C. R. Acad. Sci. Paris, Ser. I 353 (2015) 947–952

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Probability theory/Statistics

Non-parametric estimation of income distribution and poverty index in the unidimensional context with $\alpha \in]0, 1[$

Estimation non paramétrique de la distribution des revenus et de l'indice de pauvreté $\alpha \in]0, 1[$

Youssou Ciss, Galaye Dia [1], Aboubakary Diakhaby

LERSTAD, UFR de sciences appliquées et de technologie, B P 234, Université Gaston-Berger, Saint-Louis, Senegal

A R T I C L E I N F O

Article history:
Received 9 June 2014
Accepted after revision 31 August 2015
Available online 3 October 2015

Presented by Paul Deheuvels

A B S T R A C T

In this paper, we propose an estimator of Foster, Greer and Thorbecke class of measures $P(z,\alpha) = \int_0^z \left(\frac{z-x}{z}\right)^\alpha f(x)\,dx$, where $z > 0$ is the poverty line, $f(x)$ is the density distribution of the income random variable and $\alpha \in]0, 1[$ is the so-called poverty aversion. $\alpha \in]0, 1[$ remained an open problem in the work of Dia [1], where he was considering the case where $\alpha = 0$ and $\alpha \geq 1$. The estimator is constructed with the Parzen–Rosenblatt Kernel. Almost sure uniform convergence and uniform convergence in mean square error are established. Finally, the new estimator has been applied to the study of the poverty in Senegal. The study of this application indicates that our new estimator performs well.

R É S U M É

Dans cette note, nous proposons un estimateur de l'indice de pauvreté de Foster, Greer et Thorbecke, défini par : $P(z,\alpha) = \int_0^z \left(\frac{z-x}{z}\right)^\alpha f(x)\,dx$, où $z > 0$ est le seuil de pauvreté, $f(x)$ la densité de la distribution des revenus et $\alpha \in]0, 1[$ est le paramètre d'aversion de la pauvreté. Le cas $\alpha \in]0, 1[$ restait un cas ouvert dans le travail de Dia [1], où il a considéré les cas $\alpha = 0$ et $\alpha \geq 1$. L'estimateur est construit à l'aide du noyau de Parzen–Rosenblatt. La convergence uniforme presque sûre et la convergence uniforme en erreur quadratique moyenne sont établies. Enfin, notre estimateur a été appliqué à l'étude de la pauvreté au Sénégal. L'étude de cette application montre que notre estimateur est recommandé.

E-mail addresses: ciss.youssouf@gmail.com (Y. Ciss), diakhaby81@yahoo.fr, diakhaby@ugb.edu.sn (A. Diakhaby).

[1] In memory, died in May 2013.

http://dx.doi.org/10.1016/j.crma.2015.08.007
1631-073X/

Version française abrégée

Soit $F(x)$ la fonction de répartition des revenus d'une population admettant une densité $f(x)$ continue en x. La famille des indicateurs FGT (Foster, Greer, Thorbecke) [2], indexée par un réel $\alpha \geq 0$, est définie par $P(z,\alpha) = \int_0^z \left(\frac{z-x}{z}\right)^\alpha f(x)\,dx$, $z > 0$. Si $z < 0$, nous posons $P(z,\alpha) = 0$, où z est le seuil de pauvreté ou la ligne de pauvreté. L'estimateur suivant de $P(z,\alpha)$ (voir, par exemple, [9]), $\widehat{P}_n(z,\alpha) = \frac{1}{n}\sum_{i=1}^n \left(1 - \frac{X_i}{z}\right)_+^\alpha$, où $x_+ = \max(0,x)$ est l'estimateur empirique de l'indice de pauvreté FGT. Nous nous proposons de construire un estimateur de cet indice par la méthode du noyau basée sur la somme de Riemann, défini comme suit :

$$P_n(z,\alpha) = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=0}^{[z/h]} (\frac{z-ih}{z})^\alpha K\Big(\frac{X_j - ih}{h}\Big)\,dx \qquad \alpha \in]0,1[. \tag{1}$$

Nos résultats sont relatifs aux deux hypothèses suivantes sur la densité f, supposée bornée. Soit elle est uniformément continue, soit elle admet une dérivée presque partout $f' \in L_1(\mathbb{R})$. Pour chacun des cas, la convergence de l'estimateur (1) est établie. Nous aurons besoin d'hypothèses additionnelles sur K. Notons la variation totale de K sur $]-\infty; u]$, pour un réel u, par $V_{-\infty}^{u}K$, et par $V(\mathbb{R})$ sa variation totale sur $\mathbb{R}$.

Soient les hypothèses supplémentaires suivantes.

($\mathbf{A_1}$) K est à variation bornée, i.e., $V(\mathbb{R}) < \infty$.
($\mathbf{A_2}$) $\int_{\mathbb{R}} |uK(u)| < +\infty$.
($\mathbf{A_3}$) Il existe une fonction décroissante λ telle que $\lambda(\frac{u}{h}) = O(h)$ sur tout intervalle borné et telle que

$$\forall \quad (x,y) \in \mathbb{R}^2,\ |K(x) - K(y)| \leq \lambda|x-y| \quad \text{et} \quad \lambda(u) \longrightarrow 0,\ u \to 0,\ u \geq 0.$$

Nous pouvons maintenant énoncer nos principaux résultats.

Théorème 0.1. *Supposons que l'hypothèse* $\mathbf{A_1}$ *soit vérifiée et* f *uniformément continue. Alors, pour tout* $b > 0$*, l'estimateur* $P_n(z,\alpha)$ *converge uniformément presque sûrement sur* $[0,b]$ *vers* $P(z,\alpha)$ *lorsque* $n \to +\infty$*, i.e.*

$$P(\lim_{n\to+\infty} \sup_{z\in[0,b]} |P_n(z,\alpha) - P(z,\alpha)| = 0) = 1$$

pourvu que $nh^2(\log\log n)^{-1} \to +\infty$ *lorsque* $n \to +\infty$.

Théorème 0.2. *Supposons que les hypothèses* $\mathbf{A_1}$ *et* $\mathbf{A_2}$ *soient vérifiées et que* f *admette une dérivée presque partout* $f' \in L_1(\mathbb{R})$*. Alors, pour tout* $b > 0$*, l'estimateur* $P_n(z,\alpha)$ *converge uniformément presque sûrement sur* $[0,b]$ *vers* $P(z,\alpha)$ *lorsque* $n \to +\infty$*, i.e.*

$$P(\lim_{n\to+\infty} \sup_{z\in[0,b]} |P_n(z,\alpha) - P(z,\alpha)| = 0) = 1$$

pourvu que $nh^2(\log\log n)^{-1} \to +\infty$ *lorsque* $n \to +\infty$.

Théorème 0.3. *Si* f *est uniformément continue ou si l'hypothèse* $\mathbf{A_2}$ *est vérifiée et que la fonction* f *admette une dérivée presque partout* $f' \in L_1(\mathbb{R})$*, alors on a, sous l'hypothèse* $\mathbf{A_3}$ *pour tout* $b > 0$*,*

$$\lim_{n\to+\infty} \sup_{z\in[0,b]} \mathbb{E}\Big(P_n(z,\alpha) - P(z,\alpha)\Big)^2 = 0.$$

1. Introduction and definition of the estimator

Let $F(x)$ be the cumulative distribution function of the income variable X from a population with continuous density $f(x)$ at a given point x. The FGT (Foster, Greer, Thorbecke) [2] class of poverty measures indexed by the real $\alpha \geq 0$ is defined by:

$$P(z,\alpha) = \begin{cases} \int_0^z \left(\frac{z-x}{z}\right)^\alpha f(x)\,dx & \text{if } z > 0, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where z is the poverty line.

Let $X_1, \cdots, X_n$ be a random sample of size n from the income r.v. with d.f. F. The following estimator of $P(z,\alpha)$ (see [9]),

$$\widehat{P}_n(z,\alpha)=\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{X_i}{z}\right)_+^{\alpha},\qquad \text{where}\qquad x_+=\max(0,x),$$

is the empirical estimator of the FGT poverty index, which is fully useful in a large range of applications in economics (widely used in practice in econometrics and actuarial). This empirical estimator is unbiased consistent and has a limit normal-law mean $P(z,\alpha)$ and a variance equal to $n^{-1}(P(z,2\alpha)-(P(z,\alpha))^2)$. Lo et al. [5] used empirical processes and extreme-value theories to study this estimator. Seck [7], Seck and Lo [8] used some non-weighted poverty measures, viewed as stochastic processes and indexed by real numbers or monotone functions, to follow up the poverty evolution between two periods. Dia [1] propose a new kernel estimator (6), based on the Riemann sum, for $\alpha=0$ and $\alpha\geq 1$. The case $\alpha\in]0,1[$ remained an open problem.

Let us consider the classical estimator of the density $f(x)$ (Parzen–Rosenblatt):

$$\hat{f}(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{X_i-x}{h}\right)$$

where h is a function of n that tends to zero as n tends to infinity and K fulfills the following hypotheses:

$$(\mathbf{H_1})\ \sup_{-\infty<x<+\infty}|K(x)|<+\infty,\quad (\mathbf{H_2})\int_{-\infty}^{+\infty}K(x)\,\mathrm{d}x=1,\quad (\mathbf{H_3})\lim_{x\to\pm\infty}|xK(x)|=0. \tag{3}$$

Further, in this paper, we assume that the hypotheses $\mathbf{H_1},\mathbf{H_2},\mathbf{H_3}$ hold, that K is Riemann integrable, and that f is bounded with a support included in $\mathbb{R}_+$. We denote the lower bound of this support by x_0. Let us substitute f by $\hat{f}$ in (2). We obtain:

$$\tilde{J}_n(z,\alpha)=\int_0^z\left(\frac{z-x}{z}\right)^{\alpha}(nh)^{-1}\sum_{i=1}^{n}K\left(\frac{X_i-x}{h}\right)\mathrm{d}x. \tag{4}$$

Let $\Delta_{h,i}=[hi,h(i+1)[,\ 0\leq i<[\frac{z}{h}]$ be a partition of $[0,z]$ and using the Riemann sum definition of the integral, the last one corresponds to the limit of the following sum:

$$\tilde{P}_n(z,\alpha)=\frac{1}{n}\sum_{j=1}^{n}\sum_{i=0}^{[z/h]}\left(\frac{z-ih}{z}\right)^{\alpha}K\left(\frac{X_j-ih}{h}\right)\mathrm{d}x+\mathcal{V}_n(z) \tag{5}$$

where $[\frac{\cdot}{h}]$ design the integer part of $\frac{\cdot}{h}$ and

$$\mathcal{V}_n(z)=\frac{1}{n}\frac{\left(z-h[\frac{z}{h}]\right)-h}{h}\left(1-\frac{h[\frac{z}{h}]}{z}\right)^{\alpha}\sum_{j=1}^{n}K\left(\frac{X_j-[\frac{z}{h}]h}{h}\right).$$

Dia [1] showed that the term $\mathcal{V}_n(z)\to 0$ almost surely (a.s.) as $n\to+\infty$, for $\alpha=0$ and $\alpha\geq 1$.

By the properties of α-Hölderian functions, we establish that, for $\alpha\in]0,1[$, $\mathcal{V}_n(z)\to 0$ a.s. as $n\to+\infty$. Then we propose the following estimator of the FGT poverty index for $\alpha\in]0,1[$:

$$P_n(z,\alpha)=\frac{1}{n}\sum_{j=1}^{n}\sum_{i=0}^{[z/h]}\left(\frac{z-ih}{z}\right)^{\alpha}K\left(\frac{X_j-ih}{h}\right)\mathrm{d}x. \tag{6}$$

Our contribution is to prove that this kernel estimator introduced par Dia can be used for each value of index $\alpha\in[0,\infty[$, which is important since the FGT measures concern tests for poverty ordering or, equivalently, stochastic dominance and also optimal policy design (or program) for reducing poverty where the case $\alpha<1$ is specific and useful, as pointed out in [3].

Our best achievement is the complete description of the asymptotic behaviour of the kernel FGT index estimator for unexplored case of small value of α in $]0,1[$.

Additional hypotheses are made about the kernel K, that is:

($\mathbf{H_4}$) K is of bounded variation function on $\mathbb{R}$ and let $V(\mathbb{R})$ be its total variation.

($\mathbf{H_5}$) $\int_{\mathbb{R}}|uK(u)|\,\mathrm{d}u<+\infty.$

($\mathbf{H_6}$) There exists a non-increasing function λ such that $\lambda(\frac{u}{h})=O(h)$ on bounded intervals,

$$\forall(x,y)\in\mathbb{R}^2,\quad |K(x)-K(y)|\leq\lambda|x-y|\quad\text{and}\quad \lambda(u)\longrightarrow 0 \text{ when } u\to 0, \text{ and } u\geq 0.$$

Now, we can formulate our main results.

2. Convergence of the estimator

Our main results are relative to the following additional about the function $f(x)$:
$\mathbf{C_1}$: $f(x)$ is uniformly continuous.
$\mathbf{C_2}$: $f(x)$ admits almost everywhere a derivative $f'(x) \in L_1(\mathbb{R})$.

2.1. The almost-sure uniform convergence and the behaviour of the bias

Theorem 2.1. *Assume that the hypotheses* $\mathbf{H_4}$ *and* $\mathbf{C_1}$ *hold. Then, for all* $b > 0$*, estimator* $P_n(z, \alpha)$ *converges uniformly almost surely on* $[0, b]$ *to* $P(z, \alpha)$ *as* $n \to +\infty$*, i.e.*

$$P\Big(\lim_{n\to+\infty} \sup_{z\in[0,b]} |P_n(z,\alpha) - P(z,\alpha)| = 0\Big) = 1$$

provided that $n h^2 (\log\log n)^{-1} \to +\infty$ *as* $n \to +\infty$.

Theorem 2.2. *Assume that the hypotheses* $\mathbf{H_4}$, $\mathbf{H_5}$ *and* $\mathbf{C_2}$ *hold. Then, for all* $b > 0$*, the estimator* $P_n(z, \alpha)$ *converges uniformly almost surely on* $[0, b]$ *to* $P(z, \alpha)$ *as* $n \to +\infty$*, i.e.*

$$P\Big(\lim_{n\to+\infty} \sup_{z\in[0,b]} |P_n(z,\alpha) - P(z,\alpha)| = 0\Big) = 1$$

provided that $n h^2 (\log\log n)^{-1} \to +\infty$ *as* $n \to +\infty$.

For the demonstration of the theorems, we use the Theorem 2 of Kiefer [4] and the following lemmas showing that $P_n(z, \alpha)$ is uniformly asymptotic unbiased on all bounded intervals.

Lemma 2.3. *If* $\mathbf{C_1}$ *holds, then* $\forall\, b > 0$*, we have:*

$$\lim_{n\to+\infty} \sup_{z\in[0,b]} |\mathbb{E}(P_n(z,\alpha)) - P(z,\alpha)| \to 0, \quad n \to +\infty.$$

Lemma 2.4. *Let* $M = \sup_{z\in\mathbb{R}} F(z)$ *and* $A = \sup_{x\in\mathbb{R}} f(x)$*. If* $\mathbf{H_5}$ *and* $\mathbf{C_2}$ *hold, then:*

$$\sup_{z\in\mathbb{R}} |\mathbb{E}(P_n(z,\alpha)) - P(z,\alpha)| \le h \left(\Big(\int_{\mathbb{R}} |f'(x)|\,\mathrm{d}x \Big) \Big(\int_{\mathbb{R}} (|u|+1)|K(u)|\,\mathrm{d}u \Big) + 2(\alpha M + A h) \int_{-\infty}^{+\infty} |K(u)|\,\mathrm{d}u \right).$$

Remark 1. If K satisfies the hypothesis $\mathbf{H_5}$, then by using $\mathbf{H_1}$, the kernel

$$\widehat{K} = \frac{K^2}{\int_{\mathbb{R}} K^2(y)\,\mathrm{d}y} \text{ also satisfies it.}$$

From the two previous lemmas, we get the following corollaries:

Corollary 2.5. *Under the assumptions of Lemma 2.3, we have uniformly on* $[0, b]$ *(resp.* $\mathbb{R}$*)*

$$\lim_{n\to+\infty} \mathbb{E}\left(\sum_{i=1}^{[\frac{z}{h}]} \Big(1 - \frac{ih}{z}\Big)^{2\alpha} K^2\Big(\frac{X_j - ih}{h}\Big) \right) = P(z, 2\alpha) \int_{\mathbb{R}} K(y)\,\mathrm{d}y.$$

Corollary 2.6. *If the assumptions of Theorem 2.2 hold and if* $h = O(n^{-1}\log\log n)^{1/4}$*, then for all* $b > 0$*, we have almost surely:*

$$\sup_{z\in[0,b]} |\mathbb{E}(P_n(z,\alpha)) - P(z,\alpha)| = O(n^{-1}\log\log n)^{1/4}.$$

2.2. The uniform convergence in mean square error

Theorem 2.7. *If* $\mathbf{H_6}$ *and* $\mathbf{C_1}$ *hold. Then:*

1. $\lim_{n\to+\infty} n\mathbb{V}ar(P_n(z,\alpha)) = \Big(\int_{\mathbb{R}} K^2(y)\,\mathrm{d}y\Big) P(z, 2\alpha) - \Big(P(z,\alpha)\Big)^2$.
2. $\forall b > 0$, $\lim_{n\to+\infty} \sup_{z\in[0,b]} \mathbb{E}\Big(P_n(z,\alpha) - P(z,\alpha)\Big)^2 = 0$.

Table 1
Distribution of poverty by regions in the univariate case with P_n.

Région	$P_n(z_0, 0.05)$	$P_n(z_0, 0.5)$	$P_n(z_0, 0.1)$	$P_n(z_0, 0.9)$
KOLDA	0.8097132	0.7882261	0.7650018	0.5307519
TAMBA	0.7894673	0.7673535	0.7441354	0.4964766
DIOURBEL	0.6846584	0.6630282	0.6405833	0.4124965
FATICK	0.6770589	0.6584672	0.639135	0.4370827
THIES	0.6725587	0.6814292	0.6296	0.4167967
KAOLACK	0.6542437	0.6338166	0.6124937	0.3901114
SAINT-LOUIS	0.6531825	0.631532	0.6091871	0.3888169
LOUGA	0.6447103	0.6214581	0.5974451	0.3616545
ZIGUINCHOR	0.5977907	0.578627	0.5587509	0.3525811
DAKAR	0.344285	0.3288079	0.313062	0.174239

Theorem 2.8. *Assume that* $\mathbf{H_6}$ *and* $\mathbf{C_2}$ *hold. Then:*

1. $\lim_{n\to+\infty} n\mathbb{V}ar(P_n(z,\alpha)) = \Big(\int_{\mathbb{R}^2} K^2(y)\,\mathrm{d}y\Big)P(z,2\alpha) - \Big(P(z,\alpha)\Big)^2$.
2. *Moreover, if* $\mathbf{H_5}$ *holds, we have, for all* $b > 0$,

$$\lim_{n\to+\infty} \sup_{z\in[0,b]} \mathbb{E}\Big(P_n(z,\alpha) - P(z,\alpha)\Big)^2 = 0.$$

Theorem 2.9. *Assume that* $\mathbf{H_6}$ *holds. Then for all* $b > 0$,

$$\lim_{n\to+\infty} \sup_{z\in[0,b]} \sum_{0\le i\ne j\le[\frac{z}{h}]} \Big(1-\frac{ih}{z}\Big)^{\alpha}\Big(1-\frac{jh}{z}\Big)^{\alpha} \int_{\mathbb{R}} K\Big(\frac{u-ih}{h}\Big)K\Big(\frac{u-jh}{h}\Big) f(u)\,\mathrm{d}u = 0.$$

And this theorem requires the following lemma.

Lemma 2.10. *Let* $0 \le \theta_i \le 1$, $i = 1, 2$. *Then for all* x, y *and* $x \ne y$, *we have:*

$$\lim_{n\to+\infty} \sup_{(\theta_1,\theta_2)\in[0,1]^2} \Big(h^{-2}\int_{-\infty}^{+\infty} |K\Big(\frac{u-x+\theta_1}{h}\Big)K\Big(\frac{u-y+\theta_2}{h}\Big)| f(u)\,\mathrm{d}u\Big) = 0.$$

Remark 2. The estimator $P_n(z,\alpha)$ has asymptotic efficiency with respect to $\widehat{P}_n(z,\alpha)$,

$$e(z,\alpha) = \Big((P(z,2\alpha)\int_{\mathbf{R}} K^2(y)\,\mathrm{d}y) - (P(z,\alpha))^2\Big)/\Big(P(z,2\alpha) - (P(z,\alpha))^2\Big).$$

The integral $\int_{\mathbf{R}} K^2(y)\,\mathrm{d}y$ is strictly inferior to 1 for the conventional kernels [6] (p. 1068). Then we have, in this case, $e(z,\alpha) < 1$. In Theorem 2.8, the rate of convergence in mean square is estimated to $O(\frac{1}{n})$ if h is estimated at $O(\frac{1}{\sqrt{n}})$.

3. Applications

The proposed estimator has been applied to the study of poverty in Senegal. The data used is focused on ESAM/1996 (Senegal household Survey) provided by the National Statistics Agency. Incomes vary from 100 CFA to Millions CFA in the region of DAKAR. The poverty line was $z_0 = 143080$ CFA. Table 1 gives the values of $P_n(z_0,\alpha)$ for $(\alpha = 0, 05)$ $(\alpha = 0.5)$ $(\alpha = 0.1)$ and $(\alpha = 0.9)$ for the estimators. According to this table, we can classify the ten regions into six classes: **CL1:** KOLDA; **CL2:** TAMBA; **CL3:** DIOURBEL, FATICK, THIES; **CL4:** KAOLACK, SAINT-LOUIS; **CL5:** LOUGA, ZIGUINCHOR; **CL6:** DAKAR. This means that the various parameters of poverty decrease from **CL1** to **CL6**. The poverty rate also decreases along this descending order of classes. So, we can conclude that our estimator is recommended for this study of poverty in Senegal. Note that, Seck and Lo [8] use some nonweighted poverty measures to follow up the poverty evolution between two periods in Senegal.

References

[1] G. Dia, Estimation nonparamétrique de la distribution des revenus et de l'indice de pauvreté, C. R. Acad. Sci. Paris, Ser. I 346 (2008) 907–912.
[2] J.E. Foster, J. Greer, E. Thorbecke, A class of decomposable poverty measures, Econometrica 52 (1984) 761–776.
[3] J.E. Foster, J. Greer, E. Thorbecke, The Foster–Greer–Thorbecke (FGT) poverty measures: twenty five years later, J. Econ. Inequal. 8 (4) (2010) 491–524.

[4] J. Kiefer, On large deviations of the empirical distribution of vector chance variable and law of iterated logarithm, Pac. J. Math. 11 (1961) 143–154.
[5] G.S. Lo, S.T. Sall, C.T. Seck, Une théorie asymptotique générale des mesures de pauvreté, C. R. Math. Rep. Acad. Sci. Can. 31 (2) (2009) 45–52.
[6] E. Parzen, On the estimation of a probability density function and mode, Ann. Math. Stat. 33 (3) (1962) 1065–1076.
[7] C.T. Seck, Estimation non paramétrique et convergence faible des mesures de pauvreté, PhD dissertation, LSTA, Université Pierre-et-Marie-Curie, Paris, 2010.
[8] C.T. Seck, G.S. Lo, Convergence of the non-weighted poverty measures, Commun. Stat., Theory Methods 38 (2009) 3697–3704.
[9] C. Seidl, Poverty measurement: a survey, in: D. Bos, C. Seidl (Eds.), Welfare and Efficiency in Public Economics, Springer-Verlag, Heidelberg, Germany, 1988, pp. 71–147.