Répartition Ponctuelle Aléatoire des Revenus et Estimation de l'Indice de Pauvreté

Galaye Dia

9 avril 2005

Résumé

Nous nous proposons d'étendre dans le cadre des processus ponctuels l' étude des estimateurs classiques des indices de pauvreté. Dans l'étude qui va suivre, nous proposons un estimateur de l'indice de Forster-Greer-Thorbecke.

AMS : primary 60F15 ;secondary 60G25
Keywords : Processus ponctuels, superposition, seuil de pauvreté, processus aminci, panels d'observations.

[1]

1 Introduction

Il est globalement accepté que la pauvreté est aujourd'hui un phénomène mondial. Ses conséquences vont au delà de ce qui est considéré comme son aspect le plus manifeste c'est-à- dire simplement : la souffrance humaine. Elle est aujourd'hui associée à l'émigration, aux conflits militaires et plus récemment au terrorisme. Cependant ces conséquences varient selon les régions. Pour les pays riches c'est plutôt l'exclusion sociale qui prévaut, tandis que dans les pays en voie de développement c'est la malnutrition, le manque de soins primaires conduisant à la mort. Combattre la pauvreté n'est plus seulement une question de responsabilité morale mais elle est devenue une question de sécurité pour toutes les nations.

Dès lors il urge pour les gouvernements d'élaborer une politique hardie pour lutter contre ce fléau. Il est évidemment clair qu'il faudra choisir celle qui réduise au mieux ce phénomène. Avant un tel choix il faudra répondre à la question **qui est pauvre ?** et comment mesurer la pauvreté ? La pauvreté revêt plusieurs aspects : sous-alimentation, manque d'opportunités, manque de ressources, manque de liberté, manque de dignité, etc... Ainsi la pauvreté est multidimensionnelle. Cet aspect proposé par la Banque Mondiale est appelé "the Living Standards Measurement Survey "(LSMS). On peut s'intéresser à un aspect et utiliser une statistique pour étudier cet aspect ; on aura alors autant de statistiques que de composantes constituant la pauvreté. Les interventions des décideurs politiques ou des donateurs seraient dans ce cas ciblées vers ces aspects.

Une autre possibilité est d'utiliser le concept de **pauvreté absolue** c'est-à- dire le minimum nécessaire pour un être humain de vivre. Cette **ligne de pauvreté** où encore **seuil de pauvreté** est souvent fixé par les experts de la Banque Mondiale. Il est de 2.15 U.S dollars

[1]¶mail to **UFR de Sciences Appliquées et Technologie**
B.P 234 Université Gaston Berger de Saint-Louis (Sénégal) E-mail : galaye@ugb.sn
Centre de Recherches Economiques Appliquées (CREA) Univ. Ch. Anta Diop **MIMAP-II** & Laboratoire d'Etudes et de Recherches en Statistiques et Développement (LERSTAD) Univ. Gaston Berger.

de pouvoir d'achat par jour et par tête pour les pays pauvres et de 4.30 U.S dollars pour les pays de l'Europe centrale. Cependant les gouvernements peuvent déterminer eux-mêmes leur propre seuil de pauvreté. Cette détermination fait intervenir soit le revenu soit la consommation. L'approche peut être exprimée en coût d'un panier de besoins indispensables soit Z_1, en consommation d'aliments énergétiques (équilibre énergétique) ou à l'aide du coefficient d'Engel soit Z_2. On a

$$Z_1 = \sum_{i=1}^{n} p_i q_i$$

où le vecteur $(q_1, \ldots, q_n)$ représente n quantités de minimum indispensable et le vecteur $(p_1, \ldots, p_n)$ les prix correspondants. Z_2 est fourni par la méthode d'Orshanski (1965)

$$Z_2 = \frac{C_{ab}}{e_{ri}}$$

où C_{ab} est le coût des produits alimentaires de base consommés et e_{ri} la proportion du revenu total qui y a été investi. Une faible proportion du revenu utilisé pour la consommation est associée à un revenu élevé et à l'opposé à une grande proportion du revenu utlisé pour la consommation est associée un petit revenu.

L'autre concept est celui de **pauvreté relative** qui compare la position d'un individu ou d'un ménage par rapport à celle de l'ensemble de la population étudiée. Il est fixé égal à une proportion donnée de la moyenne arithmétique de la population ou de la médiane de la distribution.

Le premier concept est plus adéquat pour les pays du tiers monde où la pauvreté est endémique et implique des pénuries alimentaires et de logements tandis que le second concept est plus adapté au contexte des pays développés où on est considéré comme pauvre parce que son niveau de vie est inférieur au niveau de vie moyen du reste de la population.

A côté de ces deux concepts on parle de **pauvreté objective** et **de pauvreté subjective**. La pauvreté objective s'intéresse aux faits objectifs tels que les revenus mesurés en devises, l'alimentation exprimée en calories, l'habitat en surface occupée etc... La pauvreté subjective parle elle de degrés de satisfaction, de condition de logement. Ces deux attitudes méthodologiques peuvent être prises en compte dans la détermination du seuil de pauvreté absolue ou relative.

Une fois que le seuil de pauvreté est fixé il s'agira **d'agréger** la pauvreté c'est-à-dire de mesurer l'ampleur de la pauvreté au niveau de l'ensemble de la population considérée. A cette fin les économistes ont créé des indices utilisés par les décideurs politiques et les organismes internationaux

2 Les indices de pauvreté

Le premier indice de pauvreté qui a été longtemps utilisé appelé le *headcount ratio* ou *taux de pauvreté* est défini par

$$H = \frac{q}{n}$$

où q désigne le nombre de pauvres dans la la population et n le nombre d'individus de la population. Un tel indice ne fait pas de différence entre le pauvre juste en dessous de la ligne de pauvreté et le pauvre à l'extrême inférieure, les pauvres étant classés du plus pauvre au moins pauvre. Une politique basée sur cet indice ne limiterait pas ses efforts aux plus démunis mais consacrerait ses efforts à faire franchir le seuil de pauvreté. D'où une politique de l'aide aux moins pauvres.

Le second indice est appelé le *income gap ratio* ou **déficit moyen de revenu** des pauvres. Si y_i est le revenu du i$^{\text{ème}}$ pauvre et z le seuil de pauvreté on note $g_i = z - y_i$, il est défini par

$$I = \frac{1}{qz}\sum_{i=1}^{q} g_i$$

On remarque que I augmente si le revenu d'un pauvre quelconque diminue ce qui représente déjà un avantage par rapport à H. Mais on peut avoir le même indice I deux fois plus de pauvres dans le pays considéré ce qui n'est pas souhaitable pour une politique de diminution de la pauvreté

Ces inconvénients constatés sur ces deux indices qui ont été les tous premiers indices utilisés dans la lutte contre la pauvreté ont conduit les acteurs de la lutte contre la pauvreté á une définition rigoureuse de l'indice de pauvreté.

Exemples d'indices de pauvreté

Les indices présentés sont de la forme :

$$P_n(x,z) = A(x,z)\sum_{i=1}^{q(x)} g_i \mu_i(x,z)$$

où $A(x,z)$ est une fonction de x et z et $\mu_i(x,z)$ est une pondération sur les écarts au seuil de pauvreté.

1. **L'indice de Sen :**

$$S_n(x,z) = \frac{2}{(q+1)nz}\sum_{i=1}^{q}(z-x_i)(q+1-i)$$

2. **L'indice de Shorrocks :**

$$Sh_n(x,z) = \frac{1}{n^2 z}\sum_{i=1}^{q}(z-x_i)(2n-2i+1)$$

et enfin

3. **l'indice de Foster, Greer, Thorbecke**

$$FGT(x,z,\alpha) = \frac{1}{n}\sum_{i=1}^{q}(\frac{z-x_i}{z})^{\alpha}$$

où α est un paramètre reél positif.

- Si $\alpha = 0$ on obtient le taux de pauvreté. C'est la proportion de la population dont le revenu se situe en dessous de la ligne de pauvreté
- Si $\alpha = 1$, l'indice représente **la profondeur de la pauvreté**. Il indique le montant de revenu néceeaire pour sortir la population de la pauvreté. Le montant est rapporté au nombre d'individu de la population.

– Pour $\alpha = 2$ on a la **sévérité de la pauvreté**. Il est encore appelé **indice d'inégalité parmi les pauvres**

Dans la littérature scientifique consacrée à la recherche statistique sur la pauvreté,la distribution des revenus considérée est soit de type déterministe (G.S.LO 2003), soit stochastique (W.Ogryczak et A. Ruszczynski 2001, R. Davidson et J.I Duclos 2000). Et dans ce dernier cas la distribution des revenus est caractérisée par sa fonction de distribution F commune à tous les individus de la population étudiée.Ou encore formulée autrement, la distribution des revenus de la population considérée est une variable aléatoire de fonction de répartition F.

Il semble cependant que la nécessité d'avoir à des moments donnés l'état des revenus de la population pour une appréhension juste de la pauvreté commande de considérer la répartition des revenus dans son ensemble comme un processus ponctuel évolutif ou répartition pontuelle aléatoire de loi donnée non nécessairement connue.De ce point de vue l'étude des **panels d'observations** entre tout naturellement dans ce cadre. Nous l'envisagerons ultérieurement dans le contexte des processus fortement mélangeants.

3 Le schéma classique

Soit S_n une population de taille n. Un ensemble de revenus pour cette population est un élément x de l'espace $\mathbf{R}_+^n$ noté $x = (x_1, \ldots, x_n)$, $x_i \geq 0$, les individus étant numérotés de 1 à n.

On appelle seuil de pauvreté un réel $z \in \mathbf{R}_+$, ensemble des nombres réels strictement positifs, tel que l'individu i est considéré comme pauvre si son revenu x_i est strictement inférieur à z. On désignera par $Q(z) = \{1 \leq i \leq n : x_i < z\}$ l'ensemble des individus pauvres de la population et par q son cardinal.

La famille d'indicateurs F.G.T (Foster, Greer ,Thorbecke) indexés par $\alpha > 0$ suivants

$$P_{\alpha,H} = \frac{1}{H} \sum_{i \in Q(x)} (\frac{z - x_i}{z})^\alpha$$

a été étudiée par G.S LO (2003) pour $H = n$ et $H = q$; Il a montré en utilisant les outils d'échantillonnage que les estimateurs étaient respectivement sans biais et asymptotiquement sans biais.

Lorsque la distribution des revenus est caractérisée par une fonction de répartition F l'indicateur de F.G.T s'écrit

$$P_{z,\alpha} = \int_0^z (\frac{z - x}{z})^\alpha \, dF(x)$$

Il est estimé à l'aide d'un échantillon de taille N par

$$\widehat{P}_{(\alpha,N)} = \int_0^z (\frac{z - x}{z})^\alpha \, d\widehat{F}(x)$$

où $\widehat{F}$ est la fonction de répartition empirique des revenus

Pour deux différentes distributions de revenus Duclos et Davidson [?] étudient la dominance stochastique en utilisant P_α pour différentes valeurs entières de α. Le graphe de la courbe $P_{z,\alpha}$ est dès lors d'une utilité certaine.

4 Le schéma des processus ponctuels aléatoires

Examinons d'abord le schéma présenté par Chakravarty *et al* [1].

Soit une population de taille n. Une distribution de revenus est un vecteur de $\mathbf{R}_+^n$ l'octant positif de $\mathbf{R}^n$ soit $(x_1, \ldots, x_n)$ et l'ensemble de toutes les distributions de revenus est $\mathbf{R}_+ = \bigcup_{n \in \mathbb{N}} \mathbf{R}_+^n$. Pour tout $n \in \mathbb{N}$, ils considèrent l'ensemble des revenus S(x)=$\{x_i \quad i = 1, \ldots, n : x_i < z\}$ et désignent leur nombre par $q(z)$. Comme toute application ϕ de $\mathbf{R}_+$ dans $\mathbb{N}$ telle que l'ensemble Supp $=\{x : x \in \mathbf{R}_+, \phi(x) > 0\}$ soit fini ou dénombrable est appelée **répartition ponctuelle** par Geffroy *et al* [5] et pour toute partie A de $\mathbf{R}_+$, $\phi(A) = \sum_{x \in A} \phi(x)$ est l'effectif de la répartition ϕ sur A, il en résulte que $S(x) = \phi(]0, z])$. Si l'effectif de la population est une variable aléatoire q définie sur un espace $(\Omega, \mathcal{T})$, prenant donc ses valeurs dans $\mathbb{N}$, parler du revenu d'un i$^{\text{ème}}$ individu n'a de sens que si $q \geq i$; donc la variable associée à ce revenu n'est pas en général une variable aléatoire au sens classique car n'étant pas définie sur Ω tout entier. J.Geffoy [4] appelle une telle variable **variable aléatoire restreinte** ou **variable statistique**.

On considère une population S finie où non. Les revenus des individus sont numérotés par l'ensemble $\mathbb{N}$ des entiers naturels suivant l'ordre croissant. L'individu i est supposé avoir un revenu X_i qui est aléatoire. A la suite $X_i, i \in \mathbb{N}$, on peut associer de façon unique un processus ponctuel B.Lanuzel [**?**] que nous désignerons par q_0, engendré par les statistiques d'ordre $X_{(i)}$. La variable $X_{(i)}$ est égale au salaire du i$^{\text{ème}}$ individu dans le classement que nous venons d'évoquer. La taille N de la population est aléatoire ou certaine. On désignera par μ la mesure moyenne de q_0 .

Soit q_i $i = 1, \ldots, n$ n observations indépendantes du processus q_0 et soit $q_{(n)}$ leur superposition. On désignera par m_n l'effectif aléatoire de $q_{(n)}$ sur $\mathbf{R}_+$ et par $X_i^{(n)}$ les points de cette superposition ordonnés tels que $X_1^{(n)} \leq \ldots \leq X_{m_n}^{(n)}$. Si z est le seuil de pauvreté alors, dans le cadre des processus ponctuels, $q_0([0, z[)$ est l'effectif aléatoire de q_0 sur l'intervalle $[0, z[$ c'est- à -dire le nombre d'individus dont le revenu est strictement inférieur à z ; nous le désignerons par $q(z)$. L'indice de pauvreté d'ordre α de F.G.T est donné par

$$P(z, \alpha) = \int_0^z (1 - \frac{x}{z})^\alpha \, d\mu(x)$$

dont un estimateur naïf est

$$P_n(z, \alpha) = \frac{1}{n} \sum_{i=1}^{m_n} \left(1 - \frac{X_i^{(n)}}{z}\right)_+^\alpha$$

où $x^+ = x$ si $x > 0$ et 0 sinon

Cet estimateur est sans biais G.Dia [3] théorème 2.1 et converge presque sûrement d'après la loi forte de Kolmogorov.

Nous nous proposons de construire dans cette étude un estimateur à pas fixe de cet indice.

D'une manière générale l'observation de la répartion ponctuelle n'est que partielle pour une raison évidente de coût par exemple. Ce que l'on observe est un **processus aminci** ou **thinning process**, c'est -à-dire seulement une partie du processus, certains points sont éliminés.

5 P-amincissement

Soit donc q_0 un processus ponctuel défini sur $(\Omega, \mathcal{T}, P)$ à valeurs dans $\mathbf{R}_+$ muni de sa tribu de Borel $\mathcal{B}(\mathbf{R}_+)$. On définit sa mesure moyenne par

$$\forall A \in \mathcal{B}(\mathbf{R}_+) \qquad E(q_0(A)) = \mu(A)$$

où E désigne l'espérance mathématique.

Dans toute la suite de cette étude on supposera que le processus q_0 satisfait à l'hypothèse suivante : pour tout ouvert non vide Θ de $\mathbf{R}_+$ on a $P(q_0(\Theta) > 0) > 0$. On pose $q_0(\mathbf{R}_+) = q$

Soit p une fonction mesurable définie sur $[0, z]$ à valeur dans $[0, 1]$ et soient $U_i \quad 1 \leq i \leq q$ des variables aléatoires à valeurs dans $\{0, 1\}$ telles que

1. les U_i sont conditionnellement indépendantes de q
2. $P(U_i = 1/\mathcal{Q}) = p(X_i)$, $\mathcal{Q}$ désignant la tribu engendrée par q

Le processus ponctuel q' défini par

$$q\prime(A) = \sum_{i=1}^{q} U_i \delta_{X_i}(A) = \sum_{i=1}^{q} U_i 1_A(X_i) \quad \forall A \in \mathcal{B}(\mathbf{R}_+)$$

est appelé p-amincissement de q_0 où $\delta_{X_i}(A) := 1_A(X_i)$ est la mesure de Dirac de masse égale à 1 au point X_i.

Ce processus est obtenu de la façon suivante : X_i est retenu avec la probabilité $p(X_i)$ et élimé avec la probabilté $1 - p(X_i)$ et ceci indépendamment des autres points conditionnellement à q. Nous avons le

Théorème 1 *Soit p une fonction mesurable sur $\mathbf{R}_+$ à valeurs dans $[0, 1]$. Si $P(U_i = 1/\mathcal{Q}) = p(X_i)$ alors*

1. $P(U_i = 1/X_i = x) = p(x)$
2. $\forall A \in \mathcal{B}(\mathbf{R}) \qquad E(q\prime(A)) = \int_A p(x)\, d\mu(x) = E(\sum_{i=1}^{q_0(A)} 1_A(X_i) p(X_i))$

Preuve :

$$\forall B \in \mathcal{Q} \qquad \int_B P(U_i = 1/\mathcal{Q})\, dP = P(U_i = 1, B) = \int_B p(X)\, dP$$

Prenons $B = X_i^{-1}(A)$ il vient

$$P(U_i = 1, X_i^{-1}(A)) = \int_{X_i^{-1}(A))} p(X)\, dP = \int_A p(x)\, dF_{X_i}(x)$$

Mais

$$P(U_i = 1, X_i^{-1}(A)) = \int_A P(U_i/X_i = x) dF_{X_i}(x)$$

d'où

$$P(U_i = 1/X_i = 1) = p(x)$$

Pour montrer la deuxième égalité écrivons

$$\forall A \in \mathcal{B}(\mathbf{R}) \quad E(q\prime(A)) = E\big(E(\sum_{i=1}^{q} U_i 1_A(X_i)/q)\big) = \sum_{k=1}^{+\infty} \sum_{i=1}^{k} E(U_i 1_A(X_i)/q = k) P(q = k)$$

Dans cette somme on ne considère que les indices i pour lesquels $(U_i = 1, X_i \in A, q(A) = k)$

D'où le second membre s'écrit

$$\sum_{k=1}^{+\infty} \sum_{i=1}^{k} P(U_i = 1, X_i \in A, q = k)$$

.

$(X_i \in A, q = k)$ est élément de la tribu $\mathcal{Q}$. D'où

$$\begin{aligned} P(U_i = 1, X_i \in A, q = k) &= \int_{(X_i \in A, q=k)} p(X_i)\, dP = \int_{q=k} 1_{X_i^{-1}(A)} p(X_i)\, dP \\ &= E(1_{X_i^{-1}(A)} p(X_i)/q = k) P(q = k) \end{aligned}$$

D'où

$$\begin{aligned} E(q\prime(A)) &= \sum_{k=1}^{+\infty} \sum_{i=1}^{k} E(1_{X_i^{-1}(A)} p(X_i)/q = k) P(q = k) && (1) \\ &= E\big(\sum_{i=1}^{q} 1_{X_i^{-1}(A)} p(X_i)\big) && (2) \\ &= E(\sum_{i=1}^{q} 1_A(X_i) p(X_i)) && (3) \end{aligned}$$

Comme $1_A(X_i)p(X_i)$ est borné, d'après le théorème 2.1 de Dia [1990] l'expression (3) est égale à $\int_A p(x)\, d\mu(x)$

Pour montrer la dérnière égalité permutons les signes de sommation de (1) soit

$$\sum_{i=1}^{+\infty} \sum_{k=i}^{+\infty} P(U_i = 1, X_i \in A, q = k) \quad \text{ou encore} \sum_{i=1}^{+\infty} P(U_i = 1, X_i \in A, q \geq i)$$

On a l'égalité

$$\begin{aligned} (U_i = 1, X_i \in A, q \geq i) = (U_i = 1, X_i \in A, q_0(A) \geq i) && (4) \\ && (5) \end{aligned}$$

D'où

$$\sum_{i=1}^{+\infty} \sum_{k=i}^{+\infty} P(U_i = 1, X_i \in A, q = k) = \sum_{i=1}^{+\infty} \sum_{k=i}^{+\infty} P(U_i = 1, X_i \in A, q_0(A) = i) \qquad (6)$$

et le second membre de (6) s'écrit encore $\sum_{k=1}^{+\infty} \sum_{i=1}^{k} P(U_i = 1, X_i \in A, q_0(A) = k)$

Comme $(X_i \in A, q_0(A) = k)$ est élément de la tribu $\mathcal{Q}$, on a

$$\begin{aligned} P(U_i = 1, X_i \in A, q_0(A) = k) &= \int_{(X_i \in A, q_0(A)=k)} p(X_i)\, dP && (7) \\ &= \int_{q_0(A)=k} 1_{X_i^{-1}(A)} p(X_i)\, dP && (8) \\ &= E(1_{X_i^{-1}(A)} p(X_i)/q_0(A) = k) P(q_0(A) = k) && (9) \end{aligned}$$

D'où

$$E(q\prime(A)) = \sum_{k=1}^{+\infty} \sum_{i=1}^{k} E(1_{X_i^{-1}(A)} p(X_i)/q_0(A) = k) P(q_0(A) = k) \tag{10}$$

$$= E\Big(\sum_{i=1}^{q_0(A)} 1_{X_i^{-1}(A)} p(X_i)\Big) \tag{11}$$

$$= E(\sum_{i=1}^{q_0(A)} 1_A(X_i) p(X_i)) \tag{12}$$

Supposons que q_0 soit la répartition des revenus d'une population de taille N finie certaine. Soit Γ_k un échantillon de taille k tiré de cette population. Considérons U_i la variable aléatoiree définie de la façon suivante

$$U_i = \begin{cases} 1, & \text{si} \quad X_i \in \Gamma_k \\ 0, & \text{sinon} \end{cases}$$

Alors $p(x) = E(U_i)$ est la probabilité d'insertion de X_i dans l'échantillon. Elle est constante et égale à $k\mathcal{A}_{N-1}^{k-1}/\mathcal{A}_N^k$ soit k/N. Dans le cas où N est une variable aléatoire on supposera qu'elle admet une espérance mathématique $E(N) = \|\mu\|$ finie, $p(x)$ peut être estimé. Γ_k va nous servir de base pour construire un processus aminci $q\prime$.

On définit pour tout Borélien $A \in \mathcal{B}(\mathbf{R}_+)$

$$q\prime(A) = \text{Card}\Big\{i, \quad 1 \leq i \leq k, \quad U_i = 1 \quad X_i \in A\Big\}$$

et si z est le seuil de pauvreté

$$q\prime([0, z[) = \text{Card}\Big\{i, \quad 1 \leq i \leq k, \quad U_i = 1 \quad X_i \leq z\Big\}$$

D'après le théorème précédent $q\prime$ admet une mesure moyenne $\mu\prime$ de densité égale à $p\,d\mu$. Ainsi p étant connu, une statistique utilisant Γ_k peut estimer l'indice de pauvreté $P_\alpha(z)$ de F.G.T tel que défini précédemment.

6 l'estimateur de l'indice de F.G.T

Nous considérons les hypothéses suivantes

- $\mathbf{H_1}$) la mesure moyenne μ du processus q est finie et admet une densité f strictement positive sur $\mathbf{R}_+$
- $\mathbf{H_2}$) pour tout $x \in \mathbf{R}_+$ on

$$P(q_0([x, x + \delta x] > 1/q_0) = o(\delta x)$$

- $\mathbf{H_3}$) $P(q_0(I) = 1) \cong \mu(I)$ pour tout intervalle I de longueur arbitrairement petit

Soit k une fonction de n. On considère $[kz]$ ou $[.]$ désigne la fonction partie entière et on note

$$\begin{aligned} \Delta_{k,j} &= [\frac{j}{k}, \frac{j+1}{k}[, \quad j \in \mathbb{N} \\ \mathcal{J}_{n,j} &= \{i, i \geq 1 \quad X_i^{(n)} \in \Delta_{k,j}\} \\ \nu_{n,j} &= card\mathcal{J}_{n,j} \end{aligned}$$

$\Delta_{k,j}, j \in \mathbb{N}$ est une partition de $\mathbf{R}_+$.

Pour construire l'estimateur de $P_\alpha(z)$,nous avons besoin de l'estimateur $\widehat{f_n}$ de la densité f de μ. Il est bien connu qu'il s'exprime par :

$$\widehat{f_n}(x) = \frac{k}{n}\sum_{j=0}^{+\infty} 1_{\Delta_{k,j}}(x)\nu_{n,j}$$

Substituons cette valeur dans l'xpression intégrale de $P_\alpha(z)$. Il vient, en notant par $P_n(z,\alpha)$ l'estimateur de $P_\alpha(z)$

$$P_n(z,\alpha) = \frac{k}{n}\sum_{j=0}^{+\infty}\int_0^z (1-\frac{x}{z})^\alpha 1_{\Delta_{k,j}}(x)\nu_{n,j}\,dx$$

Comme la mesure de $\Delta_{k,j}$ est égale à $\frac{1}{k}$ et tend vers 0 lorsque $k \to +\infty$, on peut écrire

$$\int_{\Delta_{k,j}} (1-\frac{x}{z})^\alpha\,dx \cong \frac{1}{k}(1-\frac{j}{kz})^\alpha$$

Les ensembles $\Delta_{k,j}$ qui interviennent dans l'intégrale $P(z,\alpha)$ sont intérieurs au segment $[0,z]$

L'estimateur $P_n(z,\alpha)$ de $P(z,\alpha)$ que nous proposons et que nous allons étudier est défini alors par :

$$P_n(z,\alpha) = \frac{1}{n}\sum_{j=1}^{[kz]}(1-\frac{j}{kz})^\alpha\,\nu_{n,j}$$

7 Convergence de l'estimateur $P_n(z,\alpha)$

α est supposé réel supérieur ou égal à 1.

Théorème 2 *L'estimateur $P_n(z,\alpha)$ est asymptotiquement sans bias i.e*

$$\lim_{n\to+\infty} E(P_n(z,\alpha)) = P(z,\alpha)$$

Preuve : On a

$$E\Big(P_n(z,\alpha)\Big) = \frac{1}{n}\sum_{j=1}^{[kz]}(1-\frac{j}{kz})^\alpha\,n\mu(\Delta_{k,j})$$

Considérons la suite de fonctions simples

$$\widetilde{f_n}(x) = \sum_{j=1}^{[kz]} 1_{\Delta_{k,j}}(x)(1-\frac{j}{kz})^\alpha$$

On a évidemment $|\widetilde{f_n}(x)| \leq 1$ et pour tout x dans $\Delta_{k,j}$ les inégalités $\frac{j}{kz} \leq \frac{x}{z} \leq \frac{j+1}{kz}$. Ce qui implique

$$(1-\frac{j+1}{kz})^\alpha \leq (1-\frac{x}{z})^\alpha \leq (1-\frac{j}{kz})^\alpha \leq (1-\frac{j-1}{kz})^\alpha$$

d'où

$$|(1-\frac{j}{kz})^\alpha-(1-\frac{x}{z})^\alpha| \leq |(1-\frac{j-1}{kz})^\alpha-(1-\frac{j+1}{kz})^\alpha| \tag{13}$$

Considérons la fonction $g(x)=(1-\frac{x}{z})^\alpha$ en lui appliquant le théorème des accroissements finis entre $\frac{j-1}{k}$ et $\frac{j+1}{k}$, le second membre de (13) est inférieur à $\frac{2\alpha}{kz}$ quantité qui tend vers zéro si $n\to+\infty$. D'où $\widetilde{f}_n(x)$ tend vers $(1-\frac{x}{z})^\alpha$ si $n\to+\infty$. Le théorème de la convergence dominée de Lebesgue entraine que $E\Big(P_n(z,\alpha)\Big)\to\int_0^z(1-\frac{x}{z})^\alpha\,d\mu(x)$ lorsque $n\to+\infty$

L'estimateur est donc asymptotiquement sans biais.

Considérons la partition suivante de $\mathcal{J}_{n,r}$

$\mathcal{J}_{n,r}=\bigcup_{s=1}^n\mathcal{J}_{n,r}^{(s)}$ où $\mathcal{J}_{n,r}^{(s)}$ est l'ensemble des indices i tels que $X_i^{(n)}$ est élément de la s ième composante de $q_{(n)}$ soit q_s et appartient à $\Delta_{k,r}$. Soit $\nu_{n,r}^{(s)}=card\mathcal{J}_{n,r}^{(s)}$

On peut écrire $\nu_{n,j}=\sum_{s=1}^n\nu_{n,j}^{(s)}$; d'où il vient

$$P_n(z,\alpha)=\frac{1}{n}\sum_{j=1}^{[kz]}\sum_{s=1}^{n}\nu_{n,j}^{(s)}(1-\frac{j}{kz})^\alpha=\frac{1}{n}\sum_{s=1}^{n}\sum_{j=1}^{[kz]}\nu_{n,j}^{(s)}(1-\frac{j}{kz})^\alpha \tag{14}$$

Pour $s=1,\dots,n$ les variables $\zeta_{n,s}=\sum_{j=1}^{[kz]}\nu_{n,j}^{(s)}(1-\frac{j}{kz})^\alpha$ sont indépendantes et de même loi.

Posons $q_0(\Delta_{k,j})=\nu_{n,j}^{(0)}$ et considérons les hypothèses suivantes :

$$H_4)\qquad |\frac{cov(\nu_{n,j}^{(0)},\nu_{n,i}^{(0)})}{\sqrt{Var(\nu_{n,j}^{(0)})Var(\nu_{n,i}^{(0)})}}|\leq\rho(k) \text{ avec } k\rho(k)\to 0, n\to+\infty$$

où ρ est une fonction réelle positive (15)

$$H_5)\qquad E(q_0^2)<+\infty$$

Théorème 3 *Supposons que les hypothèses H_i $i=1,2,3,4,5$ soient verifiées, alors si $\frac{n}{k}\to+\infty$ on a*

$$Var(\zeta_{n,0})\to\int_0^z(1-\frac{x}{z})^{2\alpha}\,d\mu(x)\qquad n\to+\infty$$

Preuve :

$$Var(\zeta_{n,0})=\sum_{j=1}^{[kz]}Var(\nu_{n,j}^{(0)})(1-\frac{j}{kz})^{2\alpha}+\sum_{1\leq i\neq j\leq[kz]}cov(\nu_{n,j}^{(0)},\nu_{n,i}^{(0)})(1-\frac{i}{kz})^\alpha(1-\frac{j}{kz})^\alpha$$

D'après la représentation du théorème 1 on peut écrire

$$\nu_{n,j}^{(0)} = \sum_{i=1}^{q} 1_{\Delta_{k,j}}(X_i) \qquad \nu_{n,s}^{(0)} = \sum_{l=1}^{q} 1_{\Delta_{k,s}}(X_l)$$

En écrivant

$$Var(\nu_{n,j}^{(0)}) = E((\nu_{n,j}^{(0)})^2) - (E(\nu_{n,j}^{(0)}))^2$$

on a

$$E((\nu_{n,j}^{(0)})^2) = E(\sum_{i=1}^{q} 1_{\Delta_{k,j}}(X_i)) + E(\sum_{i\neq l} 1_{\Delta_{k,j}}(X_i)1_{\Delta_{k,j}}(X_l)) \quad \text{et} \tag{16}$$

$$E(\sum_{i\neq l} 1_{\Delta_{k,j}}(X_i)1_{\Delta_{k,j}}(X_l)) = \sum_{m=1}^{+\infty} \sum_{1\leq i\neq l\leq m} (1_{\Delta_{k,j}}(X_i)1_{\Delta_{k,j}}(X_l)/q=m)P(q=m) \tag{17}$$

$$= \sum_{m=1}^{+\infty} \sum_{1\leq i\neq l\leq m} P(X_i \in \Delta_{k,j}, X_l \in \Delta_{k,l}))P(q=m) \tag{18}$$

$$\leq o(\frac{1}{k})E(q(q-1)) \tag{19}$$

Les hypotèses H_1, H_2, H_3,H_5 impliquent que

$$E(\sum_{i\neq l} 1_{\Delta_{k,j}}(X_i)1_{\Delta_{k,j}}(X_l)) = o((\frac{1}{k})$$

d'où

$$E((\nu_{n,j}^{(0)})^2) = \mu(\Delta_{k,j}) + o(\frac{1}{k})$$

et donc $Var(\nu_{n,j}^{(0)}) = \mu(\Delta_{k,j}) + o(\frac{1}{k})$

L'hypothèse H_4 implique que

$$|\sum_{1\leq i\neq j\leq [kz]} cov(\nu_{n,j}^{(0)}, \nu_{n,i}^{(0)})| \leq O(k\rho(k))$$

Les coefficients $(1-\frac{j}{kz})$ étant bornés on a donc

$$Var(\zeta_{n,0}) \to \sum_{j=1}^{[kz]} \mu(\Delta_{k,j})(1-\frac{j}{kz})^{2\alpha} \qquad n \to +\infty$$

Mais

$$\sum_{j=1}^{[kz]} \mu(\Delta_{k,j})(1-\frac{j}{kz})^{2\alpha} \to \int_0^z (1-\frac{x}{z})^{2\alpha}\, d\mu(x)$$

d'où

$$Var(\zeta_{n,0}) \to \int_0^z (1-\frac{x}{z})^{2\alpha}\, d\mu(x) \qquad n \to +\infty$$

Corollaire 1 *Sous les hypothèses du théorème 2, l'estimateur $P_n(z,\alpha)$ est convergent en probabilité vers $P(z,\alpha)$ lorsque $n \to +\infty$*

Preuve : Il suffit d'écrire

$$P(|P_n(z,\alpha) - P(z,\alpha)| > \epsilon|) \leq P(|P_n(z,\alpha) - E(P_n(z,\alpha)| > \epsilon/2) + P(|E(P_n(z,\alpha)) - P(z,\alpha)| > \epsilon/2)$$

Le second terme du second membre de l'inégalité est nul pour n suffisamment grand car $P_n(z,\alpha)$ est asymptotiquement sans biais. Le premier terme est un $O(\frac{1}{n})$ d'après l'inégalité de Bienaymé Tchebichev, la décomposition (14) et la convergence de $Var(\zeta_{n,0})$.

Théorème 4 *Supposons que les hypothèses $H_i \quad i = 1,2,3,4,5$ soient verifiées, alors si $\frac{n}{k} \to +\infty$ et $\frac{n}{k^2} \to 0$ on a*

$$\sqrt{n}\frac{P_n(z,\alpha) - P(z,\alpha)}{\sqrt{P(z,2\alpha)}} \to N(0,1) \quad n \to +\infty$$

Preuve : Il suffit de démontrer que $\sqrt{n}(P(z,\alpha) - E(P_n(z,\alpha)) \to 0, \quad n \to +\infty$
On a

$$\int_0^z (1-\frac{x}{z})^\alpha \, d\mu(x) = \sum_{j=1}^{[kz]} \int_{j-1}^{j} (1-\frac{x}{kz})^\alpha \, d\mu(\frac{x}{k}) + \int_{[[kz],kz]} (1-\frac{x}{kz})^\alpha \, d\mu(\frac{x}{k}) \tag{20}$$

$$\text{et} \int_{j-1}^{j} (1-\frac{x}{kz})^\alpha \, d\mu(\frac{x}{k}) = \mu(\Delta_{k,j})(1-\frac{\theta_j}{kz})^\alpha : \theta_j \in \Delta_{k,j} \tag{21}$$

D'où

$$\sqrt{n}\Big(E(P_n(z,\alpha)) - P(z,\alpha)\Big) = \sqrt{n}\Big(\sum_{j=0}^{[kz]} ((1-\frac{j}{kz})^\alpha - (1-\frac{\theta_j}{kz})^\alpha)\mu(\Delta_{k,j}) - \int_{[[kz],kz]} (1-\frac{x}{kz})^\alpha \, d\mu(\frac{x}{k})\Big)$$

Il vient en appliquant le théorème des accroissements finis à la fonction $g(x) = (1-\frac{x}{z})^\alpha$ au points $\frac{j}{k}$ et $\frac{\theta_j}{k}$ et en majorant la dérnière intégrale que

$$\left|\sqrt{n}\Big(E(P_n(z,\alpha)) - P(z,\alpha)\Big)\right| \leq \sqrt{n}\Big(2\frac{\alpha}{k}\sum_{j=0}^{[kz]} \mu(\Delta_{k,j}) + \frac{1}{k}\Big) \leq \frac{\sqrt{n}}{k}\Big(2\alpha\mu([0,z]) + 1\Big)$$

D'où le résultat.

Remarque : Pour construire un intervalle de confiance on utilise le théorème précédant, en remplaçant le dénominateur par $P_n(z,2\alpha)$. La convergence est encore valide d'après les théorémes généraux sur les convergences de variables aléatoires. Ainsi un intervalle de confiance au niveau $1-\alpha$ de $P(z,\alpha)$ sécrit

$$\Big[P_n(z,\alpha) - n^{-\frac{1}{2}}q_{1-\alpha/2}\sqrt{P_n(z,2\alpha)} \quad , \quad P_n(z,\alpha) + n^{-\frac{1}{2}}q_{1-\alpha/2}\sqrt{P_n(z,2\alpha)}\Big]$$

où $q_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi normale standard.

Considérons l'hypothèse suivante :

$$H_6 : \qquad E(\nu_{n,j}^{(0)}\nu_{n,i}^{(0)}) = o(\frac{1}{k^2})$$

Théorème 5 *Supposons que les hypothèses* $H_i, \quad i = 1,2,3,5,6$ *soient vérifiées. Alors si* $\frac{n}{k} \to +\infty$ *on a*

$$Var(\zeta_{n,0}) \to \int_0^z (1-\frac{x}{z})^{2\alpha}\, d\mu(x) - \left(\int_0^z (1-\frac{x}{z})^{\alpha}\, d\mu(x)\right)^2 \qquad n \to +\infty$$

Preuve : il suffit de calculer

$$\sum_{1\leq i\neq j\leq [kz]} cov(\nu_{n,j}^{(0)}, \nu_{n,i}^{(0)})(1-\frac{i}{kz})^{\alpha}(1-\frac{j}{kz})^{\alpha} = \sum_{1\leq i\neq j\leq [kz]} E(\nu_{n,j}^{(0)}\nu_{n,i}^{(0)})(1-\frac{i}{kz})^{\alpha}(1-\frac{j}{kz})^{\alpha} - \sum_{1\leq i\neq j\leq [kz]} E(\nu_{n,j}^{(0)})E(\nu_{n,i}^{(0)})(1-\frac{i}{kz})^{\alpha}(1-\frac{j}{kz})^{\alpha} \tag{22}$$

Le premier terme du second membre tend vers zéro lorsque $n \to +\infty$ d'après l'hypothèse H_6 et le second membre s'écrit :

$$\sum_{1\leq i\neq j\leq [kz]} E(\nu_{n,j}^{(0)})E(\nu_{n,i}^{(0)})(1-\frac{i}{kz})^{\alpha}(1-\frac{j}{kz})^{\alpha} = \left(\sum_{i=1}^{[kz]} E(\nu_{n,i}^{(0)})(1-\frac{i}{kz})^{\alpha}\right)^2 - \sum_{i=1}^{[kz]}(E(\nu_{n,i}^{(0)}))^2(1-\frac{i}{kz})^{2\alpha} \tag{23}$$

Comme $E(\nu_{n,i}^{(0)}) = \mu(\Delta_{k,i})$, le premier terme du deuxième membre de (23) tend vers $\left(\int_0^z (1-\frac{x}{z})^{\alpha}\, d\mu(x)\right)^2$ et le second membre vers zéro lorsque $n \to +\infty$

On a le théorème 4 qui devient :

Théorème 6 *Supposons que les hypothèses* $H_i \quad i = 1,2,3,5,6$ *soient verifiées, alors si* $\frac{n}{k} \to +\infty$ *et* $\frac{n}{k^2} \to 0$ *on a*

$$\sqrt{n}\frac{P_n(z,\alpha) - P(z,\alpha)}{\sqrt{P(z,2\alpha) - (P(z,\alpha))^2}} \to N(0,1) \quad n \to +\infty$$

Examinons maintenant le cas particulier où q_0 est le processus ponctuel associé à une variable aléatoire X de la façon suivante :
pour tout borélien A de $\mathbf{R}_+$

$$q_0(A) = \begin{cases} 1 & \text{si } X \in A \\ 0 & \text{sinon} \end{cases}$$

On a alors $\mu(A) = P(X \in A)$. On retrouve le cas classique où la distribution des revenus est une variabe aléatoire de loi inconnue et superposer les n processus revient à ordonner les n variables d'un échantillon de taille n. Les hypothèses H_2, H_3, H_5 sont immédiatement vérifiées et $E(\nu_{n,j}^{(0)}\nu_{n,i}^{(0)}) = 0$. On a le

Théorème 7 *Supposons que l'hypothèse H_1 soit verifiée, alors si $\frac{n}{k} \to +\infty$ on a*

$$Var(\zeta_{n,0}) \to \int_0^z (1-\frac{x}{z})^{2\alpha} f(x)\,dx - \left(\int_0^z (1-\frac{x}{z})^{\alpha} f(x)\,dx\right)^2 \qquad n \to +\infty$$

De plus, si $\frac{n}{k^2} \to 0$ on a

$$\sqrt{n}\frac{P_n(z,\alpha)-P(z,\alpha)}{\sqrt{P(z,2\alpha)-(P(z,\alpha))^2}} \to N(0,1) \quad n \to +\infty$$

L'intervalle de confiance de niveau $1-\alpha$ pour $P(z,\alpha)$ est de la forme

$$\left[P_n(z,\alpha) - n^{-\frac{1}{2}} q_{1-\alpha/2}\sqrt{P_n(z,2\alpha)-(P_n(z,\alpha))^2} \quad , \quad P_n(z,\alpha) + n^{-\frac{1}{2}} q_{1-\alpha/2}\sqrt{P_n(z,2\alpha)-(P_n(z,\alpha))^2}\right]$$

Supposons que l'effectif de la population soit certain et égal à N. Et soit $A \in \mathbf{R}_+$. On a alors $\mu(q_0(A)) = \sum_{i=1}^{N} P(X_{(i)} \in A)$

Prenons k égale à 1. Le processus aminci résultant a pour mesure moyenne $\mu(q'_0(A)) = \frac{1}{N}\sum_{i=1}^{N} P(X_{(i)} \in A)$. Superposer n processus amincis indépendants revient à extraire de la population un échantillon ordonné de taille n d'une variable aléatoire $\widetilde{X}$ de distribution égale à $\frac{1}{N}\sum_{i=1}^{N} P(X_{(i)} \in (.))$ Le théorème suivant va nous donner de façon plus précise la définition de cette variable aléatoire.

Théorème 8 [?] *Si $q_0(\mathbf{R}_+) = N$ certain, il existe N variables aléatoires échangeables $\widetilde{X}_1, \ldots, \widetilde{X}_N$ définies sur un espace de probabilité $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ engendrant un processus ponctuel $\widetilde{q_0}$ de même distribution que q_0.*

Preuve : Désignons par $\mathcal{S}$ l'ensemble des permutations de $(1,\ldots,N)$, $\mathcal{C}$ la tribu des parties de $\mathcal{S}$ et Q la distribution uniforme sur $(\mathcal{S},\mathcal{C}$. Posons

$$(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P}) = (\Omega, \mathcal{A}, P) \times (\mathcal{S}, \mathcal{C}, \mathcal{Q})$$

Les variables $\widetilde{X}_1, \ldots, \widetilde{X}_N$ sur $\widetilde{\Omega}$ à valeurs dans $\mathbf{R}_+$ définies par

$$\forall \widetilde{\omega} = (\omega, \sigma) \in \widetilde{\Omega} \quad \widetilde{X}_j(\widetilde{\omega}) = X_{\sigma(j)}(\omega), \quad j = 1, \ldots, N$$

sont mesurables à valeurs dans $\mathbf{R}_+$ car pour tout Borélien B de $\mathbf{R}_+$ on a

$$\left\{\widetilde{\omega} \in \widetilde{\Omega} \quad \widetilde{X}_j(\widetilde{\omega}) \in B\right\} = \left\{(\omega,\sigma) \in \Omega\times\mathcal{S} \quad X_{\sigma(j)}(\omega) \in B\right\} = \bigcup_{s=1}^{N!}\left\{(\omega,\sigma_s) \in \Omega\times\{\sigma_s\} \quad X_{\sigma_s(j)}(\omega) \in B\right\}$$

$$= \bigcup_{s=1}^{N!} X^{-1}_{\sigma_s(j)}(B) \times \{\sigma_s\}.$$

Comme $X^{-1}_{\sigma_s(j)}(B)$ appartient à la tribu $\mathcal{A}$ on a donc la mesurabilité de $\widetilde{X}_j$ pour $j = 1, \ldots, N$ Ces variables aléatoires définissent un processus ponctuel $\widetilde{q}_0$.

Les deux suites $\widetilde{X}_1(\omega, \sigma), \ldots, \widetilde{X}_N(\omega, \sigma)$ et $X_1(\omega), \ldots, X_N(\omega)$ sont en relation biunivoque par définition des variables $\widetilde{X}_j(\widetilde{\omega})$. Deux éléments qui se correspondent ont la même valeur. L'effectif de chacune de ces suite sur un Borélien quelconque de $\mathbf{R}$ est donc le même. D'où q_0 et $\widetilde{q}_0$ ont la même loi.

Montrons que les variables $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_N$ sont échangeables. Soit σ_0 une permutation élément de $\mathcal{S}$.

On doit montrer que les vecteurs $(\widetilde{X}_{\sigma_0(1)}, \widetilde{X}_{\sigma_0(2)}, \ldots, \widetilde{X}_{\sigma_0(N)})$ et $(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_N)$ ont même loi.

On a par définition $\widetilde{X}_{\sigma_0(j)}(\omega, \sigma) = X_{\sigma o \sigma_0(j)}(\omega) \quad \forall(\omega, \sigma) \in \widetilde{\Omega}$

Comme la variable aléatoire $\sigma' = \sigma o \sigma_0$ est uniforme sur $(\mathcal{S}, \mathcal{C})$ et indépendante de ω les vecteurs $(X_{\sigma'(1)}, \ldots, X_{\sigma'(N)})$ et $(X_{\sigma(1)}, \ldots, X_{\sigma(N)})$ ont même loi.

Corollaire 2 *Soit $\widetilde{X}$ une variable aléatoire définie sur $(\mathcal{S}, \mathcal{C})$ de même loi que $\widetilde{X}_1$. Alors pour tout Borélien B de $\mathbf{R}+$ on a*

$$\widetilde{P}(\widetilde{X} \in B) = \frac{1}{N} \sum_{i=1}^{N} P(X_{(i)} \in B)$$

Preuve : Plongeons $(\Omega, \mathcal{A}, P)$ dans $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ en identifiant tout A de $\mathcal{A}$ à l'élément $A \times \mathcal{S}$ de $\widetilde{\mathcal{A}}$. Ainsi on a $\widetilde{P}(A \times \mathcal{S}) = P(A)$.

Alors de $\sum_{i=1}^{N} 1_B(\widetilde{X}_i) = \sum_{i=1}^{N} 1_B(X_i)$ pour tout Borélien B de $\mathbf{R}+$ il vient

$$E_{\widetilde{P}}(\sum_{i=1}^{N} 1_B(\widetilde{X}_i)) = \sum_{i=1}^{N} E_{\widetilde{P}}(1_B(X_i))$$

Le premier membre de cette égalité est égale à $N\widetilde{P}(\widetilde{X} \in B)$ car les variables aléatoires $\widetilde{X}_i, \quad i = 1, \ldots, N$ sont échangeables donc de même loi. Et le second membre égale à $\sum_{i=1}^{N} P(X_{(i)} \in B)$ car $\widetilde{P}$ est un prolongement de P

Ce corollaire nous montre que la distribution des revenus ne suit pas une loi simple. C'est un mélange de distributions de variables aléatoires ordonnées. D'où les difficultés qu'on a pour essayer de la réduire à une forme fonctionnelle [?]. Si on part déjà de l'idée que la loi de la distribution des revenus est la même pour tous les individus et si on désigne par X la variable revenu distribuée de façon indépendante sur la population alors les $X_{(i)} \quad i = 1, \ldots, N$ sont des statistiques d'ordre et on a $\sum_{i=1}^{N} P(X_{(i)} \in B) = NP(X \in B)$. D'où $\widetilde{P}(\widetilde{X} \in B) = P(X \in B)$. C'est à dire que $\widetilde{X}$ et X bien que n'étant pas définies sur le même espace de probabilité ont la même loi image. C'est précisément ce cas particulier que le modèle classique étudie.

8 Simulation

8.1 Cas de la variable aléatoire

Pour rendre compte de la qualité de notre estimateur, nous supposons qu'on dispose d'un échantillon de taille n d'une variable aleatoire de densité $f(x)$ sur $[0,1]$ définie par

$$f(x) = \begin{cases} \dfrac{1}{1-exp(-1)}exp(-x) & \text{si } x \in [0,1] \\ \\ 0 & \text{sinon} \end{cases}$$

Nous traitons le cas $\alpha = 2$. La courbe correspondante est appelée courbe de **sévérité de la pauvreté**. Le cas $\alpha = 1$ s'obtient de la même manière. Nous allons donc estimer

l'intégrale $P(z,2) = \dfrac{1}{1-exp(-1)} \displaystyle\int_0^z (1-\frac{x}{z})^2 exp(-x)\, dx$ par

$$P_n(z,2) = \frac{1}{n}\sum_{j=1}^{[kz]}(1-\frac{j}{kz})^2\,\nu_{n,j}$$

La valeur exacte de cette intégrale, obtenue par intégration par partie, est :

$$P(z,2) = (z^2 - 2z + 2 - 2exp(-z))((1-exp(-1))z^2)^{-1}$$

Nous tracerons cette courbe et la comparerons avec la courbe fournie par notre estimateur. Les étapes pour la construction de cette dernière sont les suivantes :

1. Simuler n variables uniformes sur $[0,1]$ soient U
2. Transformer ces variables pour obtenir des variables $V = -Log(1-(1-exp(-1))U)$
3. Choisir $k = \left[\sqrt{nLogn}\right]$ où [.] désigne la fonction partie entière
4. Enfin $P_n(z,2)$ est obtenue en l'égalant à une fonction $\Phi(z,V,2)$ construite en utilisant le logiciel **S-Plus**

On a choisi quatre valeurs de n. $n = 400$, $n = 10000$, $n = 20000$, $n = 100000$ et $P(z,2)$ et de $P_n(z,2)$ sont calculées en 20 points de l'intervalle $[0,1]$. Les résultats sont présentés dans le tableau suivant :

z	$P_n(z,2), n=400$	$P_n(z,2), n=10000$	$P_n(z,2), n=20000$	$P_n(z,2), n=100000$	$P(z,2)$
0.05	0.0163435	0.02169652	0.02456187	0.02450567	0.0263997
0.10	0.03610243	0.04598633	0.04947043	0.04984502	0.05144018
0.15	0.6258102	0.06979823	0.073658	0.07447239	0.07621944
0.20	0.08770182	0.09319975	0.09766862	0.09833546	0.100396
0.25	0.1123611	0.1162078	0.1212841	0.1216128	0.1239873
0.30	0.1358811	0.1387645	0.1443377	0.1444158	0.1470106
0.35	0.158514	0.1609645	0.1669593	0.1667172	0.1694823
0.40	0.180734	0.1827968	0.1891275	0.18851	0.1914183
0.45	0.2024248	0.2042819	0.2108109	0.2097748	0.212834
0.50	0.2233681	0.2254253	0.2319877	0.2305387	0.2337445
0.55	0.2436876	0.2461882	0.2526771	0.2508196	0.2541642
0.60	0.2633716	0.2665042	0.2728818	0.2706334	0.2274102
0.65	0.2824467	0.2863409	0.2926008	0.2899943	0.2935866
0.70	0.3008751	0.3056938	0.3118324	0.3089201	0.312616
0.75	0.3187211	0.3245768	0.3305817	0.327432	0.331208
0.80	0.3360172	0.3430065	0.3488573	0.3455402	03493748
0.85	0.3528611	0.3610004	0.3666753	0.3632511	0.3671284
0.90	0.3692988	0.3785657	0.3840605	0.3805731	0.3844803
0.95	0.3853703	0.33957175	0.4010256	0.39755123	0.4014416
1	0.4011382	0.4124695	0.4175889	0.4140827	0.4180233

FIG. 1 – Présentation des Coubes

La cinquième courbe est la courbe représentative de $P(z,2)$. La sixième représentation est une superposition des courbes $P(z,2)$ et $P_n(z,2)$ pour $n = 400$. On l'a aggrandie au bas de la Figure 1. On constate que l'approximation est d'autant meilleure que la taille de l'échantillon augmente. On vérifie ainsi le corollaire 1 établissant la convergence en probabilité de l'estimateur.

8.2 Cas du processus de Poisson

Considérons un processus de Poisson q_0 de mesure moyenne $\mu([0,x] = \int_0^x f(t)\,dt$ sur $[0,1]$.On prendra $f(t) = exp(t)(e-1)^{-1}$. Un tel processus vérifie toutes les hypothèses des théorèmes précédents. On simule ce processus de Poisson non homogène à l'aide d'un processus de Poisson homogène q' de paramètre $\lambda = F(1) = 1$. Si t est un point de q', il lui correspond le point x de q_0 défini par $x = F^{-1}(F(1)t) = F^{-1}(t)$. On a évidemment $F^{-1}(t) = (e-1)^{-1}Log(1+(e-1)t)$. La simulation d'un processus de Poisson homogène de paramètre λ est classique. On peut la rappeler :

1. Générer une variable uniforme U_i sur $[0,1]$
2. Calculer $W_i = \dfrac{-Log(U_i)}{\lambda}$
3. Calculer $T_{i+1} = T_i + W_i$
4. Si $T_i \geq 1$, on s'arrête sinon $i = i+1$ et on recommence par l'étape 1

Comme dans le sous-paragraphe précèdent, nous simulons des échantillons de processus de POISSON pour différentes valeurs de n ; $n = 5000$, $n = 10000$, $n = 20000$. La valeur exacte de $P(z,\alpha)$ est :

$$P(z,2) = \frac{-z^2 - 2z + 2exp(z) - 2}{(e-1)z^2}$$

On obtient le tableau suivant :

z	$P_n(z,2), n = 5000$	$P_n(z,2), n = 10000$	$P_n(z,2), n = 20000$	$P(z,2)$
0.05	0.008507191	0.008580076	0.009193228	0.00982208
0.10	0.01931172	0.01795005	0.01920905	0.01989407
0.15	0.03011824	0.02805797	0.02953111	0.03022361
0.20	0.04138718	0.03887795	0.0399563	0.40818673
0.25	0.05331028	0.05035868	0.05088993	0.05168729
0.30	0.06567026	0.06241187	0.062358	0.06283806
0.35	0.07844699	0.074993585	0.0743852	0.7427969
0.40	0.09164343	0.08796013	0.086999925	0.08602123
0.45	0.1052968	0.1015367	0.1002807	0.09807204
0.50	0.1194385	0.1156416	0.1142937	0.1104416
0.55	0.1341188	0.1302883	0.1290328	0.1231406
0.60	0.14949906	0.145562	0.1445736	0.1361787
0.65	0.165842	0.1615908	0.1609946	0.1495669
0.70	0.1834517	0.1786095	0.1786095	0.1633162
0.75	0.2021484	0.1967787	0.196603	0.1774383
0.80	0.2220548	0.2162221	0.2159865	0.1919449
0.85	0.2435869	0.2371402	0.2367299	0.2068484
0.90	0.2672591	0.2610306	0.2594378	0.2221615
0.95	0.2929506	0.2918916	0.2844706	0.2378975
1	0.3207868	0.3291846	0.3123363	0.2540699

FIG. 2 – La courbe supérieure est $P_n(z,2)$

Il est clair qu'au vu de ces valeurs notre estimateur est convergent.

8.3 Application

Les courbes d'indices de pauvreté pour les différentes régions du Sénégal peuvent être obtenues sur la base des données individuelles de l'ESAM (Enquête Sénégalaise auprès des Ménages). Les revenus, exprimés en francs CFA, étant répartis sur un large éventail allant de quelques milliers de francs à des millions de francs dans la région de DAKAR imposent un changement d'échelle pour obtenir une représentation des courbes. Autrement on aboutit à des blocages dûs à des problèmes de capacités de mémoire avec le logiciel **S-PLUS** que nous utilisons. Le théorème suivant nous permet d'effectuer ce changement sans modifier l'indice de pauvreté

Théorème 9 *Soit $T > 0$. Si X est la variable revenu de densité f. La transformation $U = \frac{X}{T}$ entraine la relation*

$$P_n(z,2) = \int_0^{\frac{z}{T}} \left(1 - \frac{u}{\frac{z}{T}}\right)^2 g(u)\,du$$

où g est la densité de U

Preuve : Il suffit d'écrire $P(U < u) = P(X < uT)$ d'où $g(u) = f(uT)T$ et par le changement de variable $u = \frac{x}{T}$ on a

$$P_n(z,2) = \int_0^{\frac{z}{T}} \left(1 - \frac{u}{\frac{z}{T}}\right)^2 Tf(uT)\,du$$

d'où le résultat.

Ce théorème nous permet de diviser tous les éléments de l'échantillon par une constante monétaire que nous avons prise égale à 10000. Nous obtenons un échantillon issu d'une loi de densité g. Le seuil de pauvreté correspondant est $\frac{z}{T}$ si z est celui de la distribution initiale. Le tableau suivant donne la sévérité de la pauvreté dans les différentes régions pour des seuils situés au premier, deuxième et troisième quartile. La dernière colonne donne la sévérité pour le seuil de pauvreté au Sénégal fourni par l'ESAM qui est de $z_0 = 143080$.

Région	Q_1	Q_2	Q_3	$P_n(Q_1,2)$	$P_n(Q_2,2)$	$P_n(Q_3,2)$	$P_n(z_0,2)$
KOLDA	29439.803	105301.787	116937.326	0.0379204	0.2835911	0.3120816	0.3685575
DAKAR	103770.3	713311.9	409457.8	0.07274386	0.4587244	0.3087006	0.1037759
ZIGUINCHOR	45519.35	103945.06	214847.93	0.08693749	0.1890754	0.339147	0.2478608
DIOURBEL	40888.91	85217.31	1668336.05	0.03634522	0.147766	0.3064992	0.2663997
SAINT-LOUIS	42435.294	87187.147	199444.722	0.05014932	0.1473584	0.3357802	0.2523313
TAMBA	36201.29	64037.04	119858.25	0.04893785	0.1295293	0.2785072	0.3286103
KAOLACK	44385.85	89617.09	196164.26	0.04741324	0.1458881	0.3298044	0.2483414
THIES	33465.432	84153.913	205000.037	0.05154648	0.177656	0.3728675	0.2864689
LOUGA	48995.81	101821.98	225814.33	0.02804116	0.1412928	0.3389804	0.2184974
FATICK	35388.82	70910.36	172206.41	0.04991782	0.1452275	0.343654	0.2983305

TAB. 1 – Sévérité de la Pauvreté

8.4 Conclusion

L'étude qui vient d'être faite pose d'autres pistes de recherches :

1. La généralisation de la méthode en utilisant la mêthode du noyau de Parzen- Rosenblatt dont la méthode du pas fixe n'est qu'un cas particulier.
2. L'efficacité de notre estimateur par rapport à celui utilisé par Kakwani [?] p.11 (nous pensons que le nôtre est plus efficace)
3. Moyennant certaines hypothèses, la convergence de $P_n(z,\alpha)$ doit être presque sûre. (Ceci nous est suggéré par l'évolution de la courbe $P_n(z,\alpha)$ lorsque n croît).
4. Notre estimateur est aussi une méthode de calcul d'intégrales. Comme nous disposons de la méthode classique de Monte Carlo pour de tels calculs, il est naturel de comparer ces deux méthodes.
5. Interpréter les résultats fournis par les différentes régions au vu des courbes obtenues
6. Etudier le cas des échantillons dépendants

Références

[1] **S.R.Chakravarty,R.Kanbur et D.Mukherjee** Population Growth and Poverty Measurement June 2002

[2] **R. Davidson and J.Y Duclos** Statistical Inference for Stochastic Dominance and for the measurement of poverty and inequality *Econometrica* Vol.68 No.6 (November 2000)1435-1468

[3] **G. Dia** Non-parametric Estimation of the Density of a Point Process *Stat.Prob.Letters N° 6,397-405*

[4] **D.Boccanfuso, B.Deculuvé, L. Savard** Poverty, Income Distribution and CGE modeling : Does the Functional Form of Distribution Matter? October 2003 *Preleminary draft*

[5] **J. Geffroy** Variables Aléatoires, Variables statistiques et Sous processus associés à un Processus ponctuel (1986) *Communication personnelle*

[6] **J.Geffroy -Quidel** Convergence stochastique des Répartitions ponctuelles aléatoires *Coimbra 1974* (Portugal)

[7] **Alan F. Kaar** Point Processes and their Statistical Inference sd Edition *Marcel Dekker, Inc*

[8] **N. Kakwani** Testing for Significance of Poverty Differences with Aplication to Côte d'Ivoire. *LSMS N° 62*

[9] **B. Lanuzel** Répartition ponctuelle aléatoire sur $\mathbb{R}$ et $\mathbb{R}^2$ *Thèse de Doctorat de $3^{ième}$cycle* (1969) Faculté des Sciences de Paris

[10] **G.S.LO** Comportement Asymptotique des Indicateurs de Pauvreté échantillonnés *Laboratoire de Méthodes qualitatives Faculté des Sciences Economiques* N° 4 2003 Univ. Ch. Anta. Diop Dakar

[11] **W.Ogriczac and Ruszczynski** Dual Stochastic Dominance and Related Mean-Risk Models *Rutcor Research Report* 10-2001, january,2001

[12] **T.Ould.Taleb** Les Mesures de la Pauvreté, de l'inégalté et l'impact de la croissance économique *Mémoire de D.E.A* U.F.R des Sciences Appliquées et de Technologie Univ. Gaston Berger Saint-Louis Janvier 2003

[13] **R-D. Reiss** A Course on Point Processes *Springer Verlag*

[14] **G ; Shakarishivili** Challenges in Measuring Poverty : Case Study of the CIS-Region *Draft Research Paper for Internatinal Policy Fellowship Program* January 2003

[15] **A. Soliz, L.Alejanro** Indices de Pauvreté Théorie et Application empirique *Mémoire de Licence Faculté des Sciences Eco et Soc Univ. Genève* Juin 1999.