1.1), est alors

$$J_n = \frac{1}{n} \sum_{j=1}^{q} c(n,q,j) d\Big(\frac{Z - Y_{j,n}}{Z}\Big), \tag{1.6}$$

où $c(n,q,j) = nw(n,q,j)/\{a(n)b(q)\}$. Il est clair que (1.6) depend des q valeurs extrêmes inférieures. Si $q/n \to 0$, nous ne serions concernés que par la queue inférieure de G. Cette situation serait typique des pays nantis, avec très peu de pauvres. Le cadre statistique naturel adapté serait la théorie des valeurs extrêmes. Mais, pour les pays pauvres que nous analysons, la prévalence empirique est souvent supérieure à 50%, si bien que l'hypothèse raisonnable est la suivante :

$$q/n \to \xi = G(Z) \in]0,1[.$$

Donc, nous serons au plus concernés par la queue et le centre de la distribution. Cela est conforme à l'hypothèse de focalisation des indices de pauvreté qui stipule que ceux-ci sont des fonctions strictes des revenus ou dépenses des pauvres. Ainsi, nous ferons appel à deux types de conditions pour G. Pour le centre, nous ferons appel à la différentiabilité. A la queue, nos conditions seront inspirées par la théorie des valeurs extrèmes, sans en être limités.

Il est aussi commode de recourir à une transformation de $Y \geq 0$ en $X = 1/(Y - y_0)$ de fonction de répartition $F(\,\cdot\,) = 1 - G(y_0 + 1/\,\cdot\,)$, avec un support non borné à droite. Cette transformation permet aussi sur le plan pratique de trouver de bons ajustements de la loi des revenus. En effet, il a été remarqué que X est bien modélisé par la loi lognormale ou celle de Singh–Maddalla. Alors (1.6) devient

$$J_n = \frac{1}{n}\sum_{j=1}^{q} c(n,q,j)\, d\Big(\frac{Z - y_0 - X_{n-j+1,n}^{-1}}{Z}\Big). \tag{1.7}$$

Nous allons décrire le comportement asymptotique de la classe générale des indicateurs de pauvreté (1.7) avec des conditions relativement douces. Afin de bien comprendre les résultats exposés ici, il faut noter que (1.1) est la forme discrète générale de l'indicateur de pauvreté, utilisée dans l'approche axiomatique. Cependant, pour une population de très grande taille (presque infinie), cet indicateur ne peut être observé que sous un échantillon, souvent très grand, de sorte que le tirage puisse être considéré comme avec remise. Il devient, de ce point de vue, une variable aléatoire sous la forme (1.7). Nos résultats permettent alors de voir que (1.7) converge en probabilité vers une quantité de la forme

$$J = \int_0^{G(Z)} L(u,F)\, d\Big(\frac{Z - y_0 - F^{-1}(1-u)}{Z}\Big) dF,$$

où L est un poids exact correspondant au poids discret de l'indicateur empirique. Ainsi J s'interprête comme la forme générale de l'indicateur de pauvreté exacte induite par la loi G (donc F) du revenu ou de la dépense. Nos résultats permettent d'estimer J par J_n par intervalle de confiance à partir des résultats de normalité asymptotique, qui se révèlent assez précis pour de petites tailles grâce à des simulations conséquentes.

2. Nos rèsultats.

Nous aurons besoin des hypothèses suivantes. Une première série d'entre elles est rélative à la fonction $d(\,\cdot\,)$:

(D1) d$(\,\cdot\,)$ admet une dérivée sur $]0,1)$.

(D2) $d'(\frac{z-y_0}{z})$ et $d(\frac{z-y_0}{z})$ sont finis.

Une deuxième série concerne la fonction de répartition, c'est-à-dire, $A(u) = 1/F^{-1}(1-u)$:

(C1) $A(\cdot)$ est dérivable sur $(0,1)$ et sa dérivée est notée $A'(u)=a(u)$.
(C2) $a(\cdot)$ est continue sur un intervalle $[a',a'']$ avec $0<a'<a''<1$.
(C3) $\exists u_0>0, \exists \eta>-3/2, \forall u\in(0,u_0), |a(u)|<C_0u^\eta\exp(\int_u^1 b(t)t^{-1}\,dt)$, où $0<b(t)\to 0$ quand $t\to 0$.

La condition (C3) veut dire que a est majorée par une fonction à variation régulière

$$S(u)=C_0u^\eta\exp\Big(\int_u^1 b(t)t^{-1}\,dt\Big)$$

d'exposant $\eta>-3/2$.

Enfin, les conditions supplémentaires concernent le poids $c(n,q,j)$. Cependant, nous ne pouvons les décrire de manière précise à cette étape. Il nous suffit de dire que nous transformerons $c(n,q,j)$ sous forme d'une fonction aléatoire $L_n(\cdot)$, de sorte que $L_n(j/n)=c(n,q,j)$ et que $L_n(\cdot)$ converge vers une fonction continue L. Nous aurons alors besoin de l'hypothèse

$$\sqrt{n}\sup_{0\le s\le 1}|L_n(s)-L(s)|=o_p(1). \tag{2.1}$$

Ces hypothèses se révèlent très douces, et sont vérifiées par l'essentiel des mesures de pauvreté courantes et les fonctions de répartition usuelles. Nous allons exposer nos résultats ci-dessus. Nous ferons un bref commentaire et donnerons par la suite quelques indications sur les preuves largement disponibles dans [10].

Théorème 1. *Supposons vraies les hypothèses* (C1)–(C3), (D1)–(D2), *et* (2.1). *Alors* $\sqrt{n}(J_n-D_n)\to\mathcal{N}(0,\theta^2)$, *avec*

$$D_n=\int_0^{q/n} L(s)d\Big(\frac{Z-y_0-1/F^{-1}(1-s)}{Z}\Big)ds,$$

$$\theta^2=Z^{-2}\int_0^{G(Z)}\int_0^{G(Z)} L(u)L(v)h(u)h(v)(u\wedge v-uv)\,dudv,$$

où $h(s)=a(s)d'\big(\frac{Z-y_0-1/F^{-1}(1-s)}{Z}\big)$. *Soit*

$$D=\int_0^{G(Z)} L(s)d\Big(\frac{Z-y_0-1/F^{-1}(1-s)}{Z}\Big)ds.$$

Alors $\sqrt{n}(J_n-D)\to\mathcal{N}(0,\vartheta^2)$, *où*

$$\vartheta^2=\theta^2+k(G(Z))^2G(Z)(1-G(Z))+\frac{2k(G(Z))}{Z}(1-G(Z))\int_{1/n}^{G(Z)} sL(s)h(s)\,ds,$$

avec

$$k(s)=L(s)d\Big(\frac{Z-y_0-1/F^{-1}(1-s)}{Z}\Big).$$

Remarque 1. Ces résultats ont été simulés avec satisfaction pour des tailles moyennes, de l'ordre de 50. Dans les données disponibles du Sénégal, l'ajustage de F avec la loi lognormal est bon, tout comme avec la loi de Singh–Maddalla. Ces résultats se révèlent efficaces dans l'évaluation et le suivi comparatif, aux plans spatial et temporel, de la pauvreté. A titre d'exemple, une étude à large échelle a été faite avec les bases de données du Sénégal [9].

3. Proofs.

Nous allons simplement brosser la preuve de nos résultats qui sont relativement techniques et disponibles [10]. Mais il est important de voir comment le poids $c(n, q, j)$ est exprimé au moyen d'une fonction $L_n(\cdot)$. Pour cela, utilisons la représentation des variables aléatoires X_i, $i \geq 1$, par $F^{-1}(1-U_i)$, $i \geq 1$, où $U_1, U_2, \ldots$ sont des variables aléatoires indépendantes uniformément distribuées sur $(0,1)$. Soit maintenant $U_n(\cdot)$ et $V_n(\cdot)$ la fonction de répartition et la fonction des quantiles empiriques basées sur $U_i, 1 \leq i \leq n$. Nous avons

$$j \geq 1,\ \frac{j-1}{n} < s \leq \frac{j}{n} \quad \Longrightarrow \quad \frac{j}{n} = U_n(V_n(s)),$$

si bien que

$$j \geq 1, \frac{j-1}{n} < s \leq \frac{j}{n} \Longrightarrow c(n, q, j) = c(n, q, nU_n(V_n(s)) \equiv L_n(s).$$

Puisque $U_n(V_n(s)) \to s$, presque sûrement quand $n \to \infty$, nous pouvons espérer que $L_n(\cdot)$ converge uniformément vers une certaine fonction $\mathrm{L}(\cdot)$, positive et continûment différentiable sur $(0,1)$, c'est-à-dire que,

$$L_n(\cdot) \to L(\cdot),\ \text{p.s. quand } n \to \infty.$$

Nous exigeons dans ce papier, en plus de (2.1) que les fonctions $L_n(\cdot)$ soient uniformément bornées par une constante $D > 0$, c'est-à-dire

$$\sup_{n \geq 1} \sup_{0 \leq s \leq 1} |L_n(s)| \leq D,\ \text{p.s.}$$

Nous pouvons aisément vérifier que (2.1) a lieu pour beaucoup de mesures de pauvreté classiques, en particulier les mesures non pondérés puisque $L_n \equiv L = 1$ pour elles. Vérifions la, pour la mesure de Shorrocks. En effet, nous avons

$$L_n(s) = c(n, q, j) = (2 - 2j/n + 1/n) \to L(s) = 2(1-s),$$

et

$$\sqrt{n}(L_n(s) - L(s)) = -2\sqrt{n}(U_n(V_n(s)) - s) + 1/\sqrt{n}.$$

Nous obtenons, via [15, p. 151], que

$$\sqrt{n} \sup_{0 \leq s \leq 1} |L_n(s) - L(s)| \leq 3/\sqrt{n}.$$

En fait, cette condition est généralement satisfaite quand le poids ne dépend pas du nombre aléatoires de pauvres q_n de l'échantillon. Les mesures de Sen et de Kakwani n'entrent pas dans ce registre. Le poids $L_n(\cdot)$ introduira une modification de la loi limite normale. Nous nous y pencherons dans un autre article.

Il nous reste simplement à indiquer la preuve du Théorème 1. Pour cela, on utilise la représentation citée et les approximations dites *hongroises*. Précisément, Csőrgö et al. [2] ont construit un espace de probabilité portant à la fois une suite de variables aléatoires $U_1, U_2, \ldots$ indépendantes et uniformément distribuées sur $(0,1)$, et une suite de ponts Browniens $(B_n)_{n\geq 1}$ telles pour $0 < \nu < 1/2$, quand $n \to \infty$,

$$\sup_{1/n \leq s \leq 1-1/n} \frac{|\beta_n(s) - B_n(s)|}{(s(1-s))^{1/2-\nu}} = O_p(n^{-\nu}), \tag{3.1}$$

et pour $0 < \nu < 1/4$,

$$\sup 1/n \leq s \leq 1 - 1/n \frac{|B_n(s) - \alpha_n(s)|}{(s(1-s))^{1/2-\nu}} = O_p(n^{-\nu}), \tag{3.2}$$

où α_n et β_n sont respectivement le processus empirique et le processus des quantiles basés sur $U_1, \ldots, U_n$. (Voir aussi [3] pour une preuve plus directe et courte, et [12] pour une version duale, en ce sens que dans [2], (3.1) est établie pour $0 < \nu < 1/2$ et (3.2) pour $0 < \nu < 1/4$, tandis que dans l'approche de [12], (3.1) est établie pour $0 < \nu < 1/4$ et (3.2) pour $0 < \nu < 1/2$). Nous nous plaçons dans cet espace et utilisons la représentation $\{X_i, i \geq 1\} = \{F^{-1}(1 - U_i), i \geq 1\}$. A partir de là, par des procédures habituelles, on peut mettre la statistique J_n sous une forme intégrale faisant intervenir la fonction empirique des quantiles. Ensuite par des techniques faisant largement appel au théorème des accroissements finis, nous serons en mesure de développer $\sqrt{n}(J_n - D)$ sous la forme

$$\sqrt{n}(J_n - D) =: N_n + \sqrt{n}(D_n - D) + R_{0,0,n} + R_{0,1,n} + R_{1,n} + R_{2,n} + R_{3,n},$$

où

$$N_n = \frac{1}{Z} \int_{1/n}^{G(Z)} L(s) B_n(s) h(s)\, ds.$$

Le principe des preuves consiste à montrer que les termes R_* tendent vers zéro en probabilité sous les hypothèses du Théorème 1, que

$$N_n \quad \text{et} \quad N_n + \sqrt{n}(D_n - D)$$

tendent vers des lois normales centrées de variances respectives θ^2 et ϑ^2. Remarquons enfin que $k(G(Z))$ est nul dès que F est bijective et cela conduit à l'égalité $\theta^2 = \vartheta^2$.

Références

1. J. A. Bishop, J. P. Formby, and B. Zheng, *Statistical inference and the Sen index of poverty.* Int. Econ. Rev. **38** (1997), no. 2, 381–387.
2. M. Csörgő, S. Csörgő, L. Horvàth, and M. Mason, *Weighted empirical and quantile processes.* Ann. Probab. **14** (1986), no. 1, 31–85.
3. M. Csörgő and L. Horvàth, *Approximations of weighted empirical and quantile processes.* Statist. Probab. Lett. **4** (1986), no. 6, 275–280.
4. G. Dia, *Répartition ponctuelle aléatoire des revenus et estimation de l'indice de pauvreté.* Afr. Stat. **1** (2005), no. 1, 47–66 .
5. J. Foster, J. Greer, and A. Shorrocks, *A class of decomposable poverty measures.* Econometrica **52** (1984), 761–766.
6. J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*. Wiley, New York, 1985.
7. N. Kakwani, *On a class of poverty measures.* Econometrica **48** (1980), no. 2, 437–446.
8. ———, *Issues on Setting Absolute Poverty Line.* In: Poverty and Social Developement. Asian Bank of Developement (ADB), 2003.
9. G. S. Lo. *Estimation asymptotique des indices de pauvreté : modélisation continue et analyse spatio-temporelle de la pauvreté au Sénégal.* J. Afric. Commun. Sci. Techn. **4**, (2009), 341–377.
10. G. S. Lo, S. T. Sall, and C. T. Seck, *A general asymptotic theory of poverty measures.* Disponible à : `http://www.ufrsat.org/lerstad/pub/ssl_general_2007.pdf`.
11. G. S. Lo and S. T. Sall, *The asymptotic theory of the poverty intensity in view of extreme value theory for two simple cases.* Afr. Stat. **2** (2007), no. 1, 41–58.
12. D. M. Mason and W. R. van Zwet, *A refinement of the KMT inequality for the uniform empirical process.* Ann. Probab. **15** (1987), no. 3, 871–884.
13. A. K. Sen, *Poverty: an ordinal approach to measurement.* Econometrica **44** (1976), 219-231.
14. A. Shorrocks, *Revisiting the Sen poverty index.* Econometrica **63** (1995), 1225–1230.
15. G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics.* Wiley-Interscience, New York, 1986.
16. B. Zheng, *Aggregate poverty measures.* J. Econ. Surv. **11** (1997), no. 2, 123–162.

LSTA,
Université Pierre et Marie Curie,
France
et
LERSTAD,
Université Gaston Berger,
Sénégal.
email: ganesamblo@ufrsat.org

Université Bambey,
Sénégal.
email: ctseck@ufrsat.org.

FASTEF,
Université Cheikh Anta Diop,
Sénégal
email: stsall@ufrsat.org