THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité : Mathématiques Appliquées

Option : Statistique

présentée par

Cheikh Tidiane SECK

Pour obtenir le grade de

DOCTEUR DE l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

Estimation Non-paramétrique et Convergence Faible des Mesures de Pauvreté

soutenue le 23 Mars 2011

devant le jury composé de :

Président :	Galaye DIA	Université Gaston Berger
Directeurs :	Paul DEHEUVELS	Université Paris 6
	Gane Samb LO	Université Gaston Berger
Rapporteurs :	Joseph Ngatchou WANDJI	Université Nancy 1
	Dietmar PFEIFER	Université de Oldenburg
Examinateurs :	Richard EMILION	Université d'Orléans
	Michel BRONIATOWSKI	Université Paris 6

Table des matières

Remerciements

Je voudrais tout d'abord remercier les professeurs Gane Samb Lô et Paul Deheuvels qui ont encadré et orienté mes travaux de recherches vers un domaine vital de la Statistique qui est la théorie des processus empiriques.

Le professeur Gane Samb Lô a été mon maître durant tout mon cursis universitaire. Il a su m'inculquer par son enthousiasme, sa rigueur et sa disponiblité, le goût à la recherche. Depuis quelques temps, il a orienté mes travaux de recherches vers ce thème de la pauvreté qui est d'une grande actualité de nos jours. Cette présente thèse est l'aboutissement de ces travaux, dont il est le principal précurseur. Qu'il me soit permis de lui exprimer toute ma reconnaissance et ma profonde gratitude.

Le professeur Paul Deheuvels m'a fait l'honneur de m'accueillir dans son laboratoire en mettant à ma disposition tous les moyens nécessaires. De par sa générosité, sa disponibilité et son humilité, il a toujours répondu positivement à nos multiples sollicitations malgré ses nombreuses charges. Il n'a cessé de nous encourager et prodiguer des conseils. Son expérience et son sens aigü de la recherche nous ont été d'un grand apport dans la réalisation de ce mémoire. Je lui exprime ici toute ma reconnaissance et ma profonde gratitude.

Les professeurs Dietmar Pfeifer et Joseph Ngatchou Wandji m'ont fait l'honneur d'établir chacun un rapport pour cette thèse. Leurs commentaires et suggestions ont permis d'améliorer la qualité du manuscrit. Ils ont réalisé cette tâche ingrate malgré leurs multiples activités. Qu'ils reçoivent mes vifs remerciements et l'expression de ma très haute reconnaissance.

Je souhaite également adresser mes sincères remerciements au Professeur Galaye Dia pour ses conseils et ses encouragements et pour m'avoir fait l'honneur de présider le jury de thèse ; qu'il trouve ici l'expression de ma plus profonde gratitude.

Les professeurs Michel Broniatowski et Richard Emilion m'ont fait l'honneur de prendre part au jury de thèse ; qu'ils veuillent accepter mes sincères remerciements.

Ce travail a été réalisé en grande partie, au sein du Laboratoire de Statistiques Théoriques et Appliquées (LSTA). Je saisis cette opportunité pour dire combien la collaboration avec les chercheurs, les doctorants et docteurs du LSTA a été fructueuse pour moi ; qu'ils en soient vivement remerciés. Leur soutien a été sans faille ; mention spéciale à Salim Bouzebda pour sa disponibilité et sa générosité légendaire. Mes remerciements vont aussi à l'endroit des autres collègues de bureau Diallo, Koné, Layal, Lucien, Ozon et Cie, Issam, Cherfi et à Phillipe Saint-pierre pour sa bonne humeur constante.

Je ne saurais oublier dans mes remerciements les secretaires du LSTA : Mme Lamart, Mme Durrande et Mme Corentine pour leur accueil chaleureux et leur gentillesse, ainsi que l'émérite bibliotécaire Pascal Epron.

Mes remerciements vont enfin à l'endroit de ceux qui, de près ou de loin, ont contribué à l'aboutissement de ce travail ou ont manifesté leur sympathie ; qu'ils trouvent ici l'expression de ma plus haute reconnaissance.

Je dédie cette thèse

à mes parents,

à mes frères et soeurs,

à ma chère épouse,

et à mes amis.

Notations

Nombres

$\lfloor x \rfloor$: partie entière d'un nombre réel x.

$N(\varepsilon, \mathcal{F}, d)$: nombre de ε-recouvrement d'une classe de fonctions $\mathcal{F}$ pour la métrique d.

$N_{[]}(\varepsilon, \mathcal{F}, d)$: nombre de ε-crochet d'une classe de fonctions $\mathcal{F}$ pour la métrique d.

$N(\varepsilon, \mathcal{F}) = \sup_Q N(\varepsilon \|F\|_{2,Q}, \mathcal{F}, d_Q)$: nombre d'entropie uniforme, où F est une enveloppe de $\mathcal{F}$ et d_Q est la métrique de $L_2(Q)$.

$card(A)$: le cardinal d'un ensemble A.

Convergences

$\rightsquigarrow$: convergence en distribution.

$\stackrel{P}{\to}$: convergence en probabilité.

$\stackrel{p.s.}{\to}$: convergence presque sûre.

Normes

$\|x\| = \sum_{i=1}^{d} |x_i|$: norme euclidienne dans $\mathbb{R}^d$.

$\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$: norme suprémum d'une fonction $f : \mathbb{R}^d \to \mathbb{R},\ d \geq 1$.

$\|f\|_I = \sup_{x \in I} |f(x)|$: norme suprémum d'une fonction $f : I \to \mathbb{R}$.

$\|\Phi(f)\|_{\mathcal{F}} = \sup_{x \in \mathbb{R}^d} |\Phi(f)|$: norme suprémum d'une fonctionnelle $\Phi : \mathcal{F} \to \mathbb{R}$.

$\|f\|_{p,Q} = (\int |f|^p dQ)^{1/p}$: norme d'une fonction f dans l'espace $L_p(Q)$.

Espaces

$C[0,1]$: espace des fonctions réelles continues sur $[0,1]$.

$D[0,1]$: espace des fonctions réelles continues à droites, avec limite à gauche sur et définies sur $[0,1]$.

$C_b(\mathbb{R})$: espace des fonctions réelles continues et bornées sur $\mathbb{R}$.

$l^\infty(T)$: espace des fonctions réelles bornées sur T.

$\mathcal{M}_b(\mathbb{R}, [0,1])$: espace des fonctions monotones sur $\mathbb{R}$ à valeurs dans $[0,1]$.

Résumé

Cette thèse introduit tout d'abord une formule générale qui englobe toutes les mesures de pauvreté uni-dimensionnelles basées sur le revenu. Nous proposons ensuite deux types d'estimateurs non-paramétriques (à noyau et de type "plug-in") pour cet indice général de pauvreté, tout en étudiant leurs propriétés asymptotiques. Notre méthodologie , basée essentiellement sur la théorie moderne du processus empirique indexé des fonctions, offre un cadre global et rigoureux qui permet d'étudier, avec la même approche, le comportement asymptotique de tous les indices de pauvreté encore disponibles jusqu'ici dans la littérature. Nous obtenons la consistance forte uniforme d'une très large classe de mesures de pauvreté incluant presque tous les modèles d'indices proposés par les économistes, décomposables comme non-décomposables. Ce résultat est utilisé pour construire des intervalles de confiance simultanés, de niveau asymptotiquement optimal (100%). Un théorème central limite uniforme fonctionnel est également établi pour cette large classe d'indicateurs de pauvreté. Comme conséquence, des procédures d'inférence robustes , basées sur le noyau de covariance et utilisant un test de Wald, sont développées afin de comparer de façon non-ambiguë deux populations différentes en termes de pauvreté.

English summary

This dissertation first presents a general representation of poverty measures that concerns all uni-dimensional poverty measures based on the income distribution. We then, deals with two types of estimators of this general poverty index : a kernel one and a plug-in one, and analyze their asymptotic properties. Our methodology, essentially based on the modern theory of empirical processes indexed by functions, offers a general and rigorous framework, which allows to study in the same approach, the asymptotic behaviour of all the income-based poverty measures that are still available yet in the literature. We obtain the strong and uniform consistency of a very broad class of poverty measures including almost all the poverty indices proposed by economists, both decomposable and non-decomposable. This result applies for building simultaneous and accurate asymptotic confidence bands for the theoritical poverty index . A uniform functional central limit theorem is also established for this wide class of poverty measures. As a consequence, robust statistical inference procedures, based upon the covariance structure, are developped using a Wald test, in order to compare in a non-ambiguous manner two different populations in terms of poverty.

Keywords : General poverty index, Kernel estimates, Plug-in estimates, Confidence bands, Empirical process indexed by classes of functions, Weak convergence.

Introduction Générale

Les travaux de l'économiste anglais Amartya Sen en 1976 ont apporté, durant les trois dernières décennies, un souffle nouveau dans la manière d'appréhender la pauvreté. En effet, un progrés considérable a été noté dans la perception et l'analyse de celle-ci. De nombreuses approches, complémentaires les unes des autres, ont été proposées par les économistes et beaucoup de mesures (ou indices) de pauvreté ont été introduites dans la littérature, en suivant l'approche axiomatique de Sen (1976). Selon ce dernier, une mesure de pauvreté acceptable doit tenir compte de trois aspects essentiels de la pauvreté : la proportion d'individus vivant en dessous du seuil de pauvreté, l'écart entre le revenu moyen des individus pauvres et le seuil de pauvreté et enfin, l'inégalité de revenu entre les individus pauvres. La prise en compte de ces trois aspects a permis à Sen (1967) d'établir un certain nombre de principes normatifs (ou *axiomes*) qu'un bon indice de pauvreté doit satisfaire. Pour une revue détaillée de ces *axiomes* de pauvreté, nous renvoyons le lecteur à Foster (1984), Chakravarty (1990) et spécialement à Zheng (1997) qui a fait une synthèse exhaustive de l'ensemble des propriétés souhaitables pour une mesure de pauvreté, ainsi que leur interaction.

La plupart des études empiriques sur la pauvreté utilisent des mesures discrètes se limitant seulement à la distribution observée. Dans une telle approche, l'évaluation de la mesure de pauvreté devient considérablement problématique lorsque la taille de la population à étudier devient importante. De plus, les mesures de pauvreté estimées avec cette approche sont exactes et non aléatoires. Ce qui ne permet pas de faire de l'inférence statistique en tenant compte des fluctuations d'échantillonnage. Dans cette thèse, nous nous intéressons à un indice de pauvreté général considéré, sous sa forme continue, comme l'espérance mathématique des formes discrètes calculées à partir d'un échantillon aléatoire de n revenus tirés dans une population, dont la distribution est définie par une fonction de répartition continue $G(y)$. Ainsi, l'indice de pauvreté que nous voulons introduire est défini sous la forme générale suivante :

$$J = \int_0^z w(G(y), G(z)) f(y,z) dG(y), \tag{1}$$

où z est un réel positif appelé *seuil de pauvreté* ou encore (*ligne de pauvreté*), c'est à dire le niveau de revenu en dessous duquel la pauvreté est effective. Si Y est une variable aléatoire représentant le revenu d'un individu pris au hasard dans la population considérée, alors $G(y) = \mathbb{P}(Y \leq y), \forall y \in \mathbb{R}$. La fonction $f(y,z)$ mesure la pauvreté d'un individu au revenu y, relativement au seuil de pauvreté z. Autrement dit, elle exprime la contribution d'un individu ayant un revenu y, à la pauvreté globale de la population. On l'appelle *fonction de déprivation individuelle.* Enfin, $w(\cdot,\cdot)$ est une fonction de pondération qui affecte une importance d'autant plus grande que l'individu considéré est plus pauvre. Par soucis de cohérence avec l'approche axiomatique de Sen, les fonctions $w(\cdot,\cdot)$ et $f(\cdot,\cdot)$ seront supposées décroisantes par rapport à la variable revenu y.

L'estimation asymptotique de l'indice J défini par l'équation (1) a fait l'objet de nombreux travaux utilisant des approches différentes et particulières. Par exemple, Bishop et *al* (1997) ont établi la consistance et la normalité asymptotique de la mesure empirique de Sen (1976) et de ses composantes - le taux de pauvreté, le déficit moyen de pauvreté et l'indice de Gini pour les pauvres - en se basant sur la théorie des U-statistiques. Ils ont montré que la mesure empirique de Sen est un estimateur consistant pour l'indice théorique

$$S = 2\int_0^z \left(1 - \frac{G(y)}{G(z)}\right)\left(\frac{z-y}{z}\right) dG(y).$$

On peut remarquer que cet indice S peut se mettre sous la forme (1) en posant

$$w[G(y), G(z)] = 2\left(1 - \frac{G(y)}{G(z)}\right), \text{ et } f(y,z) = \left(\frac{z-y}{z}\right).$$

Dans la même veine, Zheng (2001) obtient, sous des conditions très douces sur la fonction de répartition $G(y)$ de la distribution du revenu, la normalité asymptotique d'une classe d'estimateurs d'indices de pauvreté décomposables (i.e. additivement séparables). En travaillant avec des seuils de pauvreté relatifs, il a également montré que la structure de covariance fini-dimensionnelle de ces indices de pauvreté décomposables peut être estimée de façon non-paramétrique, permettant du coup, de faire de l'inférence statistique avec ces indices. Notons que ces mesures de pauvreté décomposables forment une classe particulière englobée par l'indice J et sont obtenues en choisissant une fonction de déprivation individuelle $f(y,z)$ appropriée et une fonction de pondération $w(\cdot,\cdot)$ constamment égale à 1, i.e. $w[G(y),G(z)] = 1, \ \forall y \geq 0$ et $z > 0$ fixé.

En explorant une méthode non-paramétrique pour l'estimation de l'indice de pauvreté, Dia (2008) proposa un estimateur à noyau de Parzen-Rosenblatt pour l'indice de Foster-Greer-Thorbecke (1984) défini par

$$FGT(\alpha) = \int_0^z \left(\frac{z-y}{z}\right)^\alpha dG(y), \quad \alpha \geq 0.$$

Cet indice peut se mettre sous la forme (1), en prenant pour tout $y \geq 0$,

$$w[G(y),G(z)] := 1 \text{ et } f(y,z) = \left(\frac{z-y}{z}\right)^\alpha, \ \text{ pour } \ \alpha \geq 0.$$

Dia (2008) a également étudié les propriétés asymptotiques de ce nouvel estimateur à noyau, notamment les convergences en moyenne quadratique et presque sûre , uniformément en z sur des intervalles compacts. Cet estimateur sera rappelé au chapitre 2.

On peut remarquer que tous ces travaux et d'autres encore que nous ne citerons pas ici par soucis d'espace, s'appuient sur une démarche individuelle, utilisant classe spécifique de mesures de pauvreté pour étudier ses propriétés asymptotiques. L'objectif de cette thèse, par contre, est d'établir une théorie asymptotique globale (des estimateurs de pauvreté) qui unifie toutes ces démarches individuelles en considérant l'indice général de pauvreté J défini par l'équation (1). Autrement dit, une théorie qui permet d'étudier à la fois les mesures de pauvreté décomposables et les mesures de pauvreté non-décomposables (comme celle de Sen) dites aussi linéaires par rapport au revenu. Pour ce faire, nous nous appuyerons sur deux types d'approches.

La première est basée sur l'estimation par la méthode du noyau. Avec cette approche, nous étudierons la convergence presque sûre d'une suite d'estimateurs à noyau de l'indice J, uniformément en (z,w,f), où z varie dans intervalle compact et, w et f décrivent des classes de fonctions appropriées. Dans cette approche, notre méthodologie s'inspire principalement des travaux de Einmahl et Mason (2005) qui ont élaboré une méthodologie générale basée sur la théorie moderne du processus empirique, pour établir la consistance uniforme des estimateurs de type noyau avec des fenêtres aléatoires. Les principaux outils de cette méthode consistent en des inégalités exponentielles et des bornes de moment appropriées que nous rappelerons en annexe A.3. Nous avons également obtenu, avec cette approche, une loi uniforme du logarithme en probabilité pour l'indice général de pauvreté J. Une conséquence immédiate de ce résultat est la possibilté de construire des intervalles de confiance simultanés, de niveau asymptotiquement optimal.

La seconde approche, quant à elle, utilise un estimateur direct de type "plug-in" de l'indice J, obtenu par simple remplacement de la fonction de répartition $G(y)$ par son équivalent empirique $G_n(y) = n^{-1}card\{j : Y_j \leq y\}$, où $Y_1, \cdots, Y_n$ est une suite indépendante et identiquement distribuée de variables aléatoires de fonction de répartion $G(y)$. Avec cette approche, nous discuterons de la normalité asymptotique et de la convergence faible uniforme sur des classes de fonctions bornées d'une suite d'estimateurs convenablement centrés et normalisés de l'indice J. Ici, également notre méthodologie repose sur des outils du processus empiriques telles que la notion de classe de Donsker sous des conditions d'entropie métrique uniforme ou d'entropie crochet. Cette approche nous a permis de décrire la structure de covariance globale de tous les indices de pauvreté encore disponibles jusqu'ici dans la littérature et de pouvoir proposer des tests de comparaison de pauvreté (de dominance stochastique) robustes entre deux populations aux distributions de revenu différentes.

Cette théorie asymptotique globale et unifiée que nous présentons dans cette thèse sera articulée en six chapitres.

Au premier chapitre nous rappelons quelques généralités sur les mesures de pauvreté. Nous insisterons sur la démarche générale d'agrégation qui permet de construire une mesure de pauvreté synthétique dans une population donnée. Nous donnerons également quelques exemples de mesures de pauvreté discrètes ainsi que quelques axiomes fondamentaux qui les régissent.

Dans le deuxième chapitre nous exposerons notre premier résultat qui est la consistance forte uniforme d'un estimateur général à noyau pour l'indice de pauvreté J. Cet estimateur sera construit avec le noyau de Parzen-Rosenblatt et une fenêtre variable afin de garantir une large applicabilité. Une étude de simulation de données sera réalisée pour évaluer la performance de ce résultat.

Le chapitre trois sera consacré à l'étude de lois limites uniformes du logarithme pour l'estimateur général à noyau de pauvreté étudié au deuxième chapitre. Nous déduirons de ce résultat une méthode pour construire des intervalles de confiance simultanés et asymptotiquement optimaux. Ces intervalles de confiance sont ensuite utilisés pour tester l'adéquation d'un modèle de loi paramétrique aux données de revenu afin d'estimer le degré de pauvreté.

Au chapitre quatre nous établirons la normalité asymptotique d'un estimateur de type "plug-in" pour l'indice général de pauvreté J. Des intervalles de confiance seront construits en estimant de façon non-paramétrique la variance asymptotique ; puis, comparés aux intervalles de confiances asymtotiques et simultanés, construits avec la méthode du noyau au chapitre 3.

Dans le chapitre cinq la convergence faible de l'estimateur de type "plug-in" de J, introduit au chapitre quatre sera étudiée. Plus précisément, nous montrerons que cet estimateur convenablement centré et normalisé peut être considéré comme un processus stochastique convergeant en distribution, dans un espace de fonctions réelles bornées, vers un processus limite gaussien, avec une structure de covariance complétement déterminée.

Enfin, le chapitre six concernera une application de la structure de covariance obtenue au cinquième chapitre pour tester des hypothèses de dominance stochastique en termes de pauvreté. Nous illustrerons ce test sur des données sénégalaises fournies par l'ANSD (Agence Nationale de la Statistique et de la Démographie).

Nos résultats principaux sont les suivants :

- La consistance presque sûre uniforme des mesures de pauvreté empiriques, par rapport au seuil de pauvreté z, à la fonction de pondération $w(\cdot,\cdot)$ et à la fonction de déprivation individuelle $f(\cdot,\cdot)$. Ce résultat a fait l'objet d'un papier soumis à *Journal of Nonparametric Statistics*.

- La normalité asymptotique d'un estimateur général de pauvreté incluant toutes les mesures de pauvreté basées sur le revenu et encore disponibles jusqu'ici dans la littératue. Ce résultat généralise un papier (cf. [35]) publié dans la revue *C.R. Math. Rep. Acad. Sci. Canada. Vol. 31(2) 2009, pp 45-52.*

- La convergence faible uniforme des mesures de pauvreté empiriques ainsi que leur structure de covariance asymptotique. Ce résultat généralise un autre article (cf. [34]) publié dans *Communications in Statistics- Theory and Methods, 38, 2009, 3697-3704.*

- Une méthodologie générale d'inférence non-paramétrique pour tous les indices de pauvreté basés sur le revenu. Ce résultat donne un critère de dominance (ou une relation d'ordre) assez robuste pour classer, en termes de pauvreté, deux distributions de revenu avec le même seuil de pauvreté.

Mais avant de revenir à ces résultats, leurs preuves et leurs applications, nous rappelons au chapitre 1 suivant quelques aspects importants sur l'agrégation de la pauvreté, à partir de laquelle sont construites toutes les mesures de pauvreté basées sur le revenu. Les outils techniques qui ont permis de démontrer ces résultats sont présentés en annexe, afin de faciliter la lecture du document.

Chapitre 1

Généralités sur les indices de pauvreté

1.1 Introduction

Le phénomène de pauvreté préoccupe aujourd'hui l'humanité toute entière. Il se manifeste dans tous les pays du monde. Cependant, sa perception varie d'un pays à un autre ou d'une région à une autre. La pauvreté peut être perçue comme absolue ou relative. Elle peut aussi être conçue comme un déficit de revenus ou une absence de certaines capacités fonctionnelles élémentaires. Elle est corrélée à la vulnérabilité et à l'exclusion sociale et atteint aujourd'hui des proportions inquiétantes de par le monde. Il urge alors de développer des politiques hardies pour son éradication. Pour cela, il serait nécessaire de pouvoir d'abord quantifier la pauvreté.

Mesurer la pauvreté nécessite, selon les économistes, la résolution de deux questions fondamentales. La première est d'ordre méthodologique et concerne l'identification des pauvres par la détermination d'un niveau de référence z en-dessous duquel la pauvreté est effective. La seconde question est plutôt technique et, est liée à la construction d'indicateurs synthétiques aggrégés de pauvreté sur la base des informations disponibles.

La réponse à ces deux questions se fait en analysant un certain ensemble de données pertinentes, qui est censé nous donner l'information nécessaire sur la population que l'on étudie. Ces données proviennent souvent d'enquêtes dont la nature dépend des objectifs de l'étude. Il peut s'agir d'enquêtes sur le budget des familles, sur les habitudes de consommation et de nutrition, sur les conditions de vie des ménages, sur l'évolution des prix et l'emploi, etc.

Comme la pauvreté est un phénomène qui affecte la vie des gens dans beaucoup de domaines, il serait logique de travailler avec une grande variété de statistiques qui refléteraient le niveau de vie des gens qui vivent dans la pauvreté. Par exemple des statistiques sur l'accès à l'éducation, à la santé et aux autres services publics, les biens dont dispose un individu (voiture, appareils ménagers, type de logement, etc). Une étude réalisée avec ce type d'information serait une étude multi-dimensionnelle de la pauvreté. Une autre façon de faire serait de ne travailler qu'avec un seul type de donnée qui serait censé indiquer, à lui seul, le niveau de bien-être des populations. Il existe en principe deux caractères (ou variables) globalement acceptés qui pourraient remplir cette tâche : le revenu et les dépenses de consommation. Ce type d'approche correspond à une étude uni-dimensionnelle de la pauvreté.

De prime abord, l'étude uni-dimensionnelle peut paraître plus limitée, moins complète et donc moins pertinente qu'une étude multi-dimensionnelle ; mais cela n'est pas forcément le cas. En effet, si l'on tient compte des biens dont dispose un individu par exemple, il se peut que quelqu'un ne dispose pas d'un certain bien par un simple choix personnel et non pas à cause d'une incapacité de l'obtenir. Les préférences personnelles introduisent donc un biais difficile à corriger dans l'approche multi-dimensionnelle. Prenons par exemple le cas d'un individu qui décide de ne pas avoir de voiture à cause de ses convictions écologistes.

L'approche uni-dimensionnelle, en revanche, échappe à ce type de problème puisque l'on suppose que le caractère choisi reflète le niveau de bien-être, et cela indépendamment des choix que l'individu puisse faire. L'étude uni-dimensionnelle se révélerait donc préférable pour autant que le revenu ou la consommation soient de bons indicateurs du bien-être, ce qui est globalement accepté par les économistes.

Dans cette thèse nous nous intéressons aux mesures de pauvreté uni-dimensionnelles basées sur le revenu ou les dépenses de consommation ; mais nous utiliserons le terme générique de revenu pour faire allusion aux deux caractéres. Cette approche quantitative basée sur le revenu, notons-le, est la plus fréquente dans la littérature et ultilise deux concepts généraux de pauvreté : la pauvreté absolue et la pauvreté relative. La distinction entre ces deux concepts se précise au niveau de la fixation du seuil de pauvreté.

1.2 Seuil de pauvreté

La détermination du seuil de pauvreté est une opération complexe qui dépend fondamentalement du concept de pauvreté utilisé.

La pauvreté absolue renvoie à un concept absolu qui veut dire que la pauvreté d'un ménage (ou individu) ne dépend que de sa situation et non de celle du reste de la société. Sous cet angle, un ménage est dit pauvre s'il n'est pas en mesure de satisfaire les besoins élémentaires de base tels que le fait de se nourir, de se vêtir ou de disposer d'un toit adéquat. On définit alors le seuil de pauvreté comme le revenu nécessaire pour satisfaire ces besoins élémentaires de base. Si un ménage dispose de ce revenu, il ne sera pas pauvre même si son niveau de vie est très bas par rapport au reste de la société. Cette définition du seuil est plus adéquate pour les pays sous-développés, où la pauvreté peut impliquer des pénuries alimentaires ou l'impossibilité d'avoir un logement digne. Le seuil de pauvreté est généralement fixé par les autorités gouvernementales ou des experts économistes qui estiment alors les besoins vitaux nécessaires pour la survie d'un ménage (ou individu).

Concernant la pauvreté relative, le ménage (ou l'individu) est jugé par rapport au reste de la société. C'est à dire qu'un ménage est considéré comme pauvre, non pas parce qu'il n'a pas un certain niveau de vie donné ; mais parce que son niveau de vie est très bas si on le compare à ceux des autres ménages de la société. Cette caractérisation de la pauvreté est surtout utilisée pour les pays très développés, puisque les ménages pauvres dans ces pays sont en général en mesure de satisfaire les besoins de base élémentaires ; mais sont considérés comme pauvres parce que leur niveau de vie est tout de même bien inférieur au niveau de vie moyen du pays. Le seuil de pauvreté est choisi comme un fractile de la distribution du revenu. Généralement, on prend un pourcentage du revenu médian ou du revenu moyen du pays ; mais pour accentuer les effets des inégalités dans la distribution, on choisit un pourcentage du revenu moyen car, la moyenne est plus sensible aux valeurs extrêmes.

1.3 Agrégation de la pauvreté

Lorsque le seuil de pauvreté $z > 0$ est défini, différentes formules sont proposées par les économistes pour agréger les pauvretés individuelles en une mesure synthétique qui résume la situation de pauvreté de l'ensemble de la population étudiée.

Considérons une distribution de revenus positifs $y_1, \cdots, y_n$, où $n \geq 1$ est un entier naturel. Cette distribution peut être représentée par un vecteur $y = (y_1, \cdots, y_n)$, avec $y_j \in \mathbb{R}_+ = [0, \infty)$, $\forall j = 1, \cdots, n$. y_j est le revenu d'un individu j qui est classé comme pauvre si $y_j < z$. Ainsi, $\mathcal{Y} = \{y \in \mathbb{R}_+^n,\ n \geq 1\}$ est l'ensemble des distributions de revenu définies sur $\mathbb{R}_+$. Pour toute distribution $y \in \mathcal{Y}$, on note respectivement $n = n(y)$, $\mu(y)$ et $\sigma^2(y)$ la taille, la moyenne et la variance de la population correspondant à y. On suppose aussi, sans perte de généralité, que les composantes du vecteur y sont ordonnées, i.e. $y_1 \leq y_2 \leq \cdots \leq y_n$.

Définition 1.1. *Une mesure de pauvreté (ou encore indice de pauvreté) est une application*

$$\begin{aligned} P:\quad & \mathcal{Y} \times \mathbb{R}_+ \to [0,1] \\ & (y, z) \mapsto P(y, z), \end{aligned}$$

telle que la valeur $P(y, z)$ indique le degré ou le niveau de pauvreté associé à la distribution y, où $z \in \mathbb{R}_+$ représente le seuil de pauvreté.

Un exemple très célébre de mesure de pauvreté est *l'incidence de la pauvreté* (ou "headcount ratio" en anglais), i.e. le pourcentage d'individus vivant en-dessous du seuil de pauvreté. Cette mesure est définie par le rapport

$$H(y, z) = \frac{q}{n}, \tag{1.1}$$

où $q = q(y, z)$ est le nombre de revenus dans la distribution y qui sont inférieurs à z. La mesure de l'incidence permet d'appréhender l'étendue de la pauvreté ; mais n'apporte aucune information sur l'intensité de celle-ci. Une mesure qui tient compte de cet aspect est *l'intensité moyenne de la pauvreté* appelée aussi le déficit moyen de revenu des pauvres (ou "income gap ratio" en anglais). Cette mesure est définie par

$$I(y, z) = \frac{1}{nz} \sum_{j=1}^{q} (z - y_j). \tag{1.2}$$

Ce type d'indice est également critiqué car il est parfois insensible à une redistribution des revenus au sein du groupe des pauvres. Par exemple, un transfert de revenu d'un individu pauvre vers un autre individu pauvre sans que ce dernier ne franchisse la ligne de pauvreté, laisse inchangé la valeur de l'indice $I(y,z)$. C'est pour cette raison que Sen (1976) et d'autres économistes se sont tournés vers d'autres indices qui traduisent, de façon plus correcte, le niveau de pauvreté d'une population donnée. De tels indices, par définition, sont sensibles à la distribution du revenu parmi les pauvres car leur but est de corriger les défauts que présentent les indices $H(y,z)$ et $I(y,z)$ donnés ci-dessus. Pour ce faire, Sen s'appuie sur un certain nombre de principes normatifs ou *axiomes* qui doivent régir un bon indice de pauvreté. Ces *axiomes* sont largement discutés dans l'article de Zheng (1997). Nous donnons ci-dessous les trois *axiomes* qui sont jugés les plus importants.

Soit $x = (x_1, \cdots, x_n)$ et $y = (y_1, \cdots, y_n)$ deux distributions de revenu, et $z > 0$ le seuil de pauvreté commun aux deux distributions.

- ***Axiome de focalisation*** : La mesure de pauvreté reste inchangée, si le revenu d'un individu non-pauvre augmente, i.e.

$$P(x,z) = P(y,z) \quad si \quad \exists j/y_j > z,\ \forall i \neq j, x_i = y_i \ \ et \ \ x_j = y_j + r,$$

où $r > 0$ est un accroissement de revenu.

- ***Axiome de monotonie*** : Une réduction du revenu d'un individu pauvre doit accroître la mesure de pauvreté, i.e.

$$P(x,z) > P(y,z) \quad si \quad \exists j/y_j < z,\ \forall i \neq j, x_i = y_i \ \ et \ \ x_j = y_j - r,\ r > 0.$$

- ***Axiome de transfert*** : Un transfert de revenu d'un individu pauvre vers un individu moins pauvre doit accroître la mesure de pauvreté, i.e.

$$P(x,z) > P(y,z)$$

si

$$\exists i,j/y_j < z, y_j < y_i \ \forall k \neq i,j, x_k = y_k \ \ x_j = y_j - r \ \ et \ \ x_i = y_i + r,\ r > 0.$$

Fort de ces trois axiomes et d'autres encores moins évidents, Sen (1976) proposa un indice de pauvreté qui se définit comme une moyenne pondérée des déficits de revenu des individus pauvres, i.e.

$$S(y,z) = \frac{2}{(q+1)nz} \sum_{j=1}^{q} (q+1-j)(z-y_j). \tag{1.3}$$

Cet indice est évidemment sensible à la distribution du revenu parmi les pauvres car il affecte un poids plus important aux plus pauvres. Sen a aussi montré que son indice reflétait à la fois *l'incidence de la pauvreté* $H(y,z)$, *l'intensité moyenne* $I(y,z)$ et l'inégalité parmi les pauvres et pouvait s'écrire sous la forme

$$S(y,z) = H(y,z)[I(y,z) + \frac{q}{q+1}(1 - I(y,z))G(y,z)], \tag{1.4}$$

où

$$G(y,z) = 1 - \sum_{j=1}^{q}[2(q-j)+1]\frac{y_j}{q^2\mu_p}$$

est le coefficient de Gini correspondant à la distribution du revenu parmi les pauvres, avec $\mu_p = \frac{1}{q}\sum_{j=1}^{q} y_j$ le revenu moyen des pauvres.

1.4 Exemples de mesures agrégées

Après l'étude de Sen (1976), une très grande variété de mesures de pauvreté ont vu le jour en suivant son approche axiomatique. Ces mesures de pauvreté peuvent être divisées en deux classes. La première classe contient les mesures dites non-pondérées pami lesquelles, on peut citer la célébre famille d'indices de Foster-Greer-Thorbecke (FGT) (1984). Ces indices sont définis pour $\alpha \geq 0$, par

$$P(y,z,\alpha) = \frac{1}{n}\sum_{j=1}^{q}\left(\frac{z-y_j}{z}\right)^{\alpha}. \tag{1.5}$$

La mesure $P(y,z,\alpha)$ est fréquemment utilisée dans les études empiriques sur la pauvreté. Pour $\alpha = 0$, elle est réduite à $q/n = H(y,z)$, qui est *l'incidence de pauvreté* définie dans la section précédente. Lorsque $\alpha = 1$, elle est égale à *l'intensité moyenne de la pauvreté* $I(y,z)$. Pour $\alpha = 2$, elle est interprétée comme la sévérité de la pauvreté.

La famille d'indices de Chakravarty (1983) fait également partie de cette classe des mesures non-pondérées. Elle est définie par

$$C(y,z,\beta) = \frac{1}{n}\sum_{j=1}^{q}\left[1 - \left(\frac{y_j}{z}\right)^{\beta}\right], \quad 0 < \beta < 1. \tag{1.6}$$

De même l'indice de Watts (1968) appartient à cette classe. Il est défini par

$$W(y,z) = \frac{1}{n}\sum_{j=1}^{q}\log\left(\frac{y_j}{z}\right). \tag{1.7}$$

Ces mesures non-pondérées sont généralement de la forme

$$P(y,z) = \frac{1}{n}\sum_{j=1}^{q} f(y_j, z), \tag{1.8}$$

où $f(y_j, z)$ est une fonction qui mesure la pauvreté individuelle ; c'est à dire la contribution de l'individu j à la pauvreté globale de la population. $f(y_j, z) = 0$, si $y_j \geq z$, i.e. pour les non-pauvres.

La seconde classe de mesures de pauvreté regroupe les mesures dites pondérées, dont fait partie l'indice de Sen (1976) défini ci-dessus. Dans cette classe le poids affecté aux individus varie selon leur rang dans l'échelle des pauvres. Les éléments de cette classe sont généralement des modifications de la mesure de Sen (1976). Par exemple, Kakwani (1980) généralise cette mesure de Sen sous la forme

$$K(y,z,k) = \frac{q}{n\sum_{j=1}^{q} j^k}\sum_{j=1}^{q}(q+1-j)^k\left(\frac{z-y_j}{z}\right). \tag{1.9}$$

Pour $k = 1$, on a $K(y, z, 1) = S(y, z)$.
Les mesures de pauvreté de Thon et de Shorrocks sont également des modifications de l'indice de Sen. Thon (1979) proposa la mesure suivante :

$$T(y,z) = \frac{2}{n(n+1)}\sum_{j=1}^{q}(n+1-j)\left(\frac{z-y_j}{z}\right). \tag{1.10}$$

Tandis que Shorrocks (1995) introduisit la mesure

$$Sh(y,z) = \frac{1}{n^2}\sum_{j=1}^{q}(2n-2j+1)\left(\frac{z-y_j}{z}\right). \tag{1.11}$$

En résumé, on peut remarquer que pour un seuil de pauvreté donné $z > 0$ et une distribution finie $y = (y_1, \cdots, y_n)$ telle que : $y_1 \leq y_2 \leq \cdots \leq y_n$, toutes ces mesures de pauvreté, pondérées comme non-pondérées, peuvent se mettre sous la forme générale

$$J(y,z) = \frac{1}{n}\sum_{j=1}^{q} w\left(\frac{j}{n}, \frac{q}{n}\right) f(y_j, z), \tag{1.12}$$

où $w(\cdot,\cdot)$ est une fonction de pondération et $f(\cdot,\cdot)$ une fonction mesurable appropriée. L'étude des mesures de pauvreté a souvent été restreinte dans ce cadre des distributions discrètes finies. Les résultats obtenus sont alors nécessairement limités par une absence de procédures d'inférence statistique pouvant tenir compte de l'erreur d'échantillonnage. Lorsque l'on considére les revenus $(y_1, \cdots, y_n)$ comme une réalisation d'un échantillon aléatoire $(Y_1, \cdots, Y_n)$ provenant d'une population dont la distribution est définie par une fonction de répartition continue $G(y)$, alors on peut développer une théorie asymptotique globale et unifiée pour toutes ces mesures de pauvreté, en se basant sur l'indice empirique général $J(y, z)$ donné par (1.12). Cette théorie asymptotique générale des mesures de pauvreté constitue la principale contribution de ce mémoire et fera l'objet des chapitres suivants.

Chapitre 2

Consistance uniforme de l'estimateur à noyau de pauvreté

2.1 Introduction

Les indices de pauvreté sont utilisés systématiquement par les économistes pour apprécier le niveau de pauvreté des populations. Leur forme générale est donnée par

$$J = \int_0^z w[G(y), G(z)]f(y, z)dG(y), \tag{2.1}$$

où $z > 0$ est le seuil de pauvreté ; w et f sont des fonctions bi-variables appropriées, c'est à dire continues et décroissantes par rapport à la variable y en cohérence avec l'approche axiomatique de Sen sur les mesures de pauvreté. Toutefois, des hypothèses supplémentaires peuvent être imposées à w et f pour établir certains résultats.

Dans ce chapitre nous nous intéressons à l'estimation , par la méthode du noyau, de l'indice J à partir d'un échantillon $Y_1, \cdots, Y_n$ de n revenus d'individus pris au hasard dans une population donnée. Cet indice J, comme on l'a fait remarquer dans l'introduction générale, couvre une classe très large de mesures de pauvreté parmi lesquelles, on peut citer les mesures de Sen, de Shorrocks, de Foster-Greer-Thorbecke et de Kakwani qui sont fréquemment utilisées dans les études empiriques sur la pauvreté.

Ainsi, notre objectif consiste à établir une approche globale pour étudier la convergence presque sûre uniforme de tous les estimateurs de pauvreté basés sur le revenu. Pour cela, nous travaillerons avec un estimateur général à noyau et nous

nous appuyerons sur les travaux de Einmahl et Mason (2005) qui ont élaboré une méthodologie générale basée sur le processus empirique, pour établir la consistance uniforme des estimateurs de type noyau avec des fenêtres aléatoires.

2.2 Construction de l'estimateur à noyau

Soit Y une variable aléatoire (v.a.) correspondant au revenu d'un individu, pris au hasard au sein d'une population donnée. Supposons que l'on dispose d'un échantillon indépendant de n v.a. $Y_1, \cdots, Y_n$ de même loi que la variable Y définie par sa fonction de répartition $G(y) = \mathbb{P}(Y \leq y), \forall y \in \mathbb{R}$. On suppose de plus que $G(y)$ possède une densité de probabilité notée $g(y)$ par rapport à la mesure de Lebesgue. Alors l'indice de pauvreté J défini en (2.1) peut s'écrire sous la forme

$$J = \int_0^z w[G(y), G(z)] f(y, z) g(y) dy. \tag{2.2}$$

Un estimateur bien connu, à noyau $K(\cdot)$ et de fenêtre $h > 0$, de la densité $g(y)$ est

$$g_{n,h}(y) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{Y_j - y}{h}\right).$$

Les propriétés asymptotiques de cet estimateur de la densité sont largement étudiées dans la littérature. A titre d'exemples, on peut citer les travaux de Parzen (1962), Nadaraya (1965), Silverman (1978), Stute (1982b) et les travaux récents de Einmahl et Mason (2005) traitant avec des fenêtres aléatoires du type $h = h_n(Y_1, \cdots, Y_n; y)$.

Pour construire un estimateur à noyau de J, nous procédons à un simple remplacement de la densité $g(y)$ par son estimateur $g_{n,h}(y)$ dans l'expression (2.2). Nous obtenons ainsi un estimateur à noyau de l'indice de pauvreté J défini par

$$J_{n,h} = \frac{1}{nh} \sum_{j=1}^{n} \int_0^z w[G_n(y), G_n(z)] f(y, z) K\left(\frac{Y_j - y}{h}\right) dy, \tag{2.3}$$

où $G_n(y) = n^{-1} \sum_{j=1}^{n} 1_{\{Y_j \leq y\}}$ désigne la fonction de répartition empirique correspondant à la suite $Y_1, \cdots, Y_n$ et 1_A désigne la fonction indicatrice de l'ensemble A.

Un estimateur de ce type a été proposé par Dia (2008) dans le cas de l'estimation non-paramétrique des indices de Foster-Greer-Thorbecke. Ce dernier estimateur est défini par

$$P_n(z,\alpha) = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{\lfloor \frac{z}{h} \rfloor}\left(1-\frac{ih}{z}\right)^{\alpha} K\left(\frac{Y_j - ih}{h}\right), \tag{2.4}$$

où $\lfloor \cdot \rfloor$ désigne la partie entière et $h = h_n \to 0$, lorsque $n \to \infty$. Dia (2008) a étudié les propriétés asymptotiques de l'estimateur (2.4), notamment la convergence uniforme presque sûre et en moyenne quadratique par rapport à $z \in [0, b], b > 0$. Notre estimateur $J_{n,h}$ donné par (2.3) généralise l'estimateur de Dia donné par (2.4). En plus de la convergence uniforme presque sûre par rapport à z, nous allons étudier la convergence uniforme presque sûre par rapport aux fonctions w et f appartenant à des classes de fonctions appropriées.

Le noyau $K(.)$ dans les expressions ci-dessus désigne une fonction mesurable satisfaisant aux conditions (K.1-2-3-4) suivantes :

- (K.1) $K(\cdot)$ est à variation bornée.
- (K.2) $K(t) = 0$ si $|t| \geq a$, avec $a > 0$.
- (K.3) $\int_{\mathbb{R}} K(t)dt = 1$.
- (K.4) $\|K\|_\infty := \sup_{t\in\mathbb{R}} |K(t)| < \infty$.

La fenêtre $h > 0$ est souvent considérée comme une fonction dépendant seulement de la taille n de l'échantillon. En effet, beaucoup de travaux sur l'estimation par la méthode du noyau utilisent une séquence de fenêtres déterministes, considérée comme une suite de constantes positives h_n tendant vers zéro, lorsque $n \to \infty$. A ce propos, on peut citer par exemple Parzen (1962), Stute (1982b), Silverman (1986), Devroye (1987), Bosq et Lecoutre (1987), Nadaraya (1989), Wand et Jones (1995). Comme le choix de la fenêtre est crucial pour la convergence des estimateurs à noyau, par soucis de grande applicabilité, nous présenterons nos résultats avec une fenêtre variable et aléatoire qui peut dépendre aussi bien des données $Y_1, \cdots, Y_n$ que du point local y où l'on veut estimer la densité. Plus précisément, la fenêtre $h > 0$ variera dans un intervalle dont la longueur décroît vers zéro, lorsque la taille n de l'échantillon augmente.

2.3 Consistance uniforme presque sûre

L'indice théorique J et l'estimateur $J_{n,h}$ sont considérés ici, comme des fonctionnelles dépendant de $z > 0$ et des fonctions $w \in \mathcal{W}$ et $f \in \mathcal{F}$, où $\mathcal{W}$ et $\mathcal{F}$ sont des classes de fonctions appropriées. C'est à dire

$$J_{n,h} = \{J_{n,h}(z, w, f) : z > 0, w \in \mathcal{W}, f \in \mathcal{F}\}$$

et

$$J = \{J(z, w, f) : z > 0, w \in \mathcal{W}, f \in \mathcal{F}\}.$$

Pour simplifier les notations, introduisons la classe de fonctions suivante :

$$\Phi = \{\varphi = (w, f) : w \in \mathcal{W}, f \in \mathcal{F}\} = \mathcal{W} \times \mathcal{F}. \tag{2.5}$$

Dans la suite on supposera que $\mathcal{W}$ et $\mathcal{F}$ sont des VC-classes (classe de Vapnik-Červonenkis) de fonctions convenablement mesurables possédant chacune une fonction enveloppe finie. Compte tenu des propriétés d'un indice de pauvreté, on peut sans perte de généralités, supposer de plus que $\mathcal{W}$ et $\mathcal{F}$ sont des classes de fonctions à valeurs dans $[0, 1]$, donc uniformément bornées. Dans l'annexe A.2 nous rappelons les notions de VC-classe de fonctions et de classe de fonctions convenablement mesurable ; voir aussi van der Vaart and Wellner (1996) ou Kosorok (2006) pour plus de détails sur ces notions.

Pour tous $z > 0$, et $\varphi \in \Phi$ fixés, considérons la suite de fonctions réelles définies sur $\mathbb{R}_+$ par

$$d_{n,\varphi}(y, z) = w[G_n(y), G_n(z)]f(y, z)1_{\{y<z\}}, \quad \forall n \geq 1. \tag{2.6}$$

D'après la loi des grands nombres, $G_n(y) \to G(y)$ presque sûrement pour tout réel y, alors par continuité de la fonction w, on a pour tout $y \geq 0$, $d_{n,\varphi}(y, z)$ qui converge presque sûrement vers

$$d_\varphi(y, z) = w[G(y), G(z)]f(y, z)1_{\{y<z\}}, \tag{2.7}$$

lorsque $n \to \infty$.

Avec ces notations, $J_{n,h}$ et J dépendent maintenant de z et de φ, d'où les écritures suivantes :

$$J = \int_0^z d_\varphi(y, z)g(y)dy =: J(z, \varphi), \tag{2.8}$$

et

$$J_{n,h} = \int_0^z d_{n,\varphi}(y,z) g_{n,h}(y) dy =: J_{n,h}(z,\varphi). \tag{2.9}$$

Notre premier résultat est donné par le Théorème 2.1 ci-dessous qui donne la convergence presque sûre et uniforme de l'estimateur à noyau de pauvreté $J_{n,h}$.

Théorème 2.1. *Supposons que les conditions (K.1-2-3-4) sont remplies et que $g(y)$ est une densité continue sur $\mathbb{R}$. Alors pour toutes suites de nombres réels positifs a_n et b_n satisfaisant à*

$$0 < (c \log n)/n \leq a_n < b_n < 1, \forall c > 0, \quad b_n \to 0 et n a_n / \log n \to \infty$$

et pour tout $b > 0$, nous avons lorsque $n \to \infty$,

$$\lim_{n\to\infty} \sup_{a_n \leq h \leq b_n} \sup_{z \in [0,b]} \sup_{\varphi \in \Phi} |J_{n,h}(z,\varphi) - J(z,\varphi)| = 0 \qquad p.s. \tag{2.10}$$

Preuve. Nous décomposons la différence $J_{n,h}(z,\varphi) - J(z,\varphi)$ en trois termes.

$$\begin{aligned} J_{n,h}(z,\varphi) - J(z,\varphi) &= \int_0^z [d_{n,\varphi}(y,z) - d_\varphi(y,z)] g_{n,h}(y) dy \\ &\quad + \int_0^z d_\varphi(y,z) [\mathbb{E} g_{n,h}(y) - g(y)] dy \\ &\quad + \int_0^z d_\varphi(y,z) [g_{n,h}(y) - \mathbb{E} g_{n,h}(y)] dy \\ &=: T_{1,n}(z,\varphi,h) + T_{2,n}(z,\varphi,h) + T_{3,n}(z,\varphi,h). \end{aligned}$$

Pour prouver le théorème, nous devons montrer que chacun des trois termes ci-dessus converge presque sûrement vers 0, uniformément en z, φ, h, lorsque $n \to \infty$. Le comportement des deux premiers termes $T_{1,n}(z,\varphi,h)$ et $T_{2,n}(z,\varphi,h)$ n'est pas difficile à contrôler. Par contre, pour traiter le dernier terme $T_{3,n}(z,\varphi,h)$, nous faisons recours aux travaux de Einmahl et Mason (2005) qui ont utilisé une approche basée sur le processus empirique pour établir la convergence uniforme des estimateurs à noyau indexés par des classes de fonctions. Pour une extension de ces travaux, voir Mason et Swanepoel (2010).

On peut remarquer que pour tout $\varphi \in \Phi$, la fonction $d_\varphi(\cdot, z)$ définie en (2.7) est tronquée à droite de z puisqu'elle vérifie $d_\varphi(y,z) = 0$ si $y \geq z$. La continuité des fonctions w et f implique alors que $d_\varphi(\cdot, z)$ est bornée sur l'intervalle compact $[0, z]$, et donc sur $\mathbb{R}_+$ à cause de cet argument de troncature.

Les classes $\mathcal{W}$ et $\mathcal{F}$ étant convenablement mesurables par hypothèse, donc il en est de même pour la classe de fonctions

$$\Phi = \{\varphi = (w, f) : w \in \mathcal{W}, f \in \mathcal{F}\} = \mathcal{W} \times \mathcal{F}.$$

On en déduit que $\sup_{\varphi\in\Phi} |d_\varphi(y, z)|$ existe et est fini, pour tout $y \geq 0$. Donc,

$$\sup_{\varphi\in\Phi} ||d_\varphi(\cdot, z)||_\infty \leq M, \tag{2.11}$$

où M est une constante positive. Cela implique que la classe de fonctions $\{d_\varphi : \varphi \in \Phi\}$ est uniformément bornée.

Pour compléter la preuve du théorème, nous établissons les trois lemmes suivants :

Lemme 2.1. *Nous avons*

$$\sup_{a_n\leq h\leq b_n} \sup_{z\in[0,b]} \sup_{\varphi\in\Phi} |T_{1,n}(z, \varphi, h)| \to 0 \ p.s., \quad n \to \infty. \tag{2.12}$$

Preuve. Pour tout $z \in [0, b]$, on a

$$\sup_{a_n\leq h\leq b_n} \sup_{\varphi\in\Phi} |T_{1,n}(z, \varphi, h)| \leq \int_0^z \sup_{a_n\leq h\leq b_n} \sup_{\varphi\in\Phi} |[d_{n,\varphi}(y, z) - d_\varphi(y, z)]g_{n,h}(y)|dy.$$

Rappelons que la fonction w est continue, donc en appliquant le théorème de Glivenko-Cantelli ; puis, le lemme de Slutsky, on obtient compte tenu de (2.6) que la suite de fonctions $\{d_{n,\varphi}(\cdot, z)\}_{n\geq1}$ converge uniformément sur l'intervalle $[0, z]$ pour tout $z > 0$ et tout $\varphi \in \Phi$. Puisque la densité $g(y)$ est supposée également continue par hypothèse, donc elle est uniformément continue sur $[0, z]$ qui est un intervalle compact. Cela combiné à l'hypothèse $(K.2)$ qui stipule que le noyau $K(\cdot)$ est à support compact, implique (d'après le Théorème 1 de Einmahl et Mason (2005)) que

$$\sup_{a_n\leq h\leq b_n} \sup_{y\in[0,z]} |g_{n,h}(y) - g(y)| \to 0, \quad n \to \infty.$$

On en déduit alors que la suite de fonctions $\{[d_{n,\varphi}(\cdot, z) - d_\varphi(\cdot, z)]g_{n,h}(y)\}_{n\geq1}$ converge uniformément vers 0 sur l'intervalle $[0, z]$, pour tout $z > 0$ et tout $\varphi \in \Phi$. D'où

$$\sup_{a_n\leq h\leq b_n} \sup_{\varphi\in\Phi} \int_0^z |[d_{n,\varphi}(y, z) - d_\varphi(y, z)]g_{n,h}(y)|dy \to 0, \qquad n \to \infty.$$

On obtient le lemme 2.1 en prenant le suprémum de cette quantité par rapport à $z \in [0, b]$, pour tout $b > 0$.

Lemme 2.2. *Nous avons*

$$\sup_{a_n \le h \le b_n} \sup_{z\in[0,b]} \sup_{\varphi\in\Phi} |T_{2,n}(z,\varphi,h)| \to 0 \; p.s., \, n \to \infty. \tag{2.13}$$

Preuve. Pour tout $z \in [0,b]$, on peut écrire grâce à un changement de variables $t = (u-y)/h$,

$$\begin{aligned}
\sup_{a_n \le h \le b_n} \sup_{\varphi\in\Phi} |T_{2,n}(z,\varphi,h)| &\le \int_0^z \sup_{a_n \le h \le b_n} \sup_{\varphi\in\Phi} |d_\varphi(y,z)(\mathbb{E}g_{n,h}(y) - g(y))| dy \\
&\le z \times \sup_{a_n \le h \le b_n} \sup_{\varphi\in\Phi} \|d_\varphi(y,z)\|_\infty \sup_{y\in[0,z]} |\mathbb{E}g_{n,h}(y) - g(y)|. \\
&\le z \times \sup_{\varphi\in\Phi} \|d_\varphi(y,z)\|_\infty \sup_{a_n \le h \le b_n} \sup_{y\in[0,z]} |\int_{-a}^{a} K(t)[g(y+th) - g(y)]dt| \\
&\le z \times 2a\|K\|_\infty \sup_{\varphi\in\Phi} \|d_\varphi(y,z)\|_\infty \sup_{a_n \le h \le b_n} \sup_{|x-y|<ha,\, x,y\in[0,z]} |g(x) - g(y)|.
\end{aligned}$$

Puisque la densité $g(y)$ est uniformément continue sur $[0,z]$, alors

$$\sup_{a_n \le h \le b_n} \sup_{|x-y|<ha,\, x,y\in[0,z]} |g(x) - g(y)| \to 0, \quad \text{p.s.}, \quad n \to \infty$$

car la fenêtre h décroît vers 0, quand $n \to \infty$. En rappelent la borne (2.11), la preuve du lemme 2.2 est complétée en prenant le suprémum par rapport à z de cette quantité sur l'intervalle $[0,b]$, pour tout $b > 0$.

Lemme 2.3. *Nous avons*

$$\sup_{a_n \le h \le b_n} \sup_{z\in[0,b]} \sup_{\varphi\in\Phi} |T_{3,n}(z,\varphi,h)| \to 0, \; p.s., \, n \to \infty.$$

Pour étudier ce terme $T_{3,n}(z,\varphi,h)$, nous introduisons la suite de processus indexés par $y \in [0,z]$, $\varphi \in \Phi$ et $0 < h < 1$,

$$W_{n,h}(y,\varphi) = \sum_{j=1}^{n} d_\varphi(y,z) \left[K\left(\frac{y - Y_j}{h}\right) - \mathbb{E}K\left(\frac{y - Y_j}{h}\right)\right]. \tag{2.14}$$

Alors, le comportement asymptotique de $T_{3,n}(z)$ découle de la proposition suivante qui est une version du Théorème 4 de Einmahl et Mason (2005).

Proposition 2.1. *Supposons que les conditions (K.1-2-3-4) du Théorème 3.1 sont satisfaites et que la densité $g(y)$ est continue sur $\mathbb{R}$. Alors nous avons pour tout $c > 0$ et $0 < h_0 < 1$, avec probabilité 1,*

$$\lim_{n\to\infty} \sup_{\frac{c\log n}{n}\leq h\leq h_0} \sup_{z\in[0,b]} \sup_{\varphi\in\Phi} \frac{||W_{n,h}(\cdot,\varphi)||_{[0,z]}}{\sqrt{nh(|\log h| \vee \log\log n)}} =: A(c),$$

où $A(c)$ est une constante positive .

Preuve du Lemme 2.3. Soit $z \in [0, b]$, on a

$$\begin{aligned}
\sup_{a_n\leq h\leq b_n} \sup_{z\in[0,b]} \sup_{\varphi\in\Phi} |T_{3,n}(z)| &\leq \int_0^z \sup_{a_n\leq h\leq b_n} \sup_{\varphi\in\Phi} |d_\varphi(y,z)(g_{n,h}(y) - \mathbb{E}g_{n,h}(y))|dy \\
&\leq z \times \sup_{a_n\leq h\leq b_n} \sup_{\varphi\in\Phi} \|d_\varphi(\cdot,z)(g_{n,h} - \mathbb{E}g_{n,h})\|_{[0,z]}.
\end{aligned}$$

Remarquons que

$$\frac{||W_{n,h}(\cdot,\varphi)||_{[0,z]}}{\sqrt{nh(|\log h| \vee \log\log n)}} = \frac{\sqrt{nh}||d_\varphi(\cdot,z)(g_{n,h} - \mathbb{E}g_{n,h})||_{[0,z]}}{\sqrt{(|\log h| \vee \log\log n)}}.$$

Donc en appliquant la Proposition 2.1 ci-dessus et en prenant le suprémum sur l'intervalle $[0, b]$, pour tout $b > 0$, on obtient le lemme 2.3 ; ce qui achève la preuve du théorème 2.1.

Preuve de la Proposition 2.1. Soit α_n le processus empirique basé sur $Y_1, \cdots, Y_n$ et indexé par une classe de fonctions $\mathcal{G}$, i.e. pour tout $\eta \in \mathcal{G}$, on a

$$\alpha_n(\eta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\eta(Y_i) - \mathbb{E}\eta(Y_i))$$

Pour tout $y \in [0, z]$, $\varphi \in \Phi$ and $h > 0$, définissons la fonction

$$\eta_{y,\varphi,h}(u) = d_\varphi(y,z) K\left(\frac{y-u}{h}\right), \quad u \in \mathbb{R}^+.$$

Soit

$$\mathcal{G} = \{u \mapsto \eta_{y,\varphi,h}(u) : y \in [0, z[, \varphi \in \Phi, 0 < h < 1\}.$$

D'après (2.14), on peut écrire que

$$\sup_{0<h<1} \sup_{\varphi\in\Phi} \|W_{n,h}(\cdot,\varphi)\|_{[0,z]} = \|\sqrt{n}\alpha_n\|_{\mathcal{G}}.$$

Pour établir la Proposition 2.1, la classe $\mathcal{G}$ doit satisfaire aux trois conditions (C.1), (C.2) et (C.3) suivantes :

– (C.1) $\sup_{0<h<1} \sup_{\varphi\in\Phi} \sup_{y\in[0,z]} \|\eta_{y,\varphi,h}\|_\infty < \infty$.

– (C.2) $\exists C > 0, \forall 0 < h < 1, \sup_{\varphi\in\Phi} \sup_{y\in[0,z]} \mathbb{E}\eta^2_{y,\varphi,h}(Y) \leq Ch$.

– (C.3) $\mathcal{G}$ est une classe convenablement mesurable vérifiant la condition d'entropie uniforme, i.e.
$\exists C_0 > 0, \nu_0 \geq 0, N(\epsilon, \mathcal{G}) \leq C_0\epsilon^{-\nu_0}, 0 < \epsilon < 1$.

Ces conditions nous permettent d'utiliser une borne de moment (voir Proposition 1, Einmahl et Mason (2005)) combinée à une inégalité exponentielle due à Talagrand (1994), (voir aussi, Ledoux (1996)) pour contrôler la norme suprémum du processus empirique indexé par une classe de fonctions. Ces deux résultats (rappelés dans l'annexe A.3) sont les principaux outils pour établir la Proposition 2.1. Maintenant nous allons vérifier les trois conditions (C.1), (C.2) et (C.3).
La condition (C.1) découle immédiatement de la borne (2.11), car on a

$$\sup_{g\in\mathcal{G}} \|g\|_\infty \leq M\|K\|_\infty =: \kappa < \infty.$$

Pour vérifier la condition (C.2), nous écrivons

$$\begin{aligned}
\mathbb{E}\eta^2_{y,\varphi,h}(Y) = \mathbb{E}\left[d^2_\varphi(y,z)K^2\left(\frac{y-Y}{h}\right)\right] &\leq \|d_\varphi(\cdot,z)\|^2_\infty \int_{y-h/2}^{y+h/2} K^2\left(\frac{y-u}{h}\right) g(u)du \\
&\leq \|d_\varphi(\cdot,z)\|^2_\infty h \int_{-1/2}^{1/2} K^2(t)g(y-th)dt \\
&\leq \|d_\varphi(\cdot,z)\|^2_\infty \|g\|_\infty \|K\|^2_2 h.
\end{aligned}$$

Ce qui implique que pour tout $0 < h < 1$,

$$\sup_{\varphi\in\Phi} \sup_{y\in[0,z]} \mathbb{E}\eta^2_{y,\varphi,h}(Y) \leq \sup_{z\in[0,b]} \sup_{\varphi\in\Phi} \|d_\varphi(\cdot,z)\|^2_\infty \|g\|_\infty \|K\|^2_2 h =: Ch,$$

où C est une contante positive. D'où la condition (C.2).

La classe de fonctions $\mathcal{G}$ peut s'écrire de la manière suivante :

$$\mathcal{G} = \{d_\varphi(\cdot,z) : \varphi\in\Phi, z > 0\} \cdot \left\{K\left(\frac{y-\cdot}{h}\right) : y \in \mathbb{R}^+, 0 < h < 1\right\} =: \mathcal{D}\cdot\mathcal{K} \quad (2.15)$$

Nous pouvons aussi écrire $\mathcal{D}$ sous la forme

$$\mathcal{D} = \mathcal{D}_1 \cdot \mathcal{D}_2,$$

avec

$$\mathcal{D}_1 = \{y \mapsto w[G(y), G(z)] : w \in \mathcal{W}, z > 0\}$$

et

$$\mathcal{D}_2 = \{y \mapsto f(y, z) : f \in \mathcal{F}, z > 0\}.$$

Du fait que $\mathcal{W}$ et $\mathcal{F}$ sont par hypothèse des VC-classes convenablement mesurables, en combinant le lemme 2.6.18, page 147 van der Vaart et Wellner (1996) et le lemme 8.10, page 142 Kosorok (2006), on obtient que $\mathcal{D}_1$ et $\mathcal{D}_2$ sont aussi des VC-classes de fonctions convenablement mesurables . Donc, $\mathcal{D}_1$ et $\mathcal{D}_2$ possèdent chacune un nombre de recouvrement polynomial, i.e.

$$\exists C_1, \nu_1 > 0, \quad N(\epsilon, \mathcal{D}_1) \le C_1 \epsilon^{-\nu_1}, 0 < \epsilon < 1$$

et

$$\exists C_2, \nu_2 > 0, \quad N(\epsilon, \mathcal{D}_2) \le C_2 \epsilon^{-\nu_2}, 0 < \epsilon < 1.$$

Puisque $\mathcal{W}$ et $\mathcal{F}$ sont uniformément bornées, alors $\mathcal{D}_1$ et $\mathcal{D}_2$ sont aussi uniformément bornées et admettent toutes deux une fonction enveloppe $F = 1$. Ainsi, en appliquant le lemme A.1 de Einmahl et Mason (2000), on en déduit que la classe de fonctions $\mathcal{D}$ possède un nombre de recouvrement polynomial, i.e.

$$\exists C, \nu > 0, \quad N(\epsilon, \mathcal{D}) \le C \epsilon^{-\nu}, 0 < \epsilon < 1.$$

En utilisant encore le lemme 2.6.18 de van der Vaart et Wellner (1996), on montre, compte tenu du fait que le noyau $K(\cdot)$ est à variation bornée (i.e., $K = K_1 - K_2$, où K_1 et K_2 sont des fonctions croissantes), que la classe de fonctions $\mathcal{K}$ est une VC-classe convenablement mesurable, donc possède un nombre de recouvrement polynomial.

En appliquant de nouveau le lemme A.1 de Einmahl et Mason (2000) pour les classes $\mathcal{D}$ et $\mathcal{K}$, on en déduit que la classe de fonctions $\mathcal{G}$ définie en (2.15) possède un nombre de recouvrement polynomial. De plus, $\mathcal{G}$ est convenablement mesurable, car c'est le produit de deux classes convenablement mesurables (Lemma 8.10, Kosorok (2006)). Finalement $\mathcal{G}$ satisfait à la condition (C.3).

A partir de cette étape, la preuve suit exactement la même démarche que celle de Einmahl et Mason (2005), qui est aussi celle de Mason et Swanepoel (2010). Nous aurons besoin de la proposition et du corollaire suivants dans la suite de la preuve.

Proposition 2.2. *(Einmahl et Mason (2005))*
Soit $\mathcal{G}$ une classe convenablement mesurable de fonctions réelles bornées telles que, pour des constantes $C, \nu \leq 1$ et $0 < \sigma \leq \beta$ et pour une enveloppe G_0 de $\mathcal{G}$, les conditions suivantes sont satisfaites :

(i) $\mathbb{E}[G_0^2(Y)] \leq \beta^2$;

(ii) $N(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu}, 0 < \epsilon < 1$;

(iii) $\sigma_0^2 := \sup_{g \in \mathcal{G}} \mathbb{E}[g^2(Y)] \leq \sigma^2$;

(iv) $\sup_{g \in \mathcal{G}} \|g\|_\infty \leq \frac{1}{4\sqrt{\nu}} \sqrt{n\sigma^2 / \log(C_1\beta/\sigma)}$, *où* $C_1 = C^{1/\nu} \vee 1$.

Alors pour une constante positive A, on a

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i g(Y_i)\right\|_{\mathcal{G}} \leq A\sqrt{\nu n \sigma^2 \log(C_1\beta/\sigma)},$$

où $\varepsilon_1, \cdots, \varepsilon_n$ est une suite indépendante de variables aléatoires de Rademacher, indépendantes de $Y_1, \cdots, Y_n$.

Le corollaire (2.1) suivant affaiblit la condition (*iv*) qui est parfois difficile à manipuler.

Corollaire 2.1. *(Einmahl et Mason (2005))*
Supposons que $\mathcal{G}$ satisfait aux hypothèses de la proposition (2.2), et au lieu (iv) nous supposons que la condition (v) suivante est vraie :

$$(v) \sup_{g \in \mathcal{G}} \|g\|_\infty \leq U, \quad \text{où} \quad \sigma_0 \leq U \leq C_2\sqrt{n}\beta, \quad \text{et} \quad C_2 = \frac{1}{4\sqrt{\nu \log C_1}}.$$

Alors, on a

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i g(Y_i)\right\|_{\mathcal{G}} \leq A\{\sqrt{\nu n \sigma_0^2 \log(C_1\beta/\sigma_0)} + 2\nu U \log(C_3 n (\beta/U)^2)\},$$

où $C_3 = C_1^2/16\nu$.

Suite de la preuve

Soit $j, k \leq 0$ et $c > 0$, posons $n_k = 2^k$, $h_{j,k} = (2^j c \log n_k)/n_k$ et considérons la classe de fonctions

$$\mathcal{G}_{j,k} = \{\eta_{y,\varphi,h} : y \in [0, z], \varphi \in \Phi, h_{j,k} \leq h \leq h_{j+1,k}\}.$$

Alors pour tout $h_{j,k} \leq h \leq h_{j+1,k}$, on a

$$\sigma_0^2 := \sup_{\eta \in \mathcal{G}_{j,k}} \mathbb{E}[\eta^2(Y)] \leq 2Ch_{j,k} =: \sigma_{j,k}^2.$$

Maintenant nous allons appliquer pour chaque classe $\mathcal{G}_{j,k}$, le corollaire 2.1, pour obtenir une borne supérieure de la quantité

$$\mathbb{E}\| \sum_{i=1}^{n_k} \epsilon_i \eta(Y_i) \|_{\mathcal{G}_{j,k}}.$$

D'abord, nous observons que chaque $\mathcal{G}_{j,k}$ satisfait (i) avec $G = \beta = \kappa$. Ensuite, puisque $\mathcal{G}_{j,k} \subset \mathcal{G}$, nous avons ,via (C3) que chaque $\mathcal{G}_{j,k}$ remplit la condition (ii). En posant $\sigma^2 = \sigma_{j,k}^2$, nous remarquons aussi que $\mathcal{G}_{j,k}$ satisfait à la condition (iii). Finalement, au lieu (iv), nous observons que la condition (v) : $\sup_{\eta \in \mathcal{G}} \|\eta\|_\infty \leq U$ est satisfaite avec $U = \beta = \kappa$. Les conditions (i), (ii), (iii) et (v) du corrolaire 2.1 sont toutes vérifiées, donc on peut écrire que

$$\mathbb{E}\| \sum_{i=1}^{n_k} \epsilon_i \eta(Y_i) \|_{\mathcal{G}_{j,k}} \leq A\sqrt{\nu n_k \sigma_0^2 \log(C_1\beta/\sigma_0)} + 2A\nu U \log(C_2 n_k (\beta/U)^2),$$

où A, C_1, C_2 sont des constantes positives. Notons que

$$C_1 \frac{\beta}{\sigma_0} \leq \frac{\beta^2}{\sigma_0^2} \vee C_1^2$$

et que la fonction $h \mapsto h\log(h^{-1} \vee C_1^2)$ est décroissante pour $h \geq 0$. En tenant compte de l'inégalité $\sigma_0 \leq \sigma_{j,k} =: D_0 h_{j,k}$, on obtient, pour $j, k \geq 0$,

$$\begin{aligned}
\mathbb{E}\| \sum_{i=1}^{n_k} \epsilon_i \eta(Y_i) \|_{\mathcal{G}_{j,k}} &\leq A\beta\sqrt{\frac{\nu n_k D_0 h_{j,k}}{\beta^2} \log\left(\frac{\beta^2}{D_0 h_{j,k}} \vee C_1^2\right)} + 2A\nu\beta\log(C_2 n_k) \\
&\leq D_1 \sqrt{n_k h_{j,k} \log\left(\frac{1}{D_2 h_{j,k}} \vee C_1^2\right)},
\end{aligned}$$

où $D_1 = A\sqrt{\nu D_0}$ et $D_2 = D_0/\beta^2$. Cela implique que

$$\mathbb{E}\| \sum_{i=1}^{n_k} \epsilon_i \eta(Y_i) \|_{\mathcal{G}_{j,k}} \leq D_3 \sqrt{n_k h_{j,k} \log\left(\frac{1}{D_2 h_{j,k}} \vee \log\log n_k\right)} =: D_3 a_{j,k}.$$

Maintenant appliquons l'inégalité de Talagrand (1994) pour chaque classe $\mathcal{G}_{j,k}$. En choisissant pour tous $j, k \geq 0$,

$$M = \sup \eta \in \mathcal{G}_{j,k} \|\eta\|_\infty = \kappa$$

et

$$\sigma_{\mathcal{G}}^2 = \sigma_{\mathcal{G}_{j,k}}^2 \leq D_0 h_{j,k},$$

on a pour tout $t > 0$,

$$\mathbb{P}\left\{\max_{n_{k-1}\leq n\leq n_k} \|\sqrt{n}\alpha_n\|_{\mathcal{G}_{j,k}} \geq A_1(D_3 a_{j,k} + t)\right\} \leq 2\left[\exp\left(\frac{-A_2 t^2}{D_0 n_k h_{j,k}}\right) + \exp\left(\frac{-A_2 t}{\kappa}\right)\right].$$

Etant donné $\rho > 0$, $j \geq 0$ and $k \geq 1$, posons

$$p_{j,k}(\rho) = \mathbb{P}\left\{\max_{n_{k-1}\leq n\leq n_k} \|\sqrt{n}\alpha_n\|_{\mathcal{G}_{j,k}} \geq A_1(D_3 + \rho)a_{j,k}\right\},$$

avec $t = \rho a_{j,k}$. Alors

$$p_{j,k}(\rho) \leq 2\left[\exp\left(\frac{-A_2\rho^2 a_{j,k}^2}{D_0 n_k h_{j,k}}\right) + \exp\left(\frac{-A_2\rho a_{j,k}}{\kappa}\right)\right].$$

Par définition de $a_{j,k}$, nous avons

$$a_{j,k}^2/n_k h_{j,k} \geq \log\log n_k.$$

Pour k assez grand, cela implique que

$$\begin{aligned} p_{j,k}(\rho) &\leq 2\left[\exp\left(\frac{-A_2\rho^2}{D_0}\log\log n_k\right) + \exp\left(\frac{-A_2\rho}{\kappa}\sqrt{n_k h_{j,k}\log\log n_k}\right)\right] \\ &\leq 2\left[\exp\left(\frac{-A_2\rho^2}{D_0}\log\log n_k\right) + \exp\left(\frac{-A_2\rho\sqrt{c}}{\kappa}\sqrt{\log n_k \log\log n_k}\right)\right]. \end{aligned}$$

En posant $\lambda = \frac{A_2}{D_0} \wedge \frac{A_2\sqrt{c}}{\kappa}$, on obtient

$$p_{j,k}(\rho) \leq 4\exp(-\lambda\rho\log\log n_k) = 4(\log n_k)^{-\lambda\rho}. \tag{2.16}$$

Soit $l_k = \max\{j : h_{j,k} \leq 2h_0\}$. Alors pour tout k assez grand, on a,

$$l_k \leq 2\log n_k. \tag{2.17}$$

Donc compte tenu de (2.16) et (2.17), nous avons pour k assez grand et $\rho \geq 0$,

$$P_k(\rho) := \sum_{j=1}^{l_k-1} p_{j,k}(\rho) \leq 8(\log n_k)^{1-\lambda\rho} \simeq 8\left(\frac{1}{k\log 2}\right)^{\lambda\rho}.$$

En choisissant $\rho > 3/\lambda$, nous obtenons

$$\sum_{k=1}^{\infty} P_k(\rho) \leq \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty. \tag{2.18}$$

Par définition de l_k nous remarquons que

$$2h_{l_k,k} = h_{l_k+1,k} \geq 2h_0 \Rightarrow h_{l_k,k} \geq h_0.$$

Donc pour tout $n_{k-1} \leq n \leq n_k$,

$$\left[\frac{c\log n}{n}, h_0\right] \subset \left[\frac{c\log n_k}{n_k}, h_{l_k,k}\right].$$

On en déduit que, pour k assez grand et $n_{k-1} \leq n \leq n_k$,

$$A_k(\rho) := \left\{ \max_{n_{k-1}\leq n\leq n_k} \sup_{\frac{c\log n}{n}\leq h\leq h_0} \frac{\sup_{\varphi\in\Phi} ||W_{n,h}(\cdot,\varphi)||_{[0,z]}}{\sqrt{nh(|\log h| \vee \log\log n)}} > 2A_1(D_3+\rho) \right\}$$

$$\subset \bigcup_{j=0}^{l_k-1} \left\{ \max_{n_{k-1}\leq n\leq n_k} \|\sqrt{n}\alpha_n\|_{\mathcal{G}_{j,k}} \geq A_1(D_3+\rho)a_{j,k} \right\}.$$

Cela implique que pour ρ assez grand $\mathbb{P}(A_k(\rho)) \leq P_k(\rho)$. Ce qui achève la preuve en appliquant le lemme de Borel-Cantelli compte tenu de l'Inégalité (5.11).

2.4 Simulations

Dans cette section nous menons une étude de simulation pour évaluer la bonne performance de l'estimateur à noyau de pauvreté $J_{n,h}$ dans les cas particuliers des mesures de Sen, de Shorrocks et de Foster-Greer-Thorbecke (FGT) de paramétres $\alpha = 1, 2$. Nous générons des échantillons de loi lognormale de moyenne $m = 12$ et d'écart type $\sigma = 0.75$ et tailles respectives $n = 50, 100, 1000, 5000$. La densité de probabilité de la loi lognormale est donnée par

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{y}\exp\left\{-\frac{1}{2}\left(\frac{\log y - m}{\sigma}\right)^2\right\}, \quad y > 0.$$

L'indice théorique

$$J = \int_0^z w[G(y), G(z)]f(y,z)g(y)dy$$

est calculée avec la méthode d'intégration numérique de Simpson. Tandis que l'estimateur à noyau

$$J_{n,h} = \frac{1}{nh} \sum_{j=1}^{n} \int_0^z w[G_n(y), G_n(z)] f(y,z) K\left(\frac{Y_j - y}{h}\right) dy$$

est évalué avec le noyau de Barlett-Epanechnikov donné par

$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{si } |x| \leq 1 \\ 0 & \text{sinon.} \end{cases}$$

Pour un échantillon de taille $n = 100$, nous obtenons les résultats présentés dans la Table 2.1 ci-dessous, qui contient les erreurs quadratiques moyennes de 50 échantillons générés aléatoirement avec le logiciel R. Ces erreurs sont calculées pour plusieurs valeurs du seuil z, correspondant aux valeurs quantiles suivantes : $G^{-1}(p)$, $p = 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85$.

p Index	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85
FGT(1)	0.00028	0.00023	0.00056	0.00087	0.00068	0.00081	0.00085	0.0011
FGT(2)	0.00005	0.00014	0.00023	0.00031	0.0003	0.00046	0.00061	0.00065
SEN	0.00005	0.00011	0.00012	0.00050	0.00078	0.00284	0.00634	0.0195
SHO	0.00024	0.00043	0.00054	0.00145	0.00115	0.00144	0.0121	0.0137

TABLE 2.1 – Erreurs quadratiques moyennes de l'estimateur $J_{n,h}$ calculées sur 50 échantillons indépendants générés aléatoirement.

La Table 2.1 montre qu'il y a une bonne convergence de l'estimateur $J_{n,h}$ vers l'indice théorique J dans les cas des indices classiques tels que FGT(1), FGT(2), SEN, SHORROCKS (SHO). Comme le montrent les Figures 2.1 et 2.2 ci-dessous, ces résultats sont très satisfaisants avec des tailles relativement petites de l'ordre de 50 pour les indices FGT. Pour les mesures de SEN et de SHORROCKS des tailles d'échantillon plus importantes de l'ordre de 1000 sont considérées dans la Figure 2.2.

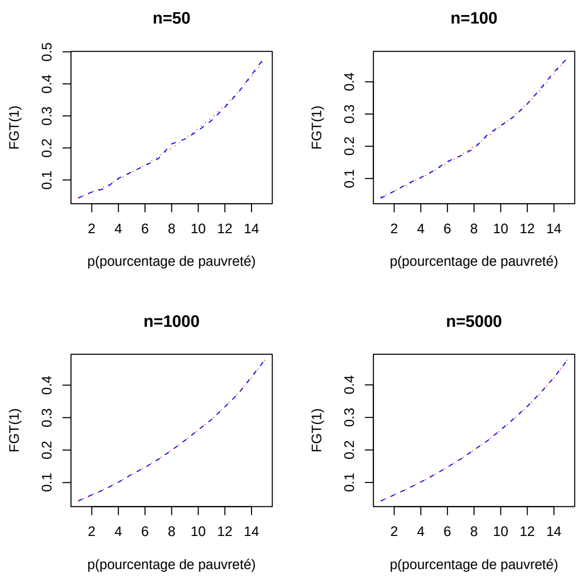

FIGURE 2.1 – *Courbes de l'estimateur $J_{n,h}(z)$ et de l'indice théorique $J(z)$ dans le cas FGT(1).*

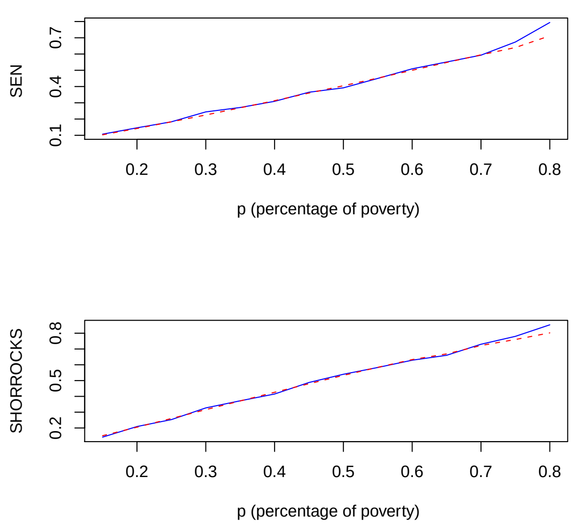

FIGURE 2.2 – *Courbes de l'estimateur $J_{n,h}(z)$ et de l'indice théorique $J(z)$ dans les cas SEN et SHORROCKS.*

Chapitre 3

Lois uniformes du logarithme pour l'estimateur à noyau de pauvreté

3.1 Introduction

Tandis que le chapitre 2 établissait la convergence uniforme presque sûre de l'estimateur à noyau de l'indice général de pauvreté

$$J = J(z, w, f) = \int_0^z w(G(y), G(z))f(y, z)dG(y), \tag{3.1}$$

le présent chapitre est consacré à l'étude de lois limites uniformes du logarithme pour J. On supposera que la fonction de répartition $G(y) = \mathbb{P}(Y \leq y), \forall y \in \mathbb{R}$ admet une densité de probabilité notée $g(y)$ par rapport à la mesure de Lebesgue. Alors l'indice J peut s'écrire sous la forme

$$J = J(z, w, f) = \int_0^z w(G(y), G(z))f(y, z)g(y)dy, \tag{3.2}$$

où $z > 0$, $w \in \mathcal{W}$ et $f \in \mathcal{F}$, $\mathcal{W}$ et $\mathcal{F}$ étant des classes de fonctions appropriées, i.e. continues et décroissantes par rapport à la variable y.

Stute (1982b) a été le premier à établir une loi uniforme du logarithme pour l'estimateur à noyau de la densité. En s'appuyant sur le comportement du module d'oscillation du processus empirique uniforme, il a montré, sous certaines conditions de régularité sur la densité $g(y)$ et sur la fenêtre h_n, que

$$\lim_{n\to\infty} \sqrt{nh_n} \sup_{y\in I} \left\{ \frac{|g_n(y) - \mathbb{E}g_n(y)|}{\sqrt{2\|K\|_2^2 g(y)\log(1/h_n)}} \right\} = 1 \quad p.s., \tag{3.3}$$

où g_n est définie pour tout $y \in \mathbb{R}$ par

$$g_n(y) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{Y_j - y}{h_n}\right), \tag{3.4}$$

avec $K(\cdot)$ une fonction noyau à support compact et à variation bornée telle que

$$0 < \int_{\mathbb{R}} K^2(t)dt = \|K\|_2^2 < \infty.$$

Deheuvels et Mason (1992) généralisent les résultats de Stute (1982a) sur le comportement du module d'oscillation du processus empirique uniforme et obtiennent une loi uniforme fonctionnelle du logarithme pour déterminer la vitesse exacte de convergence presque sûre des estimateurs à noyau de la densité $g(y)$. Mais ces résultats sont basés sur les approximations hongroises Komlós, Major et Tusnády (KMT, 1975) du processus empirique uniforme qui ne sont pas toujours valables sur le processus empirique général.

Une méthode plus générale et plus sophistiquée donnant une vitesse précise de convergence presque sûre des estimateurs de type noyau a été développée par Einmahl et Mason (2000). En effet, les résultats de ces derniers améliorent ceux de Härdle, Janssen et Serfling (1988), et retrouvent la loi uniforme du logarithme de Stute (1982b) donnée par (3.3) avec une autre approche basée sur la théorie du processus empirique indexé par des fonctions.

Dans ce chapitre nous allons utiliser cette dernière méthodologie, qui est aussi celle de Deheuvels et Mason (2004), pour établir des lois limites uniformes du logarithme en probabilité pour l'estimateur à noyau de pauvreté $J_n(z, w, f)$ défini par

$$J_n(z, w, f) = \frac{1}{nh_n} \sum_{j=1}^{n} \int_0^z w[G_n(y), G_n(z)] f(y, z) K\left(\frac{Y_j - y}{h_n}\right) dy, \tag{3.5}$$

où $G_n(y) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{Y_j \le y\}}$ est la fonction de répartition empirique d'un échantillon de variables aléatoires indépendantes $Y_1, \cdots, Y_n$ et 1_A désigne la fonction indicatrice d'un ensemble A.

Ici, nous supposons que le noyau $K(.)$ est une fonction mesurable vérifiant les conditions (K.1-2-3-4) suivantes :
- (K.1) $K(.)$ est à variation bornée sur $\mathbb{R}$.
- (K.2) $K(t) = 0$ si $|t| \geq a$, avec $a > 0$.
- (K.3) $\int_{\mathbb{R}} K(t)dt = 1$.
- (K.4) $\int_{\mathbb{R}} tK(t)dt = 0$ et $0 \neq \int_{\mathbb{R}} t^2 K(t)dt < \infty$.

Tandis que $(h_n)_{n\geq 1}$ est une suite de nombres réels positifs qui satisfait aux hypothèses (H.1-2) suivantes :
- (H.1) $h_n \to 0$, lorsque $n \to \infty$.
- (H.2) $nh_n / \log n \to \infty$, lorsque $n \to \infty$.
- (H.3) $\log(1/h_n)/ \log \log n \to \infty$ lorsque $n \to \infty$.

Nous aurons aussi besoin des hypothèses additionnelles sur la densité $g(\cdot)$ notamment que :
- (G.1) $g(\cdot)$ est 2-fois continûment dérivable.
- (G.2) $g''(\cdot)$ la dérivée seconde de la densité $g(\cdot)$ est bornée.

Pour obtenir une loi limite uniforme en z, nous allons fixer les fonctions w et f et garder la dépendance en z des quantités $J(z, w, f)$ et $J_n(z, w, f)$ de sorte que l'on notera

$$J(z) := J(z, w, f) \qquad \text{et} \qquad J_n(z) := J_n(z, w, f).$$

Afin de simplifier les notations, nous introduisons les fonctions $\theta_n(\cdot, z)$ et $\theta(\cdot, z)$ ci-dessous, définies pour tout $y \in [0, z]$, $z > 0$, par

$$\theta_n(y, z) = w[G_n(y), G_n(z)]f(y, z), \quad n \geq 1 \tag{3.6}$$

et

$$\theta(y, z) = w[G(y), G(z)]f(y, z). \tag{3.7}$$

Alors, du fait de la continuité des fonctions w et f, on peut observer en utilisant Glivenko-Cantelli, que $\theta_n(\cdot, z)$ converge presque sûrement et uniformément vers $\theta(\cdot, z)$, lorsque $n \to \infty$. D'où la convergence en probabilité de $\theta_n(\cdot, z)$ uniformément sur l'intervalle $[0, z]$, i.e. pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{y\in[0,z]} \left|\frac{\theta_n(y, z)}{\theta(y, z)} - 1\right| \geq \varepsilon\right) \to 0, \quad n \to \infty \qquad (\Theta.1).$$

Avec ces notations, l'estimateur $J_n(z)$ est réduit à

$$J_n(z) = \int_0^z \theta_n(y,z) g_n(y) dy, \tag{3.8}$$

et l'indice $J(z)$ s'écrit

$$J(z) = \int_0^z \theta(y,z) g(y) dy \tag{3.9}$$

pour tout $z > 0$.

Nous pouvons maintenant énoncer le principal résultat de ce chapitre qui est une loi faible du logarithme pour l'estimateur de pauvreté $J_n(z)$, uniformément en $z \in [0,b]$, pour tout $b > 0$.

3.2 Lois limites faibles du logarithme

Théorème 3.1. *Supposons que les hypothèses (K.1-2-3-4), (H.1-2) et (G.1-2) sont satisfaites et que la fonction $w : [0,1] \times [0,1] \to \mathbb{R}$ possède des dérivées partielles bornées, notées w'_u et w'_v. Si, de plus $h_n = Cn^{-\delta}$, où $C > 0$ et $1/5 \leq \delta < 1$, alors pour tout $b > 0$ fixé, on a lorsque $n \to \infty$,*

$$\left\{ \frac{nh_n}{2\log(1/h_n)} \right\}^{1/2} \sup_{z \in [0,b]} \pm \{J_n(z) - J(z)\} \xrightarrow{P} \sigma, \tag{3.10}$$

où

$$\sigma = \sup_{z \in [0,b]} \sigma(z) = \|K\|_2 \int_0^b \theta(y,z) g^{1/2}(y) dy,$$

avec $\sigma(z) := \|K\|_2 \sup_{z \in [0,b]} \int_0^z \theta(y,z) g^{1/2}(y) dy$ et $\|K\|_2^2 = \int_{\mathbb{R}} K^2(t) dt$.

Ici, "$\xrightarrow{P}$" désigne la convergence en probabilité.

Remarque 3.1. *En fixant z et en considérant $J_n(z,w,f) =: J_n(w,f)$ comme une fonctionnelle ne dépendant que de w et f, on obtient une loi limite uniforme fonctionnelle du logarithme pour l'estimateur $J_n(w,f)$ sur la classe de fonctions $\mathcal{W} \times \mathcal{F}$, i.e.*

$$\left\{ \frac{nh_n}{2\log(1/h_n)} \right\}^{1/2} \sup_{(w,f) \in \mathcal{W} \times \mathcal{F}} \pm \{J_n(w,f) - J(w,f)\} \xrightarrow{P} \sigma', \tag{3.11}$$

où

$$\begin{aligned}\sigma' &= \|K\|_2 \sup_{(w,f)\in\mathcal{W}\times\mathcal{F}} \int_0^z w[G(y),G(z)]f(y,z)g^{1/2}(y)dy \\ &= \|K\|_2 \sup_{(w,f)\in\mathcal{W}\times\mathcal{F}} \int_0^z \theta(y,z)g^{1/2}(y)dy,\end{aligned}$$

avec $\|K\|_2^2 = \int_{\mathbb{R}} K^2(t)dt$.

Preuve. Pour prouver ce théorème nous décomposons la différence $J_n(z) - J(z)$ en trois termes.

$$\begin{aligned}J_n(z) - J(z) &= \int_0^z \theta_n(y,z)g_n(y)dy - \int_0^z \theta(y,z)g(y)dy \\ &= \int_0^z \theta_n(y,z)[g_n(y) - \mathbb{E}g_n(y)]dy + \int_0^z [\theta_n(y,z) - \theta(y,z)]\mathbb{E}g_n(y)dy \\ &\quad + \int_0^z \theta(y,z)[\mathbb{E}g_n(y) - g(y)]dy \\ &\quad =: T_n(z) + R_{n,1}(z) + R_{n,2}(z).\end{aligned}$$

Nous allons d'abord établir que les termes $R_{n,1}(z)$ et $R_{n,2}(z)$ convergent uniformément et presque sûrement vers 0, avec une vitesse de $O(\sqrt{nh_n/2\log(1/h_n)})$. En appliquant le théorème de la moyenne, on obtient compte tenu de (3.6) et (3.7) que,

$$\begin{aligned}|R_{n,1}(z)| &\leq \int_0^z |\theta_n(y,z) - \theta(y,z)||\mathbb{E}g_n(y)|dy \\ &\leq |w'_u(\eta, G(z))||G_n(y) - G(y)||f(y,z)||\mathbb{E}g_n(y)|, \quad \eta \in [G(y), G_n(y)] \\ &\leq \sup_{0\leq t\leq 1} |w'_u(t, G(z))| \sup_{0\leq y\leq z} |f(y,z)| \sup_{0\leq y\leq z} |\mathbb{E}g_n(y)| \times \|G_n - G\|_\infty.\end{aligned}$$

On a $\sup_{0\leq y\leq z} |\mathbb{E}g_n(y)|$ borné car

$$\sup_{0\leq y\leq z} |\mathbb{E}g_n(y) - g(y)| \to 0, \quad n \to \infty.$$

En effet, on a par changement de variable,

$$\begin{aligned}\sup_{0\leq y\leq z} |\mathbb{E}g_n(y) - g(y)| &\leq \sup_{0\leq y\leq z} \int_{-a}^a |g(y - th_n) - g(y)|K(t)dt \\ &\leq \sup_{0\leq y\leq z} \sup_{|t|\leq a} |g(y - th_n) - g(y)| \int_{\mathbb{R}} K(t)dt \\ &\leq \sup_{0\leq x,y\leq z, |x-y|\leq ah_n} |g(y - th_n) - g(y)| \to 0, \ n \to \infty,\end{aligned}$$

puisque la densité $g(\cdot)$ est supposée continue sur $\mathbb{R}$, donc uniformément continue sur l'intervalle compact $[0, z]$.

De même $\sup_{0\le y\le z}|f(y,z)|$ est borné car la fonction $y \mapsto f(y,z)$ est continue sur $[0, z]$, pour tout $z > 0$ donné. On en déduit, compte tenu également du fait que w'_u est bornée, que

$$|R_{n,1}(z)| \le C_1\|G_n - G\|_\infty, \quad \forall z > 0, \tag{3.12}$$

où C_1 est une constante réelle positive.

G étant continue, en utilisant la loi du logarithme itéré de Chung (1948), l'inégalité (3.12) implique que

$$\sup_{z\in[0,b]}|R_{n,1}(z)| = O(\sqrt{n^{-1}\log\log n/2}).$$

De plus, on peut déduire de l'hypothèse (H.3) que

$$\frac{nh_n}{\log(1/h_n)} \le \frac{n}{\log\log n}.$$

D'où

$$\sqrt{\frac{nh_n}{2\log(1/h_n)}}\sup_{z\in[0,b]}|R_{n,1}(z)| \to 0, \quad n\to\infty. \tag{3.13}$$

Pour le terme $R_{n,2}(z)$ nous utilisons un développement de Taylor. On a pour tout $y \in [0, z]$

$$\begin{aligned}
\mathbb{E}g_n(y) - g(y) &= \int_{\mathbb{R}} K(u)(g(y-uh_n) - g(y))du \\
&= \int_{\mathbb{R}} K(u)[-uh_n g'(y) + \frac{u^2h_n^2}{2}g''(y-\lambda uh_n)]du, \ \ 0<\lambda<1 \\
&= \frac{h_n^2}{2}\int_{\mathbb{R}} u^2 g''(y-\theta uh_n)K(u)du.
\end{aligned}$$

Cela entraîne que

$$\sup_{y\in[0,z]}|\mathbb{E}g_n(y) - g(y)| \le \frac{h_n^2}{2}\sup_{x\in\mathbb{R}}|g''(x)|\int_{\mathbb{R}}\left|u^2K(u)\right|du \to 0, \quad n\to\infty. \tag{3.14}$$

Donc, puisque la fonction $y \mapsto \theta(y,z)$ est bornée, on en déduit que pour tout $z > 0$

$$\begin{aligned}
|R_{n,2}(z)| &\le \int_0^z |\theta(y,z)||\mathbb{E}g_n(y) - g(y)|dy \\
&\le z\times\sup_{y\in[0,z]}|\mathbb{E}g_n(y) - g(y)|\sup_{y\in[0,z]}|\theta(y,z)| \\
&\le C_2h_n^2,
\end{aligned}$$

où C_2 est une constante réelle positive. En utilisant l'hypothèse $h_n = Cn^{-\delta}, 1/5 \leq \delta < 1$, on obtient, après des calculs de routine, que

$$\sqrt{\frac{nh_n}{2\log(1/h_n)}} \sup_{z\in[0,b]} |R_{n,2}(z)| \to 0, \quad n \to \infty. \tag{3.15}$$

Quant au terme $T_n(z)$, il converge uniformément en probabilité sur l'intervalle $[0, b]$, vers une constante que nous allons déterminer. En combinant la condition (Θ.1) établie ci-dessus, avec les hypothèses (K.1-2-3), (H.1-2) et (G.1), alors on obtient, d'après le théorème 1.2 de Deheuvels et Mason (2004), que pour tout $y \in [0, z]$

$$\sqrt{\frac{nh_n}{2\log(1/h_n)}} \pm \theta_n(y,z)\{g_n(y) - \mathbb{E}g_n(y)\} \xrightarrow{P} \|K\|_2\theta(y,z)g^{1/2}(y),$$

donc

$$\int_0^z \sqrt{\frac{nh_n}{2\log(1/h_n)}} \pm \theta_n(y,z)\{g_n(y) - \mathbb{E}g_n(y)\}dy \xrightarrow{P} \|K\|_2 \int_0^z \theta(y,z)g^{1/2}(y)dy := \sigma(z).$$

D'où

$$\sup_{z\in[0,b]} \int_0^z \sqrt{\frac{nh_n}{2\log(1/h_n)}} \pm \theta_n(y,z)\{g_n(y) - \mathbb{E}g_n(y)\}dy \xrightarrow{P} \sup_{z\in[0,b]} \sigma(z) := \sigma,$$

c'est à dire

$$\sup_{z\in[0,b]} T_n(z) \xrightarrow{P} \sigma = \|K\|_2 \int_0^b \theta(y,z)g^{1/2}(y)dy. \tag{3.16}$$

Les relations (3.13), (3.15) et (3.16) prouvent entièrement le théorème.

3.3 Intervalles de confiance simultanés

Le théorème 3.1 précédent nous permet de construire des intervalles de confiance asymptotiques et uniformes en $z \in [0, b]$, pour l'indice de pauvreté théorique $J(z)$, avec un niveau de confiance asymptotiquement optimal (100%). Dans la littérature statistique, on utilise le plus souvent la normalité asymptotique et les lois qui en découlent pour construire des intervalles de confiance. Mais, il n'est pas très habituel de construire des intervalles sur la base d'une loi dégénérée. Dans cette section,

nous allons remarquer que nos lois limites uniformes du logarihme constituent une extension des lois asymptotiques normales pour construire de tels intervalles de confiance, via un mode de convergence plus fort (en probabilité).
Posons

$$L_n(z) = \sigma(z)\sqrt{\frac{2\log(1/h_n)}{nh_n}},$$

alors, nous pouvons déduire de (3.10) que pour tout $b > 0$,

$$\pm \sup_{z\in[0,b]} \frac{1}{L_n(z)}\{J_n(z) - J(z)\} \xrightarrow{P} 1, \qquad n \to \infty.$$

Donc, pour tout $\epsilon > 0$, on a

$$\mathbb{P}\left(\left|\pm\frac{1}{L_n(z)}\{J_n(z) - J(z)\} - 1\right| > \epsilon, \forall z \in [0, b]\right) \to 0,$$

ce qui équivaut à

$$\mathbb{P}\left(|\pm\{J_n(z) - J(z)\} - L_n(z)| \leq \epsilon L_n(z), \forall z \in [0, b]\right) \to 1.$$

Cela implique pour tout $\epsilon > 0$ que, lorsque $n \to \infty$,

$$\mathbb{P}(J(z) \in [J_n(z) + (1-\epsilon)L_n(z), J_n(z) + (1+\epsilon)L_n(z)], \forall z \in [0, b]) \to 1, \tag{3.17}$$

et

$$\mathbb{P}(J(z) \in [J_n(z) - (1+\epsilon)L_n(z), J_n(z) - (1-\epsilon)L_n(z)], \forall z \in [0, b]) \to 1. \tag{3.18}$$

En combinant (3.17) et (3.18), on peut écrire que pour tout $\epsilon > 0$, lorsque $n \to \infty$,

$$\mathbb{P}(J(z) \in [J_n(z) - (1+\epsilon)L_n(z), J_n(z) + (1+\epsilon)L_n(z)], \forall z \in [0, b]) \to 1. \tag{3.19}$$

Cette dernière relation (3.19) nous permet de construire des intervalles de confiance simultanés en $z \in [0, b]$ et asymptotiquement optimaux pour l'indice de pauvreté $J(z)$, de la forme

$$[J_n(z) - L_n(z), J_n(z) + L_n(z)]. \tag{3.20}$$

3.3.1 Application : Sélection d'un modèle de loi paramétrique

Du fait de leur optimalité asymptotique, nous allons utiliser les intervalles de confiance établis en (3.20), pour sélectionner un modèle de loi paramétrique adéquat afin d'analyser la pauvreté avec les données de revenu. Pour ce faire, nous considérons trois modèles paramétriques : la loi Lognormale $LN(m,\sigma)$, la loi Gamma $\gamma(a,b)$ et la loi de Singh-Maddala $SM(\alpha,\beta,q)$. Rappelons que la loi de Singh-Maddala est définie par la fonction de répartition

$$G(y) = 1 - \frac{1}{\left[1+\left(\frac{y}{\beta}\right)^{\alpha}\right]^{q}},$$

où α et q sont des paramétres de forme et β un paramétre d'échelle. Elle généralise les distributions de Paréto et de Weibull.

Les paramètres de ces lois sont estimés par la méthode du maximum de vraisemblance en utilisant des données de revenu annuel des ménages sénégalais. Ces données ont été recueillies auprès de 3278 ménages choisis au hasard, lors d'une enquête[1] d'envergure nationale, réalisée par l'Agence Nationale de la Statistique et de la Démographie en 1994. La variable revenu Y qui est étudiée ici, a été standardisée suivant l'échelle d'équivalence-adulte de la FAO. Le seuil de pauvreté annuel z était fixé à 143080 F CFA pour chaque ménage.

Afin de tester l'adéquation de ces trois modèles paramétriques pour estimer les indices de pauvreté, nous allons procéder à deux méthodes d'estimation différentes, pour les indices de Foster-Greer-Thorbecke définis par

$$FGT(\alpha) = \int_0^z \left(\frac{z-y}{z}\right)^{\alpha} dG(y), \quad \alpha \geq 0.$$

D'abord, nous calulons les intervalles de confiance construits autour de l'estimateur à noyau et donnés par (3.20), pour les indices $FGT(0)$, $FGT(1)$ et $FGT(2)$ en considérant le noyau d'Epanechnikov donné par

$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{si } |x| \leq 1 \\ 0 & \text{sinon.} \end{cases}$$

Ensuite, nous estimons ces mêmes indices par une méthode d'intégration numérique (Simpson), en supposant que la loi de la variable revenu Y suit respectivement les distributions Lognormale, Gamma et Singh-Maddala. La Table 3.1 donne les résultats

1. Enquête Sénégalaise Auprès des Ménages

de l'estimation non-paramétrique pour les indices FGT(0), FGT(1) et FGT(2) communément appelés le pourcentage de la pauvreté ($\alpha = 0$), la profondeur de la pauvreté ($\alpha = 1$) et la sévérité de la pauvreté ($\alpha = 2$).

Indice	Estimation	Borne Inf	Borne Sup
FGT(0)	0.316	0.295	0.337
FGT(1)	0.101	0.081	0.122
FGT(2)	0.048	0.027	0.068

TABLE 3.1 – Estimation non-paramétrique de l'indice FGT pour $\alpha = 0, 1, 2$.

La Table 3.2 ci-dessous donne les estimations ponctuelles de ces même indices pour chacune des trois distributions considérées.

Distribution	FGT(0)	FGT(1)	FGT(2)
Lognormal	0.311	0.106	0.050
Gamma	0.346	0.157	0.096
Singh-Maddala	0.315	0.092	0.039

TABLE 3.2 – Estimation paramétrique de l'indice FGT pour $\alpha = 0, 1, 2$.

Sur ces deux tableaux, nous remarquons que les estimations de l'indice FGT obtenues avec les modèles Lognormal et Singh-Maddala sont plus consistantes avec l'estimation non-paramétrique. De plus, elles sont comprises entre les bornes des intervalles de confiance asymptotiquement optimaux donnés par la méthode non-paramétrique, alors que celles obtenues avec la distribution Gamma ne le sont pas. Pour mieux visualiser ces remarques, nous représentons dans la Figure 3.1 ci-dessous les intervalles de confiance simultanés non-paramétriques et les courbes d'évolution de l'indice FGT(1), par rapport au seuil de pauvreté z, pour chacun des trois modèles considérés ainsi que l'indice FGT classique.

La Figure 3.1 montre que les modèles Lognormal et Singh-Maddala semblent adéquats pour estimer l'indice de pauvreté de FGT, car leurs courbes d'évolution sont entièrement dans les limites de confiance optimales établies avec la méthode non-paramétrique.

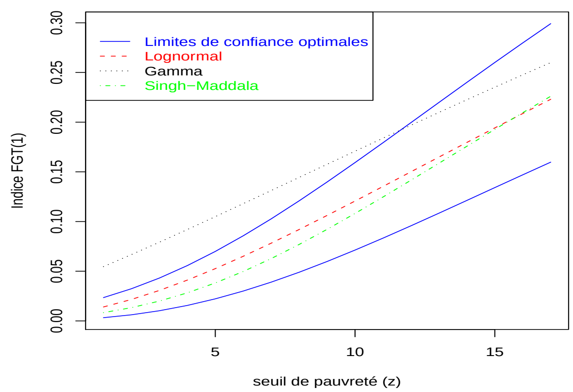

FIGURE 3.1 – *Comparaison des trois modèles paramétriques.*

Par conséquent, nous pouvons dire que, pour ces données de revenu de ménages sénégalais, l'estimation paramétrique de l'indice de pauvreté conduirait à choisir le modèle de Singh-Maddala ou le modèle Lognormal. Ce résultat est en phase avec les travaux de McDonald (1984) qui disent que dans beaucoup de situations, la distribution de Singh-Maddala apparaît comme le meilleur modèle pour ajuster les données de revenus.

Chapitre 4

Normalité asymptotique des estimateurs de type plug-in

4.1 Introduction

Dans ce chapitre le seuil de pauvreté $z > 0$ et les fonctions w et f seront tous fixés ; et on note J l'indice général de pauvreté défini par

$$J = \int_0^z w(G(y), G(z))f(y, z)dG(y), \tag{4.1}$$

où, rappelons-le, $G(y) = \mathbb{P}(Y \leq y)$ est la fonction de répartition de la variable aléatoire Y, représentant le revenu d'un individu pris au hasard dans la population étudiée.

Supposons que $Y_1, Y_2, \cdots, Y_n$ est un échantillon de variables aléatoires indépendantes et identiquement distribuées de même loi que la variable Y. Notons G_n la fonction de répartition empirique correspondante. Pour tout réel y, on a

$$G_n(y) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{Y_j \leq y\}},$$

où 1_A désigne la fonction indicatrice de l'ensemble A. Introduisons les statistiques d'ordre correspondant à l'échantillon $Y_1, Y_2, \cdots, Y_n$, notées

$$0 \leq Y_{1,n} \leq Y_{2,n} \leq \cdots \leq Y_{n,n} \leq \infty.$$

Alors si $q = q_n$ désigne le nombre de pauvres (i.e. $card\{j : j \leq n, Y_j < z\}$) dans l'échantillon de taille n, on a

$$\frac{q}{n} = G_n(z). \tag{4.2}$$

Au lieu des estimateurs à noyau, ici nous allons travailler avec des estimateurs directs de type "plug-in", i.e.

$$\begin{aligned} J_n &= \frac{1}{n}\sum_{j=1}^{q} w(G_n(Y_{j,n}), G_n(z))f(Y_{j,n}, z) &(4.3)\\ &= \frac{1}{n}\sum_{j=1}^{q} w\left(\frac{j}{n}, \frac{q}{n}\right) f(Y_{j,n}, z), &(4.4)\end{aligned}$$

où w et f sont des fonctions mesurables appropriées. Cet indice empirique J_n décrit un ensemble très vaste de mesures de pauvreté discrètes que l'on peut subdiviser en deux classes. La première classe contient les mesures dites non-pondérées pour lesquelles la fonction poids $w(\frac{j}{n}, \frac{q}{n})$ est égale à la constante 1. Cette classe contient toutes les mesures de pauvreté décomposables, en particulier celle de Foster-Greer-Thorbecke (1984). Celle-ci est définie, pour tout $\alpha \geq 0$, par

$$FGT_n(\alpha) = \frac{1}{n}\sum_{j=1}^{q}\left(\frac{z - Y_{j,n}}{z}\right)^{\alpha}. \tag{4.5}$$

La mesure $FGT(z, \alpha)$ est fréquemment utilisée dans les études empiriques de pauvreté. Pour $\alpha = 0$, elle est réduite à q/n, le pourcentage empirique d'individus pauvres. Lorsque $\alpha = 1$ et $\alpha = 2$, elle est respectivement interprétée comme l'intensité de la pauvreté et la sévérité de la pauvreté.

En dehors de cette classe, il y a la classe des mesures dites pondérées pour lesquelles le poids $w(\frac{j}{n}, \frac{q}{n})$ varie selon le rang de l'individu j dans l'échelle des personnes pauvres. Cette seconde classe contient l'ensemble des indices de pauvreté non-décomposables ou linéaires par rapport au revenu. On y trouve la fameuse mesure de Sen (1976) qui peut s'écrire sous la forme

$$S_n = \frac{1}{n}\sum_{j=1}^{q}\frac{2q}{q+1}\left(1 - \frac{G_n(Y_{j,n})}{G_n(z)} + \frac{1}{q}\right)\left(\frac{z - Y_{j,n}}{z}\right). \tag{4.6}$$

La mesure de Kakwani (1980) qui est une généralisation de (4.6) appartient aussi à cette seconde classe. Ell est définie pour $k \geq 0$ par

$$K_n(k) = \frac{1}{n} \sum_{j=1}^{q} \frac{q^{k+1}}{\sum_{j=1}^{q} j^k} \left(1 - \frac{G_n(Y_{j,n})}{G_n(z)} + \frac{1}{q}\right)^k \left(\frac{z - Y_{j,n}}{z}\right). \tag{4.7}$$

De même, Shorrocks (1995) propose une mesure pondérée définie par

$$Sh_n = \frac{1}{n} \sum_{j=1}^{q} 2 \left(1 - G_n(Y_{j,n}) + \frac{1}{2n}\right) \left(\frac{z - Y_{j,n}}{z}\right). \tag{4.8}$$

Pour plus de détails concernant ces deux classes de mesures de pauvreté, nous renvoyons le lecteur à Zheng (1997). Notre objectif dans ce chapitre, consiste plutôt à établir la normalité asymptotique de l'estimateur général de pauvreté J_n. Nous montrerons également que la variance asymptotique peut être estimée de façon non-paramétrique afin de construire des intervalles de confiance précis.

4.2 Normalité asymptotique

Soit un réel $z > 0$ fixé ; w et f deux fonctions appropriées définissant une mesure de pauvreté quelconque donnée par

$$J = \int_0^z w(G(y), G(z)) f(y, z) dG(y). \tag{4.9}$$

Nous allons d'abord montrer que l'estimateur J_n converge presque sûrement vers la quantité J.

Soit $(h_n)_{n \geq 1}$ une suite de fonctions réelles définies sur $\mathbb{R}^+$ par

$$h_n(y) = w(G_n(y), G_n(z)) f(y, z) 1_{\{y<z\}}, \quad \forall n \geq 1. \tag{4.10}$$

et h la fonction réelle définie par

$$h(y) = w(G(y), G(z)) f(y, z) 1_{\{y<z\}}, \ \ \forall y \geq 0. \tag{4.11}$$

En utilisant la loi forte des grands nombres et la continuité de la fonction w, on montre aisément que, pour tout $y \geq 0$, $h_n(y)$ converge presque sûrement vers $h(y)$, lorsque $n \to \infty$.

Introduisons quelques notations. Soit P la loi de probabilité de Y, i.e. $P =$

$\mathbb{P} \circ Y$, $\mathbb{P}_n$ la mesure empirique basée sur l'échantillon $Y_1, \cdots, Y_n$ et $\mathbb{G}_n$ le processus empirique associé. Alors pour toute fonction g mesurable, on a

$$Pg = \int g dP = \mathbb{E}g(Y), \qquad \mathbb{P}_n(g) = \frac{1}{n}\sum_{j=1}^{n} g(Y_j), \qquad \mathbb{G}_n(g) = \sqrt{n}(\mathbb{P}_n - P)g.$$

Avec ces notations, nous pouvons ré-écrire J_n en fonction de la mesure empirique $\mathbb{P}_n$,

$$J_n = \frac{1}{n}\sum_{j=1}^{n} w(G_n(Y_{j,n}), G_n(z)) f(Y_{j,n}, z) 1_{\{Y_{j,n} \leq z\}} = \mathbb{P}_n(h_n)$$

et la quantité J en fonction de la loi P,

$$J = \int_0^z w(G(y), G(z)) f(y,z) dG(y) = \int_0^\infty h(y) dG(y) = \int h dP = Ph.$$

Donc

$$J_n - J = \mathbb{P}_n(h_n) - Ph = [\mathbb{P}_n(h_n - h)] + [\mathbb{P}_n h - Ph]. \tag{4.12}$$

Puisque $Ph = J < \infty$, en appliquant de nouveau la loi des grands nombres pour les deux derniers termes de l'égalité (4.12), on obtient finalement que

$$J_n \xrightarrow{p.s.} J, \quad n \to \infty.$$

Une condition nécessaire pour appliquer le théorème central limite est que $\mathbb{E}h^2(Y) < \infty$. Cette condition est vérifiée ici , puisque

$$\mathbb{E}h^2(Y) = Ph^2 = \int_0^{+\infty} h(y) dG(y) = \int_0^z w^2(G(y), G(z)) f^2(y,z) dG(y) < \infty, \tag{4.13}$$

car les fonctions w et f sont supposées continues par hypothèse donc intégrables sur l'intervalle compact $[0, z]$ pour tout $z > 0$ fixé.

Maintenant nous allons décrire la normalité asymptotique de l'estimateur de type "plug-in" J_n dans le Théorème 4.1 suivant, où $\rightsquigarrow$ désigne la convergence faible.

Théorème 4.1. *Soit z un réel positif fixé ; $w : [0,1] \times [0,1] \to \mathbb{R}_+$ et $f : \mathbb{R}^2 \to \mathbb{R}_+$ deux fonctions continues et décroissantes chacune par rapport à sa première variable. Si de plus, la fonction w admet des dérivées partielles premières w'_u et w'_v bornées, alors lorsque $n \to \infty$, on a*

$$\sqrt{n}[J_n - J] \rightsquigarrow N(0, \sigma^2),$$

où $N(0,\sigma^2)$ *est la loi d'une variable aléatoire normale de variance*

$$\sigma^2 = \sigma_1^2 + \sigma_2^2,$$

avec

$$\sigma_1^2 = \int_0^z w^2(G(y),G(z))f^2(y,z)dG(y) - \left(\int_0^z w(G(y),G(z))f(y,z)dG(y)\right)^2$$

et

$$\begin{aligned}\sigma_2^2 &= \int_0^z\int_0^z \varphi(x,z)\varphi(y,z)[G(x\wedge y) - G(x)G(y)]dG(x)dG(y)\\ &\quad + G(z)(1-G(z))\left(\int_0^z \phi(y,z)dG(y)\right)^2\\ &\quad + 2(1-G(z))\left(\int_0^z \varphi(y,z)G(y)dG(y)\right).\left(\int_0^z \phi(y,z)dG(y)\right).\end{aligned}$$

où $\varphi(y,z) = w'_u(G(y),G(z))f(y,z)$ *et* $\phi(y,z) = w'_v(G(y),G(z))f(y,z)$.

Remarque 4.1. *Avec des choix appropriés pour les fonctions* $w(\cdot,\cdot)$ *et* $f(\cdot,\cdot)$*, ce théorème nous permet d'obtenir la normalité asymptotique des mesures empiriques de Sen, de Shorrocks, de Kakwani, de Foster-Greer-Thorbecke et de beaucoup d'autres mesures existant dans la littérature.*

Preuve. Par décomposition, on obtient

$$\begin{aligned}\sqrt{n}[J_n - J] &= \sqrt{n}[\mathbb{P}_n h_n - Ph]\\ &= \sqrt{n}[\mathbb{P}_n(h_n - h) - P(h_n - h)] + \sqrt{n}P(h_n - h) + \sqrt{n}(\mathbb{P}_n - P)h\\ &= \mathbb{G}_n(h_n - h) + \mathbb{G}_n(h) + \sqrt{n}P(h_n - h),\end{aligned}$$

où h_n et h sont les fonctions définies précédemment en (4.10) et (4.11).

Pour prouver le théorème 4.1, nous allons maintenant établir que le premier terme de cette décomposition $\mathbb{G}_n(h_n - h)$ tend vers 0 presque sûrement, alors que les deuxième et troisième termes $\mathbb{G}_n(h)$ et $\sqrt{n}P(h_n - h)$ convergent chacun, en distribution, vers une variable gaussienne, lorsque $n \to \infty$.
Pour toute fonction réelle φ définie dans un domaine D, on note

$$\|\varphi\|_D = \sup_{y\in D}|\varphi(y)| \quad \text{et} \quad \|\varphi\|_\infty = \|\varphi\|_{\mathbb{R}}.$$

Comme précédemment, une application de la loi des grands nombres implique que $h_n(y)$ converge presque sûrement vers $h(y)$, pour tout $y \geq 0$. D'après le théorème de Glivenko-Cantelli, cette convergence est uniforme sur $\mathbb{R}$. Donc, pour tout $\delta > 0$ et n assez grand, on a

$$\|h_n - h\|_\infty \leq \frac{\delta}{\sqrt{n}}, \quad \delta > 0.$$

D'où,

$$|\mathbb{G}_n(h_n - h)| \leq \mathbb{G}_n\|h_n - h\|_\infty \leq \mathbb{G}_n(\frac{\delta}{\sqrt{n}}) = (\mathbb{P}_n - P)(\delta).$$

Or $(\mathbb{P}_n - P)(\delta) \to 0$, d'après la loi des grands nombres. Donc le premier terme $\mathbb{G}_n(h_n - h)$ converge presque sûrement vers 0, lorsque $n \to \infty$.

Pour le second terme, nous appliquons directement le théorème central limite puisque $Ph^2 = \mathbb{E}h^2(Y) < \infty$ d'après (4.13). On obtient

$$\mathbb{G}_n(h) \rightsquigarrow \mathbb{G}(h),$$

où $\mathbb{G}(h)$ est une variable aléatoire gaussienne de variance,

$$\sigma_1^2 = Ph^2 - (Ph)^2 = \int_0^z w^2(G(y), G(z)) f^2(y,z) dG(y) - \left(\int_0^z w(G(y), G(z)) f(y,z) dG(y)\right)^2.$$

Pour le troisième terme, nous appliquons un développement de Taylor à la fonction bi-variable w. On peut alors écrire, avec $B(\cdot)$ désignant un pont brownien standard,

$$\begin{aligned}
\sqrt{n}\,\{w(G_n(y), G_n(z)) - w(G(y), G(z))\} &= w'_u(\zeta)\sqrt{n}(G_n(y) - G(y)) + w'_v(\zeta)\sqrt{n}(G_n(z) - G(z)) \\
&= w'_u(\zeta)\{\sqrt{n}(G_n(y) - G(y)) - B(G(y))\} \\
&\quad + w'_v(\zeta)\{\sqrt{n}(G_n(z) - G(z)) - B(G(z))\} \\
&\quad + w'_u(\zeta)B(G(y)) + w'_v(\zeta)B(G(z)) \\
&\quad + o(|G_n(y) - G(y)|) + o(|G_n(z) - G(z)|),
\end{aligned}$$

où $\zeta = \zeta(y,z) = (G(y), G(z))$. Posons

$$\varphi(y,z) = w'_u(\zeta) f(y,z) \text{ et } \phi(y,z) = w'_v(\zeta) f(y,z).$$

Alors, rappelant les définitions de h_n et h en (4.10) et (4.11) on obtient, en rajoutant et en retranchant une intégrale du pont brownien, que

$$\begin{aligned}
\sqrt{n}P(h_n - h) &= \int_0^z \varphi(y,z)\{\sqrt{n}(G_n(y) - G(y)) - B(G(y))\}dG(y) \\
&\quad + \int_0^z \phi(y,z)\{\sqrt{n}(G_n(z) - G(z)) - B(G(z))\}dG(y) \\
&\quad + \int_0^z \varphi(y,z)B(G(y))dG(y) \\
&\quad + \int_0^z \phi(y,z)B(G(z)dG(y) + o(1) \\
&=: R_{n,1} + R_{n,2} + N_1 + N_2 + o(1).
\end{aligned}$$

Les termes $R_{n,1}$ et $R_{n,2}$ convergent presque sûrement vers 0, lorsque $n \to \infty$. En effet, nous avons

$$\begin{aligned}
|R_{n,1}| &= \int_0^z \varphi(y,z)\{\sqrt{n}(G_n(y) - G(y)) - B(G(y))\}dG(y) \\
&\leq (G(z) - G(0)) \sup_{y\in[0,z]} |\varphi(y,z)| \times \|\sqrt{n}(G_n - G) - B \circ G\|_\infty
\end{aligned}$$

et

$$\begin{aligned}
|R_{n,2}| &= \int_0^z \phi(y,z)\{\sqrt{n}(G_n(z) - G(z)) - B(G(z))\}dG(y) \\
&\leq (G(z) - G(0)) \sup_{y\in[0,z]} |\phi(y,z)| \times \|\sqrt{n}(G_n - G) - B \circ G\|_\infty,
\end{aligned}$$

où $''\circ''$ désigne l'opérateur de composition de fonctions.

D'après les approximations hongroises (KMT, 1975), on peut choisir le pont brownien $B(\cdot)$ de sorte que $\|\sqrt{n}(G_n - G) - B \circ G\|_\infty$ converge presque sûrement vers 0, lorsque $n \to \infty$. Alors en combinant la continuité de f avec les dérivées partielles bornées de w, nous en déduisons que les termes $R_{n,1}$ et $R_{n,2}$ tendent vers 0 presque sûrement, quand $n \to \infty$.

Nous remarquons enfin que les deux autres termes N_1 et N_2 sont des intégrales du pont brownien $B(\cdot)$, donc des variables aléatoires gaussiennes car, ils peuvent s'écrire comme des transformations linéaires d'un vecteur gaussien. La somme $N_1 + N_2$ est aussi une intégrale du pont brownien $B(\cdot)$, donc c'est une variable aléatoire gaussienne. Ainsi, lorsque $n \to \infty$, le terme $\sqrt{n}P(h_n - h)$ converge en distribution vers une variable aléatoire gaussienne de variance

$$\begin{aligned}\sigma_2^2 &= \int_0^z \int_0^z \varphi(x,z)\varphi(y,z)[G(x\wedge y)-G(x)G(y)]dG(x)dG(y)\\ &+G(z)(1-G(z))\left(\int_0^z \phi(y,z)dG(y)\right)^2\\ &+2(1-G(z))\left(\int_0^z \varphi(y,z)G(y)dG(y)\right)\left(\int_0^z \phi(y,z)dG(y)\right).\end{aligned}$$

Pour avoir la variance asymptotique σ^2 de $\sqrt{n}[J_n - J]$, on peut remarquer que la covariance des termes $\mathbb{G}_n(h)$ et $\sqrt{n}P(h_n - h)$ est nulle pour toute fonction h mesurable. En effet, on a

$$\begin{aligned}\mathrm{cov}(\mathbb{G}_n(h), \sqrt{n}P(h_n-h)) &= \mathrm{cov}(\sqrt{n}(\mathbb{P}_n h - Ph), \sqrt{n}(Ph_n - Ph))\\ &= n\mathrm{cov}(\mathbb{P}_n h - Ph, Ph_n - Ph)\\ &= n[\mathrm{cov}(\mathbb{P}_n h, Ph_n) - \mathrm{cov}(\mathbb{P}_n h, Ph) - \mathrm{cov}(Ph, Ph_n) + \mathrm{cov}(Ph, Ph)]\\ &= n\mathrm{cov}(\mathbb{P}_n h, Ph_n) \quad \text{car } \mathrm{cov}(X,a)=0 \quad \text{si } a \text{ est une constante}\\ &= n\mathrm{cov}(\frac{1}{n}\sum_{j=1}^n h(Y_j), \mathbb{E}h_n(Y))\\ &= \sum_{j=1}^n \mathrm{cov}(h(Y_j), \mathbb{E}h_n(Y)) = 0.\end{aligned}$$

Finalement $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

Remarque 4.2. *Lorsque la variance σ^2 est estimée, le théorème 4.1 nous permet de calculer des intervalles de confiance asymptotiques pour l'indice de pauvreté J. Par exemple, pour construire un intervalle de confiance de niveau $100(1-\alpha)\%$, il suffit de trouver un estimateur consistant s_n^2 de σ^2 et d'appliquer la formule*

$$J_n \pm \frac{s_n}{\sqrt{n}} u_{1-\alpha/2}, \tag{4.14}$$

où $u_{1-\alpha/2}$ $(0<\alpha<1)$ désigne le quantile d'ordre $(1-\alpha/2)$ de la loi normale centrée réduite.

4.3 Simulations

Pour appliquer ce résultat de normalité asymptotique dans l'analyse de la pauvreté, il est nécessaire de savoir dans quelles conditions il est vrai. Pour cela, nous allons procéder à des simulations sur des échantillons de données de tailles $n = 100$, $n = 500$, $n = 1000$ et $n = 50000$.

La distribution lognormale est une loi très connue dans l'ajustement des données de revenu. Cette loi présente un avantage sur la distribution de Paréto car ses paramétres sont faciles à estimer et à interpréter. D'autres distributions comme celle de Singh-Maddala, sont aussi adéquates pour ajuster des données de revenu. Cependant, pour des raisons d'espace, nous allons conduire nos simulations, en supposant que la fonction de répartition G suit la loi lognormale $LN(m, \sigma)$ dont la densité de probabilité est donnée par

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{y}\exp\left\{-\frac{1}{2}\left(\frac{\log y - m}{\sigma}\right)^2\right\} \quad y > 0.$$

Nous travaillons ici dans le cas de la mesure de Sen où, l'indice théorique J est égal à

$$J = 2\int_0^z \left(1 - \frac{G(y)}{G(z)}\right)\left(1 - \frac{y}{z}\right) g(y)dy.$$

Cet indice J ainsi que la variance σ^2 dans le théorème 4.1 sont calculés avec la méthode d'intégration numérique de Simpson, en prenant G égale à la distribution lognormale $LN(12, 0.75)$.

Nous générons 100 échantillons de taille n selon la loi $LN(12, 0.75)$. Pour chaque échantillon i, nous effectuons les deux étapes suivantes :
- Calcul de l'indice empirique $J_{n,i}$ dans la cas de Sen qui est égal à :

$$J_{n,i} = \frac{2}{n}\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\left(1 - \frac{Y_{j,n}}{z}\right),$$

où q est le nombre de pauvres dans l'échantillon.
- Calcul de la quantité $z_{n,i} = \frac{\sqrt{n}(J^z_{n,i} - J)}{\sigma}$.

Après ce procédé, nous traçons un graphique quantile-quantile (QQ-plot) pour comparer la série $z_{n,i}, i = 1, \cdots, 100$ à la loi normale centrée réduite. Les résultats sont donnés par la figure 4.1 ci-dessous. Sur le graphique nous observons une linéarité

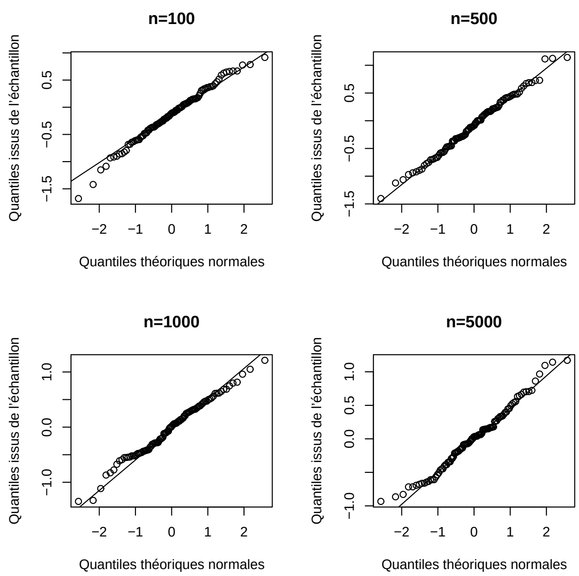

FIGURE 4.1 – *Simulation de la normalité asymptotique de l'estimateur "plug-in" J_n dans le cas de Sen.*

des points ; ce qui témoigne de la normalité asymptotique de l'estimateur "plug-in" J_n.

4.4 Estimation de la variance asymptotique

Puisque la variance σ^2 dépend de la distribution G, on peut utiliser une méthode non-paramétrique pour l'estimer. Pour cela, nous remplaçons la distribution théorique G par son équivalent empirique G_n partout dans l'expression de σ^2. En rappelant les statistiques d'ordre de l'échantillon $Y_1, \cdots, Y_n$:

$$Y_{1,n} \leq \cdots \leq Y_{q,n} < z \leq Y_{q+1} \leq \cdots \leq Y_{n,n},$$

on peut définir un estimateur empirique de la variance σ^2, par

$$s_n^2 = s_{1,n}^2 + s_{2,n}^2 + s_{3,n}^2 + s_{4,n}^2,$$

où

$$s_{1,n}^2 = \frac{1}{n}\sum_{j=1}^{q} w^2\left(\frac{j}{n}, \frac{q}{n}\right) f^2(Y_{j,n}, z) - \left(\frac{1}{n}\sum_{j=1}^{q} w(\frac{j}{n}, \frac{q}{n}) f(Y_{j,n}, z)\right),$$

$$s_{2,n}^2 = \frac{1}{n^2}\sum_{i=1}^{q}\sum_{j=1}^{q}\varphi(Y_{i,n})\varphi(Y_{j,n})(\frac{i}{n}\wedge\frac{j}{n}-\frac{i}{n}\frac{j}{n}),$$

$$s_{3,n}^2 = \frac{q}{n}\left(1-\frac{q}{n}\right)\left(\frac{1}{n}\sum_{j=1}^{q}\phi(Y_{j,n},z)\right)^2,$$

$$s_{4,n}^2 = 2\left(1-\frac{q}{n}\right)\left(\frac{1}{n}\sum_{j=1}^{q}\varphi(Y_{j,n},z)\right)\left(\frac{1}{n}\sum_{j=1}^{q}\phi(Y_{j,n},z)\right).$$

Sous les conditions du Théorème 4.1, il a été prouvé (cf. Zitikis (2002)) que s_n^2 est un estimateur fortement consistant pour σ^2 si la fonction poids $w(\cdot,\cdot)$ est telle que l'application $a : u \mapsto a(u) = w(u, G(z))$ pour tout $z > 0$ fixé, vérifie :

$$|a(u)| \leq cu^{\alpha-1}(1-u)^{\beta-1}, \quad 0 < u < 1, \tag{4.15}$$

pour des constantes $\alpha, \beta > 0$ et $0 < c < \infty$.

Notons que la condition (4.15) est satisfaite pour la plupart des indices de pauvreté. Voici quelques exemples :

Indice FGT. Nous avons $a(u) = 1$, et la condition (4.15) est vérifiée pour $\alpha = \beta = c = 1$.

Indice de Shorrocks. $a(u) = 2(1-u), \quad 0 < u < 1$. Nous avons

$$|a(u)| \leq 2(1-u)^{2-1},$$

et (4.15) est vraie pour $\alpha = 1, \beta = 2$ et $c = 2$.

Indice de Sen. $a(u) = 2(1-u/G(z)), \quad 0 < s < 1$. On a

$$|a(u)| \leq \frac{2}{G(z)}(1-u)^{2-1},$$

et (4.15) est vraie pour $\alpha = 1, \beta = 2$ et $c = 2/G(z)$.

Indice de Kakwani. $a(u) = (k+1)(1-u/G(z))^{k+1}, \quad 0 < s < 1$. On a

$$|a(u)| \leq \frac{k+1}{(G(z))^{k+1}}(1-u)^{(k+2)-1},$$

et (4.15) est vraie pour $\alpha = 1, \beta = k + 2$ and $c = k + 1/(G(z))^{k+1}$.

Avec cette estimation non-paramétrique de la variance σ^2, nous pouvons alors utiliser la formule (4.14) pour estimer l'indice de pauvreté J à partir d'intervalles de confiance de niveau $100(1-\alpha)\%$, avec $0 < \alpha < 1$. Travaillant avec les données de revenu des ménages sénégalais de 1994, nous avons comparé à l'aide de la Figure 4.1 suivante, les intervalles de confiance de niveau $100(1-\alpha)\%$ avec ceux du chapitre 3, ne nécessitant pas de fixer un seuil de signification α et construits autour d'un estimateur à noyau.

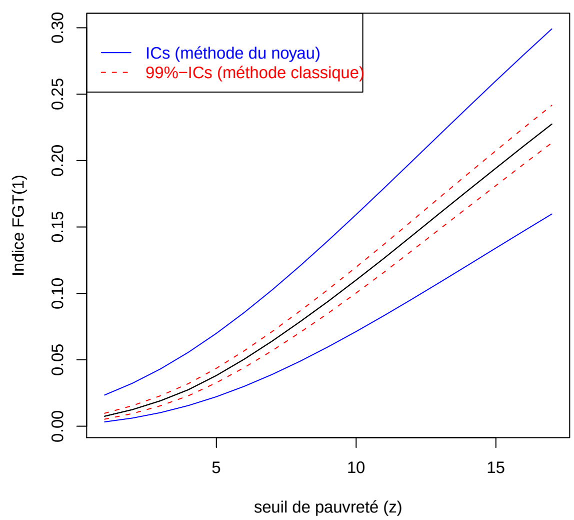

FIGURE 4.2 – *Intervalles de confiance simultanés de l'indice de pauvreté FGT(1). Le trait continu représente les intervalles de confiance asymptotiquement optimaux construits autour de l'estimateur à noyau étudié au chapitre 3 ; le trait discontinu représente des intervalles de confiance uniformes de niveau 99% obtenus avec la normalité asymptotique de l'estimateur "plug-in". Le trait foncé représente la vraie courbe.*

Nous observons que les intervalles de confiance construits autour de l'estimateur à noyau sont plus larges que ceux construits, à l'aide du résultat de normalité asymptotique, autour de l'estimateur "plug-in". Cela paraît naturel car le niveau de confiance est plus élevé avec la méthode du noyau. Cependant, les intervalles de confiances issus de la normalité asymptotique sont plus précis comme le montre la figure 4.1.

Chapitre 5

Convergence faible des estimateurs de type plug-in

5.1 Introduction et Résultat

La théorie de la convergence faible des processus stochastiques joue un rôle fondamental en statistique asymptotique et, particulièrement dans l'approximation des lois de statistiques de test pour vérifier des hypothèses. Il s'agit de la convergence en loi d'une suite de variables aléatoires $(\mathbb{X}_n(\omega, t))$, t variant dans un ensemble T, vers une fonction aléatoire $\mathbb{X}$, à valeurs dans un même espace métrique probabilisable. Lorsque l'ensemble d'indices T est une classe de fonctions, les processus considérés sont dits indexés par des fonctions. Un exemple particulièrement important est le processus empirique indexé par une classe de fonctions, dont la convergence en distribution vers le pont brownien est à l'origine de nombreuses applications statistiques. En effet, le théorème central limite uniforme permet de contrôler simultanément le comportement de la déviation d'une infinité de variables aléatoires par rapport à leurs espérances et du coup, d'établir des bornes d'approximations ou des intervalles de confiances uniformes pour un paramétre fonctionnel donné.

L'intérêt de l'étude de la convergence faible pour les mesures de pauvreté réside dans le fait qu'elle permet d'unifier tous les résultats asymptotiques, établis jusqu'ici avec des approches différentes, dans un cadre unique global afin de développer des méthodes d'inférence robustes, pouvant se baser sur n'importe quelle classe d'indices de pauvreté. Le chapitre 6 suivant traitera de l'inférence. Par contre, dans ce chapitre nous nous intéressons à la convergence faible uniforme de l'estimateur direct de type "plug-in" J_n introduit au chapitre 4, et considéré ici, pour $z > 0$ fixé, comme un processus stochastique indexé par les fonctions $w \in \mathcal{W}$ et $f \in \mathcal{F}$, où $\mathcal{W}$ et $\mathcal{F}$ sont des classes de fonctions appropriées définies ci-dessous. L'estimateur J_n est donné par

$$J_n = \frac{1}{n}\sum_{j=1}^{n} w[G_n(Y_{j,n}), G_n(z)]f(Y_{j,n}, z)1_{\{Y_{j,n}<z\}},$$

et les classes de fonctions $\mathcal{W}$ et $\mathcal{F}$ définies comme suit :

$$\mathcal{W} = \{w : [0,1]\times[0,1] \to \mathbb{R}^+ /\ w \quad \text{continue, et} \;\; u \mapsto w(u,\cdot) \;\; \text{est décroissante}\},$$

et

$$\mathcal{F} = \{f : \mathbb{R}^2 \to \mathbb{R}^+ /\ f \quad \text{continue,} \quad \text{et} \;\; y \mapsto f(y,\cdot) \;\; \text{est décroissante}\}.$$

Pour toutes fonctions $w \in \mathcal{W}$ et $f \in \mathcal{F}$, notons

$$J_n(w,f) := J_n = \frac{1}{n}\sum_{j=1}^{n} w[G_n(Y_{j,n}), G_n(z)]f(Y_{j,n}, z)1_{\{Y_{j,n}<z\}}$$

et

$$J(w,f) := J = \int_0^\infty w[G(y), G(z)]f(y,z)1_{\{y<z\}}dG(y).$$

Posons

$$h_{w,f,G_n} = w[G_n(y), G_n(z)]f(y,z)1_{\{y<z\}} \tag{5.1}$$

et

$$h_{w,f,G} = w[G(y), G(z)]f(y,z)1_{\{y<z\}}. \tag{5.2}$$

Alors en utilisant les notations de la section 4.2 du chapitre 4 et une décomposition similaire à (4.12), on obtient

$$\sqrt{n}[J_n(w,f) - J(w,f)] = \mathbb{G}_n(h_{w,f,G_n} - h_{w,f,G}) + \mathbb{G}_n(h_{w,f,G}) + \sqrt{n}P(h_{w,f,G_n} - h_{w,f,G}), \tag{5.3}$$

où $\mathbb{G}_n$ désigne le processus empirique et P la loi commune des Y_j.

Ainsi, la convergence faible du processus centré normalisé $\sqrt{n}[J_n(w,f) - J(w,f)]$ va dépendre du comportement asymptotique des trois termes du membre de droite de l'égalité (5.3).

Théorème 5.1. *Soit z un réel positif fixé et G une fonction de répartition assosiée à la loi P, admettant une densité de probabilité g. Alors le processus $\{\sqrt{n}[J_n(w,f) - J(w,f)] : w \in \mathcal{W}, f \in \mathcal{F}\}$ converge en distribution dans $l^\infty(\mathcal{W}\times\mathcal{F})$ vers un processus gaussien centré de covariance*

$$\begin{aligned}\Sigma((w_1,f_1);(w_2,f_2)) &= \int_0^z w_1[G(y),G(z)]f_1(y,z)w_2[G(y),G(z)]f_2(y,z)dG(y) \\ &- \int_0^z w_1[G(y),G(z)]f_1(y,z)dG(y)\int_0^z w_2[G(y),G(z)]f_2(y,z)dG(y) + \vartheta,\end{aligned}$$

où

$$\begin{aligned}\vartheta &= \int_0^z\int_0^z (a_1(x,y)+a_2(x,y)+a_3(x,y)+a_4(x,y))\times \\ &\quad (G(x\wedge y)-G(x)G(y))dG(x)dG(y),\end{aligned}$$

avec

$$\begin{aligned}a_1(x,y) &= (w_1)'_u[G(x),G(z)]f_1(x,z)\cdot(w_2)'_u[G(y),G(z)]f_2(y,z) \\ a_2(x,y) &= (w_1)'_u[G(x),G(z)]f_1(x,z)\cdot\frac{w_2[G(y),G(z)]}{g(y)}\frac{\partial f_2}{\partial y}(y,z) \\ a_3(x,y) &= \frac{w_1[G(x),G(z)]}{g_(x)}\frac{\partial f_1}{\partial x}(x,z)\cdot(w_2)'_u[G(y),G(z)]f_2(y,z) \\ a_4(x,y) &= \frac{w_1[G(x),G(z)]}{g(x)}\frac{\partial f_1}{\partial x}(x,z)\cdot\frac{w_2[G(y),G(z)]}{g(y)}\frac{\partial f_2}{\partial y}(y,z).\end{aligned}$$

5.2 Preuve

La preuve de ce théorème sera divisée en trois parties. Dans la première partie, nous montrerons que le processus empirique $\mathbb{G}_n$ indexé par la classe de fonctions $\mathcal{H} = \{h_{w,f,G} : w \in \mathcal{W}, f \in \mathcal{F}\}$, où $h_{w,f,G}$ est une fonction indexée définie par l'équation (5.2), converge en distribution vers un processus gaussien $\mathbb{G}$ dans $l^\infty(\mathcal{H})$. Dans la deuxième, nous prouverons que

$$\sup_{w\in\mathcal{W},f\in\mathcal{F}} |\mathbb{G}_n(h_{w,f,G_n} - h_{w,f,G})| \xrightarrow{P} 0, \quad n\to\infty. \tag{5.4}$$

Et enfin dans la troisième et dernière partie nous établirons, en utilisant la delta-méthode fonctionnelle (cf. Annexe A.4) que, le terme $\sqrt{n}P(h_{w,f,G_n} - h_{w,f,G})$ converge aussi en distribution vers un processus gaussien $\mathbb{W}$ indexé par $(w,f) \in \mathcal{W}\times\mathcal{F}$.

5.2.1 1^{ere} partie

Pour cette partie, il suffit de montrer que la classe de fonctions $\mathcal{H}$ définie par

$$\begin{aligned}\mathcal{H} &= \{h_{w,f,G} : w \in \mathcal{W}, f \in \mathcal{F}\} \\ &= \{y \mapsto w[G(y), G(z)]f(y,z)1_{\{y<z\}} : w \in \mathcal{W}, f \in \mathcal{F}\}\end{aligned}$$

est P-Donsker, où rappelons-le, $P = \mathbb{P} \circ Y$ est la loi commune des Y_j. Nous pouvons écrire $\mathcal{H}$ sous la forme $\mathcal{H} = \mathcal{H}_1 \cdot \mathcal{H}_2$, où

$$\mathcal{H}_1 = \{y \mapsto w[G(y), G(z)] : w \in \mathcal{W}\}$$

et

$$\mathcal{H}_2 = \{y \mapsto f(y,z)1_{\{y<z\}} : f \in \mathcal{F}\}.$$

Ainsi, $\mathcal{H}$ sera une classe de Donsker si et seulement si $\mathcal{H}_1$ et $\mathcal{H}_2$ sont des classes de Donsker, car $\mathcal{H}_1$ et $\mathcal{H}_2$ sont constituées de fonctions supposées à valeurs dans $[0,1]$, donc uniformément bornées. Pour montrer que $\mathcal{H}_i, i = 1,2$ est P-Donsker nous allons établir les trois conditions (C.1),(C.2) et (C.3) suivantes :

(C.1) : $P^*H_i < \infty$, où H_i est une enveloppe mesurable de la classe $\mathcal{H}_i$. Ici, P^* désigne une mesure extérieure telle que pour toute fonction g mesurable

$$P^*g = Pg = \int g dP.$$

(C.2) : $\mathcal{H}_i$ est P-mesurable.

(C.3) : $\mathcal{H}_i$ vérifie la condition d'entropie uniforme, i.e.

$$\int_0^1 \sup_Q \sqrt{\log N(\epsilon \left\|H_i\right\|_{2,Q}, \mathcal{H}_i, L_2(Q))} d\epsilon < \infty, \qquad (EU)$$

où Q décrit l'ensemble des mesures de probabilité sur $\mathbb{R}$ et $N(\epsilon, \mathcal{H}_i, L_2(Q)), i = 1,2$ est le nombre de $\epsilon-$ recouvrements de $\mathcal{H}_i$, i.e. le nombre minimal de boules, par rapport à la norme $L_2(Q)$, de rayon ϵ nécessaire pour couvrir $\mathcal{H}_i$.

Pour $z > 0$ fixé, et G une fonction de répartition donnée, les fonctions $y \mapsto w[G(y), G(z)]$ et $y \mapsto f(y,z)1_{\{y<z\}}$ sont décroissantes par hypothèse, en vertu des propriétés normatives d'une mesure de pauvreté. Sans perte de généralité, on peut les considérer comme prenant leurs valeurs dans $[0,1]$. Ainsi, les classes $\mathcal{H}_1$ et $\mathcal{H}_2$

possédent une fonction enveloppe commune qui est la constante $H = 1$.
Nous observons que

$$\mathcal{H}_1 = \{y \mapsto w[G(y), \cdot] : w \in \mathcal{W}\} \subset \mathcal{M}_b(\mathbb{R}, [0,1]),$$

où $\mathcal{M}_b(\mathbb{R}, [0,1])$ est l'ensemble des fonctions monotones bornées sur $\mathbb{R}$ à valeurs dans l'intervalle $[0,1]$. Donc, en vertu du Théorème 2.7.5, page 159, van der Vaart et Wellner (1996), on a pour $r = 2$, que

$$\sup_Q \sqrt{\log N_{[]}(\epsilon, \mathcal{H}_1, L_2(Q))} < K\epsilon^{-1}, \quad 0 < \epsilon < 1,$$

où le suprémum est pris sur toutes les mesures de probabilité Q sur $\mathbb{R}$ et K est une constante positive. L'inégalité (5.5) suivante est bien connue (cf. van der Vaart et Wellner (1996)). Pour toute classe de fonctions $\mathcal{G}$ on a

$$N(\epsilon, \mathcal{G}, L_2(Q)) \leq N_{[]}(2\epsilon, \mathcal{G}, L_2(Q)) \tag{5.5}$$

Alors on en déduit que

$$\sup_Q \sqrt{\log N(\epsilon, \mathcal{H}_1, L_2(Q))} < \sup_Q \sqrt{\log N_{[]}(2\epsilon, \mathcal{H}_1, L_2(Q))}.$$

Puisque $\|H\|_{2,Q} = 1$, cela implique que

$$\sup_Q \sqrt{\log N(\epsilon\|H\|_{2,Q}, \mathcal{H}_1, L_2(Q))} < \frac{K}{2}\epsilon^{-1}, \quad 0 < \epsilon < 1.$$

D'où

$$\int_0^1 \sup_Q \sqrt{\log N(\epsilon\|H\|_{2,Q}, \mathcal{H}_1, L_2(Q))} d\epsilon \leq \sqrt{\frac{K}{2}} \int_0^1 \epsilon^{-1/2} d\epsilon = \sqrt{2K} < \infty.$$

Donc $\mathcal{H}_1$ satisfait à la condition d'entropie uniforme (C.3). De manière analogue, on montre que $\mathcal{H}_2$ vérifie aussi cette condition (C.3).

Pour vérifier la $P-$ mesurabilité (condition C.2), nous allons montrer que la classe $\mathcal{H}_1$ est convenablement mesurable, i.e. admet une sous-classe $\mathcal{H}_1^0$ dénombrable et dense (voir Annexe A.2). Pour cela, il suffit de prouver (voir Exemple 2.3.4, page 110, van der Vaart et Wellner (1996)) que la classe $\mathcal{H}_1$ est séparable pour la norme suprémum. Une condition suffisante pour que $\mathcal{H}_1$ soit ainsi est que $\mathcal{H}_1$ soit totalement bornée pour la norme suprémum. Rappelons que $\mathcal{H}_1$ est une classe de fonctions décroissantes sur $\mathbb{R}$ à valeurs dans $[0,1]$. Donc si $-\mathcal{H}_1$ composée de fonctions croissantes est totalement bornée, $\mathcal{H}_1$ sera aussi totalement bornée.

Notons N_∞ le nombre minimal de boules de rayon $\delta > 0$, par rapport à la norme suprémum $\|\cdot\|_\infty$, nécessaires pour couvrir une classe de fonctions $\mathcal{G}$, alors $\log N_\infty$ est appelé δ-entropie de $\mathcal{G}$ pour la norme suprémum.

Proposition 5.1. *van de Geer (2000)*
Si $\mathcal{G}$ est une classe de fonctions croissantes $g : \mathbb{R} \to [0,1]$ et F un sous-ensemble fini de $\mathbb{R}$ de cardinal n, alors

$$\log N_\infty(\delta, \mathcal{G}) \leq \lfloor \frac{1}{\delta} \rfloor \log \left(n + \lfloor \frac{1}{\delta} \rfloor \right), \quad \forall \delta > 0,$$

où $\lfloor x \rfloor$ désigne la partie entière d'un nombre réel x.

Remarque 5.1. *Cette proposition nous permet d'établir que la $\delta-$entropie de $\mathcal{H}_1$ par rapport à la norme suprémum est finie pour tout $\delta > 0$. Ce qui entraîne immédiatement que $\mathcal{H}_1$ est totalement bornée pour la norme suprémum.*

Preuve. La preuve de cette proposition est condensée dans van der Geer (2000). Nous la détaillons ici pour mettre le lecteur à l'aise. Supposons que les points du sous-ensemble F soient ordonnés comme suit : $x_1 \leq x_2 \leq \cdots \leq x_n$. Pour tout $g \in \mathcal{G}$, posons

$$M_i = \lfloor \frac{g(x_i)}{\delta} \rfloor, \quad i = 1, \cdots, n$$

avec δ un réel positif donné, et définissons la fonction

$$\tilde{g}(x_i) = \delta M_i, \quad i = 1, \cdots, n.$$

Alors on a, pour tout $i = 1, \cdots, n$,

$$|\tilde{g}(x_i) - g(x_i)| = |\delta \lfloor \frac{g(x_i)}{\delta} \rfloor - \frac{\delta g(x_i)}{\delta}| = \delta |\lfloor \frac{g(x_i)}{\delta} \rfloor - \frac{g(x_i)}{\delta}| \leq \delta, \quad \forall \delta > 0.$$

Comme g est croissante, on a

$$0 \leq M_1 \leq M_2 \leq \cdots \leq M_n \leq \lfloor 1/\delta \rfloor,$$

car $g(x) \leq 1, \forall x \in \mathbb{R}$. Le nombre de fonctions $\tilde{g}$ que l'on peut construire avec ce procédé est :

$$C_{n+\lfloor 1/\delta \rfloor}^{\lfloor 1/\delta \rfloor} = \frac{(n + \lfloor 1/\delta \rfloor) \times \cdots \times (n+1)}{\lfloor 1/\delta \rfloor \times (\lfloor 1/\delta \rfloor - 1) \times \cdots \times 1}$$

donc

$$\log C_{n+\lfloor 1/\delta\rfloor}^{\lfloor 1/\delta\rfloor} = \sum_{j=1}^{\lfloor 1/\delta\rfloor} \log(n+j) - \log(j) \leq \sum_{j=1}^{\lfloor 1/\delta\rfloor} \log(n+j) \leq \lfloor 1/\delta\rfloor \log(n+\lfloor 1/\delta\rfloor).$$

Comme la $\delta-$entropie de $\mathcal{G}$ pour la norme suprémum est le logarithme du nombre de boules de centre $\tilde{g}$ et de rayon δ, nécessaire pour couvrir $\mathcal{G}$, alors on a

$$\log N_\infty(\delta, \mathcal{G}) = \log C_{n+\lfloor 1/\delta\rfloor}^{\lfloor 1/\delta\rfloor} \leq \lfloor 1/\delta\rfloor \log(n+\lfloor 1/\delta\rfloor).$$

Ce qui achève la preuve de la proposition.

On peut donc déduire de cette proposition que pour tous $n \in \mathbb{N}$ et $\delta > 0$, fixés

$$\log N_\infty(\delta, \mathcal{G}) < \infty.$$

D'où $\mathcal{G}$ est totalement bornée pour la norme suprémum, donc $\mathcal{G}$ est séparable pour cette norme. Par conséquent, $\mathcal{G}$ est convenablement mesurable. En appliquant cette proposition à $-\mathcal{H}_1$, on en déduit que $\mathcal{H}_1$ est convenablement mesurable, donc $P-$mesurable, i.e. satisfait à la condition (C.2). Avec le même raisonnement, on montre que $\mathcal{H}_2$ satisfait aussi à cette condition (C.2).

La condition (C.1) est immédiate. En effet, comme $\mathcal{H}_1$ et $\mathcal{H}_2$ sont des classes de fonctions à valeurs dans $[0,1]$, on peut prendre comme fonction enveloppe commune la constane $H = 1$ qui est mesurable, donc

$$P^*H^2 = PH^2 = 1 < \infty.$$

Nous venons de montrer que les classes de fonctions $\mathcal{H}_1$ et $\mathcal{H}_2$ vérifient les conditions (C.1), C.2) et (C.3), donc $\mathcal{H} = \mathcal{H}_1 \cdot \mathcal{H}_2$ est $P-$Donsker, i.e. le processus $\{\mathbb{G}_n(h_{w,f,G}) : h_{w,f,G} \in \mathcal{H}\}$ converge en distribution vers un processus gaussien $\mathbb{G}$ dans $l^\infty(\mathcal{H})$, de fonction de covariance

$$\begin{aligned} \mathrm{cov}(\mathbb{G}(h_{w_1,f_1,G}), \mathbb{G}(h_{w_2,f_2,G})) &= Ph_{w_1,f_1,G}h_{w_2,f_2,G} - Ph_{w_1,f_1,G}Ph_{w_2,f_2,G} \\ &= \int_0^z w_1[G(y),G(z)]f_1(y,z)w_2[G(y),G(z)]f_2(y,z)dG(y) - \\ &\quad \int_0^z w_1[G(y),G(z)]f_1(y,z)dG(y) \int_0^z w_2[G(y),G(z)]f_2(y,z)dG(y). \end{aligned}$$

Par construction, la classe $\mathcal{H}$ peut être identifiée à la classe $\mathcal{W} \times \mathcal{F}$, donc on a

$$\mathbb{G}_n \rightsquigarrow \mathbb{G} \quad \text{dans} \quad l^\infty(\mathcal{W} \times \mathcal{F}). \tag{5.6}$$

5.2.2 $2^{ème}$ partie

Dans cette partie on note G^{-1} la fonction des quantiles associée à la fonction de répartition G qui admet une densité, donc G^{-1} est dérivable et on suppose que sa dérivée est bornée. De même, on admet aussi que les fonctions bi-variables w et f possèdent des dérivées partielles premières bornées.
Pour $z > 0$ fixé, $w \in \mathcal{W}$ et $f \in \mathcal{F}$, considérons la fonction θ définie par

$$\theta(u) = w(u, G(z))f(G^{-1}(u), z)1_{\{G^{-1}(u)<z\}}, \quad u \in [0, 1].$$

Alors

$$h_{w,f,G}(y) = w(G(y), G(z))f(y, z)1_{\{y<z\}} = \theta(G(y)). \tag{5.7}$$

Pour simplifier les notations au niveau des indices, on peut identifier le couple de fonctions (w, f) à la fonction θ en posant

$$h_{\theta,G}(y) := h_{w,f,G}(y) = \theta(G(y)).$$

Soit

$$\Theta = \{u \mapsto \theta(u) = w(u, G(z))f(G^{-1}(u), z)1_{\{G^{-1}(u)<z\}} : w \in \mathcal{W}, f \in \mathcal{F}\}. \tag{5.8}$$

Pour établir (5.4), il suffit de montrer que

$$\sup_{\theta\in\Theta} |\mathbb{G}_n(h_{\theta,G_n} - h_{\theta,G})| \xrightarrow{P} 0, \ n \to \infty. \tag{5.9}$$

Pour cela, nous allons utiliser les techniques développées dans l'article de van der Vaart et Wellner (2007). La classe de fonctions Θ doit alors satisfaire la condition d'entropie uniforme pour une fonction enveloppe donnée, mais aussi, elle doit être convenablement mesurable.

w et f étant des fonctions à valeurs dans $[0, 1]$, donc la constante $H = 1$ est une fonction enveloppe pour les classes $\mathcal{W}$ et $\mathcal{F}$. Par définition de la classe Θ, $H = 1$ est aussi une fonction enveloppe pour Θ. Ainsi la décroissance des fonctions $u \mapsto w(u, \cdot)$ et $u \mapsto f(G^{-1}(u), \cdot)$ implique que Θ est un sous-ensemble de $\mathcal{M}(\mathbb{R}, [0, 1])$, ensemble des fonctions monotones $g : \mathbb{R} \to [0, 1]$. Par conséquent, on peut monter comme dans la partie I de la preuve que, Θ satisfait à la condition d'entropie uniforme

$$\int_0^1 \sup_Q \sqrt{\log N(\epsilon\|H\|_{2,Q}, \Theta, L_2(Q))} d\epsilon < \infty \tag{5.10}$$

et que Θ est convenablement mesurable.

Rappelons que si G_n désigne la fonction de répartition empirique associée à la distribution G, alors d'après le principe d'invariance de Doob-Donsker, $\sqrt{n}(G_n - G)$ converge en distribution vers $B \circ G$ dans $l^\infty(\mathbb{R})$, l'ensemble des fonctions réelles bornées sur $\mathbb{R}$. B étant un pont brownien défini sur $C[0,1]$, l'espace des fonctions continues sur $[0,1]$.
Posons

$$D = l^\infty(\mathbb{R}) = \{d : \mathbb{R} \to \mathbb{R}/ \ \|d\|_\infty < \infty\}$$

et

$$D_0 = C_b(\mathbb{R}) = \{d : \mathbb{R} \to \mathbb{R}/ \ \|d\|_\infty < \infty \text{ et } d \text{ continue}\}.$$

Alors comme B est à trajectoires dans $C[0,1]$, $B \circ G$ est à trajectoires dans D_0.

Pour tous $d_0 \in D_0, \theta \in \Theta$ et $\delta > 0$ fixés, introduisons la suite de classes de fonctions

$$\mathcal{F}_n(\theta, d_0, \delta) = \{h_{\theta, G+n^{-1/2}d} - h_{\theta, G+n^{-1/2}d_0} : d \in D, \|d - d_0\|_\infty < \delta\},$$

avec $F_n(\theta, d_0, \delta)$ une fonction enveloppe mesurable de $\mathcal{F}_n(\theta, d_0, \delta)$ pour tout $n \geq 1$. Pour établir (5.9), il suffit de vérifier (cf. Théorème 2.3, van der Vaart et Wellner (2007)) les trois conditions (i), (ii) et (iii) suivantes pour la classe de fonctions Θ :

(i) $\sup_{\theta\in\Theta} |\mathbb{G}_n(h_{\theta, G+n^{-1/2}d_0} - h_{\theta,G})| \xrightarrow{P} 0, \ n \to \infty \quad \forall d_0 \in D_0.$

(ii) $\sup_{\theta\in\Theta} |\mathbb{G}_n F_n(\theta, d_0, \delta)| \xrightarrow{P} 0, \ n \to \infty \quad \forall \delta > 0, \forall d_0 \in D_0.$

(iii) $\sup_{\theta\in\Theta} \sup_{d_0 \in K} F_n(\theta, d_0, \delta_n) \xrightarrow{P} 0, \ n \to \infty \quad \forall \delta_n \to 0, \forall K \subset D_0, K$ compact.

Vérifions maintenant ces trois conditions pour la classe de fonctions Θ. Pour $z > 0$ fixé, posons

$$a(u) = w(u, G(z)), \quad u \in [0,1]$$

et

$$b(u) = f(G^{-1}(u), z), \quad u \in [0,1].$$

Alors $a(\cdot)$ et $b(\cdot)$ sont des fonctions dérivables car $w(\cdot,\cdot)$ et $f(\cdot,\cdot)$ sont différentiables par hypothèse. En appliquant le théorème de la moyenne, on obtient compte tenu de la relation (5.7), que pour tout $y \in [0, z]$,

$$\begin{aligned}
|h_{\theta,G+n^{-1/2}d}(y) - h_{\theta,G+n^{-1/2}d_0}(y)| &= |\theta(G(y) + n^{-1/2}d(y)) - \theta(G(y) + n^{-1/2}d_0(y))| \\
&= |w[G(y) + n^{-1/2}d(y), G(z)]f[G^{-1}(G(y) + n^{-1/2}d(y)), z] - \\
&\quad w[G(y) + n^{-1/2}d_0(y), G(z)]f[G^{-1}(G(y) + n^{-1/2}d_0(y)), z]| \\
&\leq |a(G(y) + n^{-1/2}d(y)) - a(G(y) + n^{-1/2}d_0(y))| \times \\
&\quad |b(G(y) + n^{-1/2}d(y))| \\
&\quad + |b(G(y) + n^{-1/2}d(y)) - b(G(y) + n^{-1/2}d_0(y))| \times \\
&\quad |a(G(y) + n^{-1/2}d_0(y))| \\
&\leq \sup_{u\in[0,1]} |b(u)| \sup_{u\in[0,1]} |a'(u)| \times |\frac{d(y) - d_0(y)}{\sqrt{n}}| + \\
&\quad \sup_{u\in[0,1]} |a(u)| \sup_{u\in[0,1]} |b'(u)| \times |\frac{d(y) - d_0(y)}{\sqrt{n}}|.
\end{aligned}$$

Donc pour tout $y \in [0, z]$,

$$|h_{\theta,G+n^{-1/2}d}(y) - h_{\theta,G+n^{-1/2}d_0}(y)| \leq \frac{M}{\sqrt{n}} \times \|d - d_0\|_\infty, \tag{5.11}$$

où

$$M = \sup_{u\in[0,1]} |b(u)| \sup_{u\in[0,1]} |a'(u)| + \sup_{u\in[0,1]} |a(u)| \sup_{u\in[0,1]} |b'(u)|.$$

Ainsi, pour tout n, la classe $\mathcal{F}_n(\theta, d_0, \delta)$ admet une fonction enveloppe constante de la forme

$$F_n(\theta, d_0, \delta) = \frac{\delta M}{\sqrt{n}}. \tag{5.12}$$

Cela implique, en notant $\|\mathbb{G}_n\|_\Theta = \sup_{\theta\in\Theta} |\mathbb{G}_n(\theta)|$, que, lorsque $n \to \infty$

$$\sup_{\theta\in\Theta} |\mathbb{G}_n F_n(\theta, d_0, \delta)| \leq \|\mathbb{G}_n\|_\Theta \frac{\delta M}{\sqrt{n}} = \delta M \|\mathbb{P}_n - P\|_\Theta \to 0,$$

pour tout $d_0 \in D_0$ et tout $\delta > 0$ car Θ est une classe de Glivenko-cantelli d'après (5.10). D'où la condition (ii).

La condition (iii) est immédiate, car d'après (5.12), on a

$$\sqrt{n} P F_n(\theta, d_0, \delta_n) = \delta_n M \to 0,$$

quand $\delta_n \to 0$. Par contre, la condition (i) nécessite le lemme suivant :

Lemme 5.1. *Supposons que Φ est une classe convenablement mesurable de fonctions mesurables admettant une fonction enveloppe $\bar{\phi}$ telle que*

$$\int_0^1 \sup_Q \sqrt{\log N(\epsilon \|\bar{\phi}\|_{2,Q}, \Phi, L_2(Q))} d\epsilon < \infty, \tag{5.13}$$

$$P(\bar{\phi} \circ G)^2 < \infty \tag{5.14}$$

et que pour tout $d_0 \in D_0$,

$$P(\bar{\phi} \circ (G + n^{-1/2} d_0))^2 = O(1), \quad P(\bar{\phi} \circ (G + n^{-1/2} d_0))^2 1_{\{\bar{\phi} \circ (G + n^{-1/2} d_0) \geq \epsilon \sqrt{n}\}} \to 0, \tag{5.15}$$

pour tout $\epsilon > 0$. Si de plus,

$$\sup_{\phi \in \Phi} P(\phi \circ (G + n^{-1/2} d_0) - \phi \circ G)^2 \to 0, \quad n \to \infty \tag{5.16}$$

pour tout $d_0 \in D_0$. Alors la condition (i) est vraie.

Preuve. (voir Lemme 3.1, page 241, van der Vaart et Wellner (2007)). Précèdemment, on a établi que la classe de fonctions Θ définie en (5.8) satisfait à la condition d'entropie uniforme et admet une enveloppe constante égale à 1. Donc les conditions (5.13), (5.14) et (5.15) du lemme 5.1 sont remplies. Pour appliquer ce lemme à la classe Θ, il suffit de vérifier la condition (5.16). En appliquant le théorème de la moyenne, comme précédemment, on montre que pour tout $y \in [0, z]$

$$\begin{aligned} |\theta \circ (G(y) + n^{-1/2} d_0(y)) - \theta \circ G(y)| &\leq |w[G(y) + n^{-1/2} d_0(y), G(z)] \times \\ &\quad f[G^{-1}(G(y) + n^{-1/2} d_0(y)), G(z)] \\ &\quad - w[G(y), G(z)] f[G^{-1} \circ G(y), G(z)]| \\ &\leq \frac{M \|d_0\|_\infty}{\sqrt{n}}. \end{aligned}$$

Donc

$$P(\theta \circ (G + n^{-1/2} d_0) - \theta \circ G)^2 \leq \frac{M \|d_0\|_\infty}{n} \to 0, \quad n \to \infty$$

pour tout $d_0 \in D_0$. D'où la relation (5.16). En appliquant le lemme 5.1 à la classe de fonctions Θ, on obtient la condition (i).

Finalement, on déduit de (i), (ii) et (iii) que (5.9) est vraie, i.e.

$$\sup_{w \in \mathcal{W}, f \in \mathcal{F}} |\mathbb{G}_n(h_{w,f,G_n} - h_{w,f,G})| = \sup_{\theta \in \theta} |\mathbb{G}_n(h_{\theta,G_n} - h_{\theta,G})| \xrightarrow{P} 0, \ n \to \infty. \tag{5.17}$$

5.2.3 $3^{ème}$ partie

Dans cette partie nous appliquons la delta-méthode fonctionnelle (voir Annexe A.4) à l'application $\phi : G \mapsto \phi(G) = Ph_{\theta,G}$, où θ décrit la classe de fonctions Θ, définie en (5.8), i.e.

$$\theta(u) = w(u, G(z))f(G^{-1}(u), z)1_{\{G^{-1}(u)<z\}}, \quad u \in [0,1].$$

Pour tout $y \geq 0$, on rappelle que

$$h_{\theta,G}(y) = \theta(G(y)) = w(G(y), G(z))f(y,z)1_{\{y<z\}} = h_{w,f,G}(y). \tag{5.18}$$

Soit $(t_n)_{n\geq 1}$ une suite de réels et $(d_n)_{n\geq 1}$ une suite d'éléments de $D_0 = C_b(\mathbb{R})$ telles que

$$G + t_n d_n \in D_0, \quad t_n \to 0, \quad d_n \to d_0 \in D_0.$$

En rappelant les fonctions $a(u) = w(u, G(z))$ et $b(u) = f(G^{-1}(u), z)$ définies dans la deuxième partie de la preuve, on peut écrire pour tout $y \in [0, z]$, que

$$\begin{aligned}
h_{\theta,G+t_nd_n}(y) - h_{\theta,G}(y) &= \theta(G(y) + t_nd_n(y)) - \theta(G(y)) \\
&= w[G(y) + t_nd_n(y), G(z)]f[G^{-1}(G(y) + t_nd_n(y)), z] \\
&\quad -w[G(y), G(z)]f[G^{-1}(G(y)), z] \\
&= a[G(y) + t_nd_n(y)]b[G(y) + t_nd_n(y)] - a[G(y)]b[G(y)] \\
&= \{a[G(y) + t_nd_n(y)] - a[G(y)]\}b[G(y) + t_nd_n(y)] \\
&\quad +a[G(y)]\{b[G(y) + t_nd_n(y)] - b[G(y)]\}.
\end{aligned}$$

Ainsi, pour n assez grand, nous avons

$$\begin{aligned}
h_{\theta,G+t_nd_n}(y) - h_{\theta,G}(y) &= a'(G(y))b(G(y))t_nd_n(y) + b'(G(y))a(G(y))t_nd_n(y) + o(t_nd_n(y)) \\
&= \theta'(G(y))t_nd_n(y) + o(t_nd_n(y)),
\end{aligned}$$

où $a'(\cdot), b'(\cdot)$ et $\theta'(\cdot)$ désignent respectivement les fonctions dérivées de $a(\cdot), b(\cdot)$ et $\theta(\cdot)$. Donc, lorsque $n \to \infty$ on a

$$\frac{1}{t_n}P(h_{\theta,G+t_nd_n} - h_{\theta,G}) \to \int_0^z \theta'(G(y))d_0(y)dG(y) =: [\phi'(G) \cdot d_0](\theta),$$

pour tout $\theta \in \Theta$.

Puisque $\sqrt{n}(G_n - G)$ converge en distribution vers le pont brownien $B \circ G$, qui est séparable et à trajectoires dans $C_b(\mathbb{R}) = D_0$, en appliquant le Théorème 3.9.4 (Delta-méthode, voir page 374, van der Vaart et Wellner (1996)) pour l'application $\phi(G) = Ph_{\theta,G}$, on obtient pour tout $\theta \in \Theta$, lorsque $n \to \infty$, que

$$\sqrt{n}(\phi(G_n) - \phi(G)) = \sqrt{n}(Ph_{\theta,G_n} - Ph_{\theta,G}) \rightsquigarrow \int_0^z \theta'(G(y))B(G(y))dG(y) =: \mathbb{W}(\theta).$$

Considérons maintenant le processus

$$\{\mathbb{W}_n(\theta) = \sqrt{n}(Ph_{\theta,G_n} - Ph_{\theta,G}) : \theta \in \Theta\}.$$

Rappelons qu' on a $Ph^2_{\theta,G} < \infty,\ \forall\theta \in \Theta$. Pour montrer que le processus $\mathbb{W}_n$ converge en distribution vers le processus gaussien $\mathbb{W}$, il suffit de prouver que Θ est une classe de Donsker ; ce qui est le cas. En effet, on a déjà montré en (5.10) que Θ vérifie la condition d'entropie uniforme

$$\int_0^1 \sup_Q \sqrt{\log N(\epsilon\|H\|_{2,Q}, \Theta, L_2(Q))}d\epsilon < \infty \tag{5.19}$$

et que Θ est convenablement mesurable, donc P-mesurable. De plus, son enveloppe $H = 1$ satisfait à $P^*H^2 = PH^2 = 1 < \infty$. Donc Θ est P-Donsker, i.e.

$$\mathbb{W}_n \rightsquigarrow \mathbb{W} \quad \text{dans} \quad l^\infty(\Theta).$$

$\mathbb{W}$ est alors un processus gaussien de covariance

$$\mathrm{cov}(\mathbb{W}(\theta_1), \mathbb{W}(\theta_2)) = \int_0^z \int_0^z \theta_1'(G(x))\theta_2'(G(y))[G(x \wedge y) - G(x)G(y)]dG(x)dG(y). \tag{5.20}$$

D'après l'égalité (5.18), la fonction θ peut être identifiée au couple de fonctions (w, f). Donc, on peut écrire

$$\mathbb{W}_n \rightsquigarrow \mathbb{W} \quad \text{dans} \quad l^\infty(\mathcal{W} \times \mathcal{F}), \tag{5.21}$$

et la covariance en (5.20) s'écrit

$$\begin{aligned}\mathrm{cov}(\mathbb{W}(w_1, f_1); \mathbb{W}(w_2, f_2)) \quad = \quad & \int_0^z \int_0^z (a_1(x,y) + a_2(x,y) + a_3(x,y) + a_4(x,y)) \times \\ & (G(x \wedge y) - G(x)G(y))dG(x)dG(y),\end{aligned}$$

avec

$$\begin{aligned}
a_1(x,y) &= (w_1)'_u[G(x),G(z)]f_1(x,z)\cdot(w_2)'_u[G(y),G(z)]f_2(y,z)\\
a_2(x,y) &= (w_1)'_u[G(x),G(z)]f_1(x,z)\cdot\frac{w_2[G(y),G(z)]}{g(y)}\frac{\partial f_2}{\partial y}(y,z)\\
a_3(x,y) &= \frac{w_1[G(x),G(z)]}{g_(x)}\frac{\partial f_1}{\partial x}(x,z)\cdot(w_2)'_u[G(y),G(z)]f_2(y,z)\\
a_4(x,y) &= \frac{w_1[G(x),G(z)]}{g(x)}\frac{\partial f_1}{\partial x}(x,z)\cdot\frac{w_2[G(y),G(z)]}{g(y)}\frac{\partial f_2}{\partial y}(y,z).
\end{aligned}$$

Récapitulation de la preuve :
Nous avons que

$$\begin{aligned}
\sqrt{n}[J_n(w,f)-J(w,f)] &= \sqrt{n}(\mathbb{P}_n h_{w,f,G_n} - Ph_{w,f,G})\\
&= \mathbb{G}_n(h_{w,f,G_n}-h_{w,f,G}) + \sqrt{n}(\mathbb{P}_n h_{w,f,G} - Ph_{w,f,G})\\
&\quad + \sqrt{n}P(h_{w,f,G_n}) - h_{w,f,G})\\
&= \mathbb{G}_n(h_{w,f,G_n}-h_{w,f,G}) + \mathbb{G}_n(h_{w,f,G}) + \mathbb{W}_n(h_{w,f,G}).
\end{aligned}$$

On vient d'établir que :

- $\sup_{w\in\mathcal{W},f\in\mathcal{F}}\mathbb{G}_n(h_{w,f,G_n}-h_{w,f,G}) \xrightarrow{P} 0$, en (5.17).
- $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ dans $l^\infty(\mathcal{W}\times\mathcal{F})$, en (5.6).
- $\mathbb{W}_n \rightsquigarrow \mathbb{W}$ dans $l^\infty(\mathcal{W}\times\mathcal{F})$, en (5.21).

Donc $\sqrt{n}[J_n^z(w,f)-J^z(w,f)]$ qui est la somme de ces trois processus est asymptotiquement tendu dans $l^\infty(\mathcal{W}\times\mathcal{F})$. De plus, les marges finies de chaque processus convergent vers celles d'un processus gaussien, alors les marges de la somme vont converger aussi vers celles d'un processus gaussien. D'après le théorème de Prohorov (voir Annexe A.1) et compte tenu du lemme 5.2 ci-dessous, le processus $\{\sqrt{n}(J_n^z(w,f)-J^z(w,f)) : w\in\mathcal{W}, f\in\mathcal{F}\}$ converge faiblement dans $l^\infty(\mathcal{W}\times\mathcal{F})$ vers un processus gaussien tendu de covariance

$$\Sigma((w_1,f_1);(w_2,f_2)) = \mathrm{cov}(\mathbb{G}(h_{w_1,f_1,G},\mathbb{G}(h_{w_2,f_2,G})) + \mathrm{cov}(\mathbb{W}(w_1,f_1),\mathbb{W}(w_2,f_2)).$$

La covariance croisée des deux processus est nulle comme on l'a montré dans la preuve du résultat de normalité asymptotique au chapitre 4.

Lemme 5.2. *Deux processus tendus X et Y à valeurs dans $l^\infty(T)$ sont identiques en loi si toutes leurs distributions marginales finies sont identiques en loi.*

Chapitre 6

Inférence avec les indices de pauvreté

6.1 Test de dominance de pauvreté

L'une des principales raisons pour mesurer la pauvreté est d'établir des comparaisons dans le temps et dans l'espace. Dans ce chapitre nous allons présenter un test de comparaison entre deux distributions, lequel test sera basé sur la structure de covariance établie au chapitre 5. Rappelons que pour $z > 0$ fixé, notre indice général de pauvreté J peut s'exprimer comme une fonctionnelle de w et de f

$$J = \int_0^z w(G(y), G(z)) f(y, z) dG(y) =: J(w, f),$$

où w et f appartiennent à des classes de fonctions appropriées. Notons que chaque couple de fonctions (w, f) définit un indice de pauvreté spécifique, de sorte que la covariance du processus J pour deux couples différents (w_1, f_1) et (w_2, f_2) peut être assimilée à la covariance entre deux mesures de pauvreté différentes. Cette covariance est asymptotiquement définie par

$$\begin{aligned}
\Sigma((w_1, f_1); (w_2, f_2)) &= \int_0^z w_1(G(y), G(z)) f_1(y, z) w_2(G(y), G(z)) f_2(y, z) dG(y) \\
&- \int_0^z w_1(G(y), G(z)) f_1(y, z) dG(y) \times \\
&\quad \int_0^z w_2(G(y), G(z)) f_2(y, z) dG(y) + \vartheta,
\end{aligned}$$

où

$$\begin{aligned}\vartheta &= \int_0^z \int_0^z (a_1(x,y)+a_2(x,y)+a_3(x,y)+a_4(x,y)) \times \\ &\quad (G(x \wedge y) - G(x)G(y))dG(x)dG(y),\end{aligned}$$

avec

$$\begin{aligned}a_1(x,y) &= (w_1)'_u[G(x),G(z)]f_1(x,z)\cdot(w_2)'_u[G(y),G(z)]f_2(y,z)\\ a_2(x,y) &= (w_1)'_u[G(x),G(z)]f_1(x,z)\cdot\frac{w_2[G(y),G(z)]}{g(y)}\frac{\partial f_2}{\partial y}(y,z)\\ a_3(x,y) &= \frac{w_1[G(x),G(z)]}{g_(x)}\frac{\partial f_1}{\partial x}(x,z)\cdot(w_2)'_u[G(y),G(z)]f_2(y,z)\\ a_4(x,y) &= \frac{w_1[G(x),G(z)]}{g(x)}\frac{\partial f_1}{\partial x}(x,z)\cdot\frac{w_2[G(y),G(z)]}{g(y)}\frac{\partial f_2}{\partial y}(y,z).\end{aligned}$$

Il est donc clair que cette covariance dépend de la distribution $G(y)$. Pour faire de l'inférence avec, il convient de l'estimer de façon consistante. Cela peut être réalisé avec deux types d'approches différentes.

La première consiste à choisir la distribution $G(y)$ parmi l'un des modèles paramétriques connus. Dans ce cas la covariance est une fonction des paramétres inconnus du modèle choisi. En remplaçant ces paramétres inconnus par leurs estimateurs de maximum de vraisemblance ou d'autres estimateurs consistants, on obtient un estimateur consistant de la covariance.

L'autre approche qui est celle que nous utilisons ici, correspond au cas non-paramétrique. Si nous supposons que $G(y)$ admet une densité de probabilité $g(y)$ par rapport à la mesure de Lebesgue, on peut exprimer la covariance Σ en fonction de $g(y)$. En disposant d'un estimateur non-paramétrique consistant de cette densité g, on peut obtenir facilement un estimateur consistant de la covariance Σ.

Dans la littérature il existe beaucoup de méthodes non-paramétriques pour l'estimation de la densité. Silverman (1986) a fait une revue détaillée de ces méthodes allant des veilles méthodes d'histogramme aux plus sophistiquées telle que la méthode du noyau. Cette dernière est probablement la mieux connue et la plus populaire car un nombre très important de travaux sont consacrés à l'estimation de la densité par la méthode du noyau. L'estimateur à noyau de la densité de probabilité $g(y)$ est généralement défini par

$$g_n(y) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{y - Y_j}{h}\right),$$

où $K(\cdot)$ est une fonction noyau et h est un paramétre de lissage qui dépend de la taille n de l'échantillon $Y_1, \cdots, Y_n$. Sous certaines conditions de régularité sur $K(\cdot)$, Parzen (1962) a montré que $g_n(y)$ converge en probabilité vers $g(y)$, pour tout réel y. La consistance forte de $g_n(y)$ pour $g(y)$ a été également établie (voir par exemple, Silverman (1978)). Ainsi, chaque terme de la covariance Σ, dépendant de la seule inconnue $g(y)$, peut être estimé de manière consistante. En appliquant le lemme de Slutsky, on obtient finalement un estimateur consistant de la covariance Σ.

Les méthodes statistiques pour comparer deux distributions en termes de pauvreté, ou *Partial poverty orderings* en anglais, sont nombreuses et variées dans la littérature. Ce sont des tests d'hypothèses de dominance ou de non-dominance stochastique.

Considérons deux distributions de revenu F et G. Soit $\mathcal{P}$ une classe d'indices de pauvreté. On peut définir, à l'aide de ces indices, une relation de dominance ou d'ordre partiel entre F et G. On dit que F domine G si pour tout seuil de pauvreté $z > 0$ fixé,

$$J(F, z) \leq J(G, z), \quad \forall J \in \mathcal{P},$$

où $J(F, z)$ et $J(G, z)$ sont les mesures de pauvreté correspondant respectivement aux distributions F et G.

Beach et Richmond (1985) ont développé une technique pour tester cette hypothèse de dominance contre l'hypothèse alternative de non-dominance. Cette technique fut généralisée par Bishop, Formby et Thistle (1992) qui proposèrent un test dit d'union-intersection. Howes (1994) propose une méthode inverse dite test d'intersection-union pour l'hypothèse nulle de dominance contre l'hypothèse alternative de non-dominance. Tous ces deux types de test sont faciles à mettre en oeuvre ; mais comportent des lacunes car ils sont basés sur une grille finie de points choisis arbitrairement. De plus, ils sont impuissants devant les tests qui utilisent la structure de covariance des indices de pauvreté considérés. Pour pallier ces défauts, on peut utiliser le test généralisé de Wald qui a été décrit par Kodde et Palm (1986) et Wolak (1989). La procédure de test est la suivante. Supposons que l'on veuille comparer deux distributions F et G en utilisant d indices de pauvreté donnés. On note par J_F et J_G les vecteurs de $\mathbb{R}^d$ dont les composantes sont respectivement les d indices de pauvreté théoriques correspondant à la distribution F et à la distribution G. La méthode généralisée de Wald (cf. Zheng (2001)) nous permet de tester les types d'hypothèses

$$H_0 : J_F = J_G \qquad \text{contre} \qquad H_1 : J_F \succeq J_G$$

et

$$H_0 : J_F \succeq J_G \qquad \text{contre} \qquad H_1 : J_F \not\succeq J_G$$

où $J_F \succeq J_G$ signifie que la population correspondant à la distribution F est moins pauvre que celle correspondant à la distribution G, i.e. F domine G.

Supposons que deux échantillons aléatoires de tailles respectives n_F et n_G sont tirés indépendamment des deux populations étudiées. Les estimateurs des vecteurs d'indices J_F et J_G sont notés $\hat{J}_F$ et $\hat{J}_G$; leurs matrices de covariances respectives sont estimées par $\hat{\Sigma}_F$ et $\hat{\Sigma}_G$. Posons

$$\Delta\hat{J} = J_F - J_G,$$

alors la variance de $\Delta\hat{J}$ peut être estimée par

$$\hat{\Sigma} = \frac{1}{n_F}\hat{\Sigma}_F + \frac{1}{n_G}\hat{\Sigma}_G.$$

Pour exprimer les statistiques de test de Wald, il faut résoudre le problème de minimisation suivant :

$$\min_{v\geq 0}(\Delta\hat{J} - v)'\hat{\Sigma}^{-1}(\Delta\hat{J} - v). \tag{6.1}$$

Si $\tilde{v}$ est une solution de ce problème de minimisation, alors les statistiques de Wald sont :

$$c_1 = (\Delta\hat{J})'\hat{\Sigma}^{-1}(\Delta\hat{J}) - (\Delta\hat{J} - \tilde{v})'\hat{\Sigma}^{-1}(\Delta\hat{J} - \tilde{v}),$$

pour le test de l'égalité ($H_0 : J_F = J_G$) contre la dominance ($H_1 : J_F \succeq J_G$) et,

$$c_2 = (\Delta\hat{J} - \tilde{v})'\hat{\Sigma}^{-1}(\Delta\hat{J} - \tilde{v}),$$

pour le test de la dominance ($H_0 : J_F \succeq J_G$) contre la non-dominance ($H_1 : J_F \not\succeq J_G$).

Il a été prouvé que, sous l'hypothèse nulle H_0, ces statistiques sont asymptotiquement distribuées suivant une somme pondérée de variables du χ^2 à différents degrés de liberté. A ce propos, nous référons à Kodde et Palm (1986), Wolak(1989), Gouriéroux et al. (1982), Kudô (1963) et Perlman (1969).

Pour conclure le test, il faudra comparer les valeurs de c_1 ou c_2 avec les bornes inférieure q_l et supérieure q_u de la valeur critique du test considéré pour un seuil de signification donné $0 < \alpha < 1$. Une table de ces valeurs critiques est disponible dans Kodde et Palm (1986). q_l est obtenue en choisissant un niveau de signification α et en posant les degrés de liberté $dl = 1$, i.e.

$$\alpha = \frac{1}{2}\mathbb{P}(\chi^2(1) \geq q_l).$$

Pour obtenir q_u, on pose les degrés de liberté $dl = d$, i.e.

$$\alpha = \frac{1}{2}\mathbb{P}(\chi^2(d-1) \geq q_u) + \frac{1}{2}\mathbb{P}(\chi^2(d) \geq q_u).$$

La règle de décision du test s'énonce alors comme suit :
- Si $c_1(c_2) < q_l$, alors on accepte H_0
- Si $c_1(c_2) > q_u$, alors on rejette H_0
- Si $q_l < c_1(c_2) < q_u$, on ne peut conclure. Des simulations de Monte Carlo sont alors nécessaires pour compléter l'inférence (voir par exemple page 215, Wolak (1989) oubien Fisher et al. (1998)).

6.2 Illustration empirique

Maintenant, nous allons appliquer ce test pour comparer la situation de pauvreté au Sénégal entre deux périodes : 1994 et 2001. Le seuil de pauvreté était supposé le même, fixé par les experts de la Banque mondiale à 1 $ US/jour/personne. Pour ces deux périodes le Sénégal dispose de données de dépenses annuelles récoltées aucours d'enquêtes menées auprès des ménages par l'Agence Nationale de la Statistique et de la Démographie (ANSD) du Sénégal. Deux échantillons aléatoires simples ont été recueillis indépendamment durant ces deux périodes. Pour 1994 on a observé 3278 ménages, alors que pour la période 2001, 6594 ménages ont été interrogés.

Considérons maintenant les indices de pauvreté de Kakwani (1980) qui sont de la forme

$$J^z(k) = (k+1)\int_0^z \left(1 - \frac{G(y)}{G(z)}\right)^k \left(\frac{z-y}{z}\right) dG(y), \quad k \geq 0. \tag{6.2}$$

$z > 0$, désigne le seuil de pauvreté et $G(y)$ est la distribution du revenu dans la population étudiée. Cette classe d'indices de pauvreté est certainement la forme la plus évoluée des indices, l'outil le plus récent et le plus complet. En effet, on y retrouve la mesure de Foster-Greer-Thorbecke (1984) pour $k = 0$ et la célèbre mesure de Sen (1976) pour $k = 1$. Pour $k \geq 2$, la mesure de Kakwani possède toutes les propriétés désirables pour un indice de pauvreté. Elle a aussi l'avantage de donner un poids plus grand aux plus pauvres. Cela implique qu'une baisse de cet indice indique, en général, une amélioration significative du bien-être des plus pauvres.

Pour des raisons d'espace, nous allons travailler avec $d = 5$ indices de pauvreté de Kakwani correspondant respectivement aux valeurs de $k = 0, 1, 2, 3, 4$. Notons J_{94} le vecteur constitué de ces 5 indices de pauvreté évalués en 1994 et J_{01} celui constitué de ces 5 indices évalués pour la période 2001. Utilisons la procédure généralisée de Wald pour tester l'hypothèse nulle $H_0 : J_{01} \succeq J_{94}$ qui veut dire que le Sénégal est moins pauvre en 2001 qu'en 1994 contre l'hypothèse alternative $H_1 : J_{01} \not\succeq J_{94}$. C'est à dire,

$$H_0 : J_{01} \succeq J_{94} \qquad \text{contre} \qquad H_1 : J_{01} \not\succeq J_{94}.$$

La statistique de test est alors

$$c_2 = (\Delta\hat{J} - \tilde{v})'\hat{\Sigma}^{-1}(\Delta\hat{J} - \tilde{v}).$$

La valeur obtenue pour cette statistique, après résolution du problème de minimisation par la fonction *ConstrOptim* du logiciel R, est $c_2 = -25.78$, avec $\tilde{v} = (0.02, 0.042, 0.04, 0.023, 2.4e - 11)$. Pour un seuil de signification de $\alpha = 1\%$, la borne inférieure q_l de la valeur critique du test est telle que $0.01 = \frac{1}{2}\mathbb{P}(\chi^2(1) \geq q_l)$. D'après la table de Kodde et Palm (1986), on a $q_l = 5.412$, valeur positive qui est largement supérieure à $c_2 = -25.78$. Donc on accepte l'hypothèse H_0 selon laquelle le Sénégal est moins pauvre en 2001 qu'en 1994 au seuil de signification de 1%. Si on effectue le test d'hypothèses

$$H_0 : J_{01} = J_{94} \qquad \text{contre} \qquad H_1 : J_{01} \succeq J_{94}.$$

La statistique utilisée est

$$c_1 = (\Delta\hat{J})'\hat{\Sigma}^{-1}(\Delta\hat{J}) - (\Delta\hat{J} - \tilde{v})'\hat{\Sigma}^{-1}(\Delta\hat{J} - \tilde{v})$$

La valeur obtenue c_1, après résolution du problème de minimisation, est $c_1 = 143,78$. La borne supérieure q_u de la valeur critique, au seuil de 1%, vérifie

$$0.01 = \frac{1}{2}\mathbb{P}(\chi^2(4) \geq q_u) + \frac{1}{2}\mathbb{P}(\chi^2(1) \geq q_u).$$

D'après la table de Kodde et Palm, on a $q_u = 12.483$, valeur qui est largement inférieure à $c_1 = 143,78$. Donc ici, on rejette H_0 ; ce qui confirme la conclusion du premier test.

Ce test de comparaison multiple peut aussi être réalisé avec divers types d'indices de pauvreté. Tout dépend des fonctions w et f qui définissent la forme de l'indice de pauvreté. Ainsi, la structure de covariance décrite au chapitre 5 permet de tester la dominance entre deux distributions avec n'importe quelle classe de mesures de pauvreté. Cela donne un critère de dominance robuste et uniforme, par rapport aux indices de pauvreté, pour classer des distributions de revenu, pourvu que le seuil de pauvreté z soit fixe.

Annexes

A.1 Convergence faible et mesures extérieures

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Pour toute application $U : \mathcal{X} \to \bar{\mathbb{R}}$ mesurable, on pose $U^+ = \max(U, 0)$ et $U^- = \min(U, 0)$.L'application U est quasi-intégrable si au moins $\mathbb{E}U^+$ ou $\mathbb{E}U^-$ est finie ; on écrit alors $\mathbb{E}U = \mathbb{E}U^+ - \mathbb{E}U^-$. Soit $T : \Omega \to \bar{\mathbb{R}}$ une application quelconque. L'intégrale extérieure de T par rapport à $\mathbb{P}$ est définie par

$$\mathbb{E}^*T = \inf\{\mathbb{E}U : U \geq T,\ U \text{ mesurable et quasi-intégrable}\}.$$

De même, pour toute partie $B \subset \Omega$ on définit la mesure extérieure de B par

$$\mathbb{P}^*(B) = \inf\{\mathbb{P}(A) : A \supset B,\ A \in \mathcal{A}\} = \mathbb{E}^*(1_B),$$

où 1_B est la fonction indicatrice de l'ensemble B. Notons que si T est une application mesurable et quasi-intégrable, alors $\mathbb{E}^*T = \mathbb{E}T$. De même, pour toute partie B mesurable (i.e. $B \in \mathcal{A}$), on a $\mathbb{P}^*(B) = \mathbb{P}(B)$.

De façon similaire, on définit également les notions d'intégrale intérieure et de mesure intérieure. Si $T : \Omega \to \bar{\mathbb{R}}$ est une application quelconque, l'intégrale intérieure de T par rapport à $\mathbb{P}$, notée $\mathbb{E}_*T$ est définie à partir de la relation suivante :

$$\mathbb{E}_*T = -\mathbb{E}^*(-T),$$

i.e.

$$\mathbb{E}_*T = \sup\{\mathbb{E}U : U \leq T,\ U \text{ mesurable et quasi-intégrable}\}.$$

Aussi la mesure intérieure de $B \subset \Omega$ est aussi définie par

$$\mathbb{P}_*(B) = \sup\{\mathbb{P}(A) : A \subset B, A \in \mathcal{A}\} = \mathbb{E}_*(1_B).$$

Ces notions d'intégrale et de mesure extérieures où la condition de mesurabilité n'est pas nécessaire sont à la base de la théorie moderne de la convergence faible (appelée aussi convergence vague). En effet, la théorie classique traitée par Billingsley (1968) s'appuie sur la mesurabilité des processus considérés ; ce qui limite considérablement les applications statistiques. Cette nouvelle théorie, en revanche, donne lieu à de nombreuses applications aussi bien en statistique paramétrique que non-paramétrique. Par exemple, elle s'est révélée particulièrement intéressante pour l'étude de la convergence faible des processus empiriques qui sont des processus non mesurables, car ayant souvent leurs trajectoires dans des espaces de Banach non séparables.

Dans ce qui suit, nous rappelons quelques points essentiels de cette nouvelle théorie de la convergence faible sans mesurabilité.

Soit $\{X_n, n \geq 1\}$ une suite de processus non nécessairement mesurables définis sur des espaces de probabilité $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ à valeurs dans un même espace métrique $(\mathcal{M}, \mathcal{T}_{bor}, d)$ équipé de ses boréliens, d étant une métrique. Soit X un processus mesurable à valeurs dans $(\mathcal{M}, \mathcal{T}_{bor}, d)$.

Définition 6.1. *X_n converge faiblement vers X si*

$$\lim_{n\to\infty} \mathbb{E}^* f(X_n) = \mathbb{E} f(X), \quad \forall f : \mathcal{M} \to \mathbb{R} \textit{ continue bornée.}$$

On dénote ce type de convergence par $X_n \rightsquigarrow X$.

Beaucoup de résultats établis avec la théorie classique de Billingsley tels que : le théorème de portmanteau, le théorème de l'application continue, le théorème de Prohorov, ainsi que les critères de tension restent encore valables pour la convergence faible sans mesurabilité. Pour ces résultats, nous renvoyons le lecteur à van der Vaart et Wellner (1996). Toutefois, signalons que le théorème de Prohorov nécessite deux nouveaux concepts dans la nouvelle théorie : la tension et la mesurabilité asymptotiques.

Définition 6.2. *- La suite X_n est dite asymptotiquement mesurable si*

$$\lim_{n\to\infty} \mathbb{E}^* f(X_n) - \mathbb{E}_* f(X_n) = 0, \quad \forall f : \mathcal{M} \to \mathbb{R} \textit{ continue bornée.}$$

- La suite X_n est dite asymptotiquement tendue si pour tout $\varepsilon > 0$, il existe un compact $K \subset \mathcal{M}$ tel que

$$\liminf_{n\to\infty} \mathbb{P}_*(X_n \in K^\delta) \geq 1 - \varepsilon, \quad \textit{pour tout } \delta > 0,$$

où $K^\delta = \{y \in \mathcal{M} : d(y,K) < \delta\}$ *est une* δ*-extension de* K.

Remarque. Cette notion de tension asymptotique est moins forte que celle utilisée dans la théorie classique. En effet, $\mathbb{P}_*(X_n \in K^\delta) \geq 1-\varepsilon$ n'est nécessaire que quand on va vers la limite, c'est à dire lorsque $n \to \infty$; contrairement à la tension uniforme, utilisée dans la théorie classique de Billingsley (1968) qui exige que cette probabilité soit supérieure à $\geq 1-\varepsilon$ pour tout n, i.e.

$$\forall \varepsilon > 0,\ \exists K \text{ compact},\ \mathbb{P}(X_n \in K) \geq 1-\varepsilon, \quad \forall n \geq 1.$$

Cependant, si les X_n sont mesurables et définis dans un espace polonais, les notions de tension asymptotique et de tension uniforme sont équivalentes car toute mesure de probabilité dans un espace polonais est tendue (cf. Lemma 1.3.2, van der Vaart et Wellner (1996)).

Lemme 6.1. *i) Si* $X_n \rightsquigarrow X$, *alors* X_n *est asymptotiquement mesurable.*
ii) Si $X_n \rightsquigarrow X$, *alors* X_n *est asymptotiquement tendue ssi* X *est tendue.*

La version suivante du théorème de Prohorov donne la réciproque du lemme 6.1.

Théorème 6.1. *Si* X_n *est une suite de processus asymptotiquement tendue et asymptotiquement mesurable, alors il admet une sous-suite* X_{n_k} *qui converge faiblement vers un processus* X *Borel mesurable et tendu.*

Sous des conditions de tension et de mesurabilité asymptotiques, le théorème de Prohorov montre l'existence d'une limite faible, mesurable et tendue pour une suite de processus non mesurables. L'unicité de la limite découle du fait qu'une mesure de Borel L est déterminée de manière unique par l'application $f \mapsto \int f dL$, $f \in C_b(\mathcal{M})$. Pour plus de détails, voir page 25, van der Vaart et Wellner (1996). Le lemme suivant montre aussi l'unicité de la limite X, si $X \in l^\infty(T)$, l'ensemble des fonctions réelles bornées sur T.

Lemme 6.2. *Deux processus tendus* X *et* Y *à valeurs dans* $l^\infty(T)$ *sont identiques en loi si toutes leurs distributions marginales finies sont identiques en loi.*

A.2 Classes de Donsker

Dans cette section, nous rappelons quelques propriétés des classes de Donsker. Soit $X, X_1, \cdots, X_n$ une suite indépendante de variables aléatoires définies dans un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ à valeurs dans un espace mesurable $(\mathcal{X}, \mathcal{A})$. On définit la mesure empirique $\mathbb{P}_n$ associée à cette suite comme une combinaison linéaire de mesures de Dirac aux points X_i, i.e.

$$\mathbb{P}_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j}.$$

Etant donné une classe $\mathcal{F}$ de fonctions mesurables $f : \mathcal{X} \to \mathbb{R}$, la mesure empirique $\mathbb{P}_n$ induit une application de $\mathcal{F}$ dans $\mathbb{R}$ définie par

$$f \mapsto \mathbb{P}_n f.$$

Pour toute fonction mesurable f et toute mesure signée Q, on note $Qf = \int f dQ$. Soit $P = \mathbb{P} \circ X^{-1}$ la loi de probabilité commune des X_j. On appelle processus empirique indexé par $\mathcal{F}$, l'application $\mathbb{G}_n : \mathcal{F} \mapsto \mathbb{R}$ définie par

$$f \mapsto \mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(f(X_j) - Pf).$$

Cette définition généralise la notion du processus empirique classique indexé par $t \in \mathbb{R}^d$, défini par

$$\alpha_n(t) = \sqrt{n}(G_n(t) - G(t)),$$

où $G_n(t) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{X_j \le t\}}$ est la distribution empirique basée sur $X_1, \cdots, X_n$. En effet, on a

$$\alpha_n(t) = \sqrt{n}(G_n(t) - G(t)) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(1_{\{X_j \le t\}} - \mathbb{E}1_{\{X_j \le t\}})$$

ce qui équivaut à

$$\alpha_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(1_{]-\infty,t]}(X_j) - \mathbb{E}1_{]-\infty,t]}(X_j)) = \mathbb{G}_n f_t,$$

où $f_t(x) = 1_{]-\infty,t]}(x)$ pour tout $t \in \mathbb{R}^d$.

Du fait de ses nombreuses applications, la théorie du processus empirique constitue un outil fondamental en statistique asymptotique. Elle est à la base de nombreuses lois limites car beaucoup de statistiques peuvent s'exprimer comme des fonctionnelles de la distribution empirique. Les propriétés asymptotiques du processus empirique classique sont principalement étudiées dans l'espace $D[0,1]$ et celles du processus des sommes partielles de variables aléatoires i.i.d dans $C[0,1]$. La théorie de la convergence faible sans mesurabilité est d'une grande importance dans la mesure où elle permet de considérer des espaces plus adaptés au processus empirique, qui a naturellement ses trajectoires dans un espace de Banach non-séparable.

L'espace le plus convoité est $l^\infty(T)$, l'espace des fonctions réelles bornées et définies sur T. Si $T = \mathcal{F}$, où $\mathcal{F}$ est une classe de fonctions mesurables $f : \mathcal{X} \to \mathbb{R}$, des versions fonctionnelles de la loi des grands nombres et du théorème centrale limite peuvent être établies sous certaines conditions. Ainsi, une classe de fonctions $\mathcal{F}$ dans laquelle le théorème central limite a lieu uniformément sera appelée classe de Donsker.

Soit $l^\infty(\mathcal{F})$ l'espace des fonctions bornées $H : \mathcal{F} \to \mathbb{R}$, muni de la norme

$$\|H\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |H(f)|.$$

On suppose que $Pf = \mathbb{E}f(X) < \infty$ pour $f \in \mathcal{F}$ et qu'il existe une fonction enveloppe F telle que $\sup_{f\in\mathcal{F}} |f(x)| \leq F(x) < \infty$, $\forall x \in \mathcal{X}$. Alors,

$$\sup_{f\in\mathcal{F}} |f(x) - Pf| < 2F(x) < \infty, \quad \forall x \in \mathcal{X}.$$

Ainsi, le processus empirique $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ sera à trajectoires dans $l^\infty(\mathcal{F})$.

Définition 6.3. *Une classe $\mathcal{F}$ de fonctions mesurables est dite P-Donsker s'il existe un processus tendu $\mathbb{G} : \mathcal{F} \to \mathbb{R}$ tel que*

$$\mathbb{G}_n \rightsquigarrow \mathbb{G} \quad \textit{dans } \; l^\infty(\mathcal{F}).$$

Une classe $\mathcal{F}$ de fonctions mesurables est dite P-Glivenko-Cantelli si

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0 \quad \mathbb{P}^* - p.s.$$

Remarque. Par continuité de la norme, on a $\mathbb{G}_n \rightsquigarrow \mathbb{G} \Rightarrow \|\mathbb{G}_n\|_{\mathcal{F}} \rightsquigarrow \|\mathbb{G}\|_{\mathcal{F}}$. Cela implique que $n^{-1/2}\|\mathbb{G}_n\|_{\mathcal{F}}$ converge en loi vers 0, donc converge en probabilité vers 0 et finalement,

$$n^{-1/2}\|\mathbb{G}_n\|_{\mathcal{F}} = \|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0 \quad \mathbb{P}^* - p.s.$$

Ainsi, toute classe de Donsker est une classe de Glivenko-Cantelli. Mais la réciproque est fausse.

Les propriétés de classe de Donsker et de Glivenko-Cantelli dépendent essentiellement de la complexité (taille) des classes de fonctions considérées. En effet, une classe finie de fonctions de carré intégrable est toujours Donsker, alors que l'espace $\mathcal{L}^2$ constitué de l'ensemble des fonctions de carré intégrable n'est presque jamais une classe de Donsker. Un moyen simple pour contrôler la complexité d'une classe de fonctions $\mathcal{F}$ est la notion d'entropie qui se définit comme le logarithme du nombre de boules (ou crochets) de rayon (ou longueur) ε, nécessaires pour recouvrir la classe $\mathcal{F}$.
Supposons que $(\mathcal{F}, d)$ est espace muni d'une métrique d.

Définition 6.4. *Le nombre de recouvrement, noté $N(\varepsilon, \mathcal{F}, d)$ est le nombre minimal de boules de rayon ε, par rapport à d, nécessaires pour recouvrir $\mathcal{F}$.*
On appelle alors entropie métrique, le logarithme du nombre de recouvrement.

Soit l et u deux fonctions mesurables. On définit un crochet $[l, u]$ comme l'ensemble des fonctions f telles que $l \leq f \leq u$. Un ε-crochet est un crochet $[l, u]$ tel que $d(l, u) < \varepsilon$.

Définition 6.5. *Le nombre de crochet, noté $N_{[]}(\varepsilon, \mathcal{F}, d)$ est le nombre minimal de ε-crochets nécessaires pour recouvrir $\mathcal{F}$.*
On appelle alors entropie crochet, le logarithme du nombre de crochet.

Dans les théorèmes de caractérisation des classes de Donsker ou de Glivenko-Cantelli, on utilise souvent les normes $L_r(Q), r \geq 1$ définies par

$$\|f\|_{r,Q} = \left(\int |f|^r dQ \right)^{1/r},$$

où Q est une mesure de probabilité quelconque.
On définit l'entropie uniforme, par rapport à la norme $L_r(Q)$ d'une classe de fonctions $\mathcal{F}$ possédant une enveloppe notée F par le nombre réel noté

$$N(\varepsilon, \mathcal{F}) = \sup_Q \log N(\varepsilon \|F\|_{r,Q}, \mathcal{F}, L_r(Q)),$$

où le suprémum est pris sur toutes les mesures de probabilité Q telles que $0 < QF^2 < \infty$. Notons que pour les résultats relatifs à l'entropie crochet, on n'a besoin d'aucune hypothèse de mesurabilité. Tandis que les résultats relatifs à l'entropie métrique (uniforme) nécessitent la notion de P-mesurabilité suivante :

Définition 6.6. *Une classe $\mathcal{F}$ de fonctions mesurables f définies sur un espace de probabilité $(\mathcal{X}, \mathcal{A}, P)$ est dite P-mesurable si l'application*

$$(X_1, \cdots, X_n) \mapsto \| \sum_{i=1}^{n} e_i f(X_i) \|_{\mathcal{F}}$$

est mesurable sur la complétion de l'espace produit $(\mathcal{X}^n, \mathcal{A}^n, P^n)$ pour tout $n \geq 1$ et pour tout vecteur $(e_1, \cdots, e_n) \in \{-1, 1\}^n$.

Proposition 6.1. *Si $\mathcal{F}$ est une classe de fonctions mesurables telle que*

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))} d\varepsilon < \infty, \qquad (EC)$$

alors $\mathcal{F}$ est P-Donsker.

Proposition 6.2. *Soit $\mathcal{F}$ une classe de fonctions mesurables possédant une enveloppe F telle que*

$$\int_0^\infty \sup_Q \sqrt{\log N(\varepsilon \|F\|_{2,Q}, \mathcal{F}, L_2(Q))} d\varepsilon < \infty, \qquad (EU)$$

et pour tout $\delta > 0$, la classe de fonctions

$$\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \|f - g\|_{2,Q} < \infty\}$$

est P-mesurable. Si de plus, $P^* F^2 < \infty$ alors $\mathcal{F}$ est P-Donsker.

Remarques.

- Dans la condition d'entropie uniforme (EU), le rayon des boules de recouvrement dépend de la grosseur de l'enveloppe F dans $L_2(Q)$; donc plus F y est grosse, plus il est facile de recouvrir $\mathcal{F}$. Pour $\varepsilon \geq 1$, une seule boule suffit pour recouvrir $\mathcal{F}$, donc $\log N(\varepsilon \|F\|_{2,Q}, \mathcal{F}, L_2(Q)) = 0$. D'où la condition (EU) peut être aussi reformulée de la manière suivante :

$$\int_0^1 \sup_Q \sqrt{\log N(\varepsilon \|F\|_{2,Q}, \mathcal{F}, L_2(Q))} d\varepsilon < \infty,$$

- La condition d'entropie crochet (EC) n'est basée que sur la seule loi de probabilité P. Contrairement à la condition (EU) qui est basée sur le suprémum d'un grand nombre d'entropies métriques. Cela est dû au fait que le nombre de crochets est bien plus grande que le nombre de recouvrement. Par conséquent, les deux conditions ne sont pas comparables.

Dans les applications la notion de P-mesurabilité introduite ci-dessus n'est pas étudiée directement. On utilise un concept beaucoup plus fort de *classe convenablement mesurable* ou *pointwise measurable class* en anglais.

Définition 6.7. *Une classe $\mathcal{F}$ de fonctions mesurables est dite convenblement mesurable s'il existe une sous-classe dénombrable $\mathcal{G} \subset \mathcal{F}$ telle que pour tout $f \in \mathcal{F}$, il existe une suite $\{g_m, m \geq 1\} \subset \mathcal{G}$, avec $g_m(x) \to f(x)$, $\forall x \in \mathcal{X}$.*

Conséquence. Si $\mathcal{F}$ est convenablement mesurable, alors $\mathcal{F}$ est P-mesurable pour toute loi de probabilité P. En effet, le supremum sur la classe $\mathcal{F}$ peut être ramené sur une sous-classe dénombrable, d'où la mesurabilité.

Exemples de classes de fonctions P-mesurables.

- Les fonctions indicatrices d'intervalles ou de boules, les fonctions séparables pour la norme suprémum sont des classes convenablement mesurables donc P-mesurables pour toute probabilité P.
- Si $\mathcal{F}$ est un espace topologique de Suslin admissible, alors $\mathcal{F}$ est P-mesurable pour toute probabilité P.

Tandis que l'entropie crochet est vérifiée par des classes de fonctions assez régulières telles que les fonctions monotones bornées, les fonctions convexes et les indicatrices d'ensembles fermés convexes, un outil remarquable pour contrôler l'entropie uniforme est la notion de VC-classe ou classe de Vapnik-$\check{C}$ervonenkis.

Soit $E = \{x_1, \cdots, x_n\}$ un ensemble quelconque, $\mathcal{C}$ une collection de parties de $\mathcal{X}$. On définit la trace de $\mathcal{C}$ sur E par

$$\mathcal{T}^E(\mathcal{C}) = \{E \cap C : C \in \mathcal{C}\}.$$

On dit que l'ensemble E est *complètement tracé* par $\mathcal{C}$ si toute partie de E est dans la trace de $\mathcal{C}$ sur E, i.e.

$$\forall A \subset E, \quad A \in \mathcal{T}^E(\mathcal{C}).$$

On appelle *indice VC* de la collection d'ensembles $\mathcal{C}$ le plus petit entier n tel que aucun ensemble de cardinal n n'est *complètement tracé* par $\mathcal{C}$. Il est noté $\mathcal{V}(\mathcal{C})$.

Définition 6.8. *Une collection d'ensembles* $\mathcal{C}$ *est une VC-classe d'ensemble (ou une classe de Vapnik-Červonenkis) si son indice* $\mathcal{V}(\mathcal{C}) < \infty$.

Cette notion de VC-classe peut également s'étendre aux classes de fonctions. Soit une fonction $f : \mathcal{X} \to \mathbb{R}$. Le sous-graphe de f est défini par

$$S_f = \{(x,t) \in \mathcal{X} \times \mathbb{R} : t < f(x)\}.$$

Définition 6.9. *Soit* $\mathcal{F} = \{f : \mathcal{X} \to \mathbb{R}\}$ *une classe de fonctions et* $\mathcal{C} = \{S_f : f \in \mathcal{F}\}$ *l'ensemble des sous-graphes des fonctions de* $\mathcal{F}$. *On dit que* $\mathcal{F}$ *est une VC-classe de fonctions si* $\mathcal{C}$ *est une VC-classe d'ensembles dans* $\mathcal{F} \times \mathbb{R}$. *On a l'indice VC de* $\mathcal{F}$, *noté* $\mathcal{V}(\mathcal{F}) = \mathcal{V}(\mathcal{C})$.

Le théorème suivant montre que les VC-classes de fonctions sont des classes à recouvrement polynomial dans la mesure où leur nombre de recouvrement est borné supérieurement par un polynôme en $(1/\varepsilon)$.

Théorème 6.2. *Soit* $\mathcal{F}$ *une VC-classe de fonctions d'indice* $\mathcal{V}(\mathcal{F}) = \nu$ *et* F *une enveloppe mesurable de* $\mathcal{F}$. *pour toute mesure de probabilité* Q *telle que* $0 < \|F\|_{r,Q} < \infty$, $r \geq 1$, *on a*

$$N(\varepsilon\|F\|_{r,Q}, \mathcal{F}, L_r(Q))d\varepsilon < K\nu(16e)^{\nu}\left(\frac{1}{\varepsilon}\right)^{r(\nu-1)},$$

où K *est une constante positive et* $0 < \varepsilon < 1$.

Preuve. (voir page 141, van der Vaart et Wellner (1996)).
Ce théorème permet de dire que toute VC-classe de fonctions est P-Donsker pour toute probabilité P si elle est P-mesurable et admet une enveloppe de carré intégrable.

La propriété de classe de Donsker est stable pour certaines opérations. Cela évite de calculer l'entropie à chaque fois qu'on doit montrer qu'une classe de fonctions $\mathcal{F}$ est Donsker.

Proposition 6.3. *Soit* $\mathcal{F}$ *et* $\mathcal{G}$ *deux classes de fonctions* P*-Donsker. Alors les classes suivantes sont aussi* P*-Donsker :*
i) $\mathcal{F} \cup \mathcal{G}$
ii) $\mathcal{F} + \mathcal{G} = \{f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$
iii) $\mathcal{F} \vee \mathcal{G} = \{f \vee g : f \in \mathcal{F}, g \in \mathcal{G}\}$ *et* $\mathcal{F} \wedge \mathcal{G} = \{f \wedge g : f \in \mathcal{F}, g \in \mathcal{G}\}$
iv) $\mathcal{F} \cdot \mathcal{G} = \{f \cdot g : f \in \mathcal{F}, g \in \mathcal{G}\}$ *si* $\mathcal{F}$ *et* $\mathcal{G}$ *sont uniformément bornées.*

v) $\phi \circ \mathcal{F} = \{\phi \circ f : f \in \mathcal{F}\}$ *si* ϕ *est une fonction réelle lipshitzienne et qu'il existe au moins* $f \in \mathcal{F}$ *telle que* $\|\phi \circ f\|_{2,P} < \infty$
vi) $\mathcal{F} \cdot g = \{f \cdot g : f \in \mathcal{F}\}$ *si* g *est une fonction mesurable bornée et* $\|P\|_{\mathcal{F}}$.

Inégalité exponentiellle et borne de moment

Soit $X_1, \cdots, X_n$ une suite i.i.d de variables aléatoires et $\varepsilon_1, \cdots, \varepsilon_n$ une suite indépendante de variables aléatoires de Rademacher indépendantes des X_i. L'inégalité suivante est due à Talagrand (1994).

Proposition 6.4. *Soit* $\mathcal{F}$ *une classe convenablement mesurable de fonctions* $f : \mathcal{X} \to \mathbb{R}$ *telle que pour tout* $g \in \mathcal{G}$, $\|f\|_\infty \leq M$*. Alors pour tout* $t > 0$*, on a*

$$\mathbb{P}\left\{\max_{1\leq m\leq n} \|\sqrt{m}\alpha_m\|_{\mathcal{F}} \leq A_1 \left(\mathbb{E}\|\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathcal{F}} + t\right)\right\} \leq 2\left[\exp(-\frac{A_2 t^2}{n\sigma_{\mathcal{F}}^2}) + \exp(-\frac{A_2 t}{M})\right],$$

où $A_1 > 0$ *et* $A_2 > 0$ *sont des constantes universelles et* $\sigma_{\mathcal{F}}^2 = \sup_{f\in\mathcal{F}} Var(f(X))$.

La borne de moment suivante est aussi crucial dans nos preuves. Elle est due à Einmahl et Mason (2005).

Proposition 6.5. *Soit* $\mathcal{F}$ *une classe convenablement mesurable de fonctions* $f : \mathcal{X} \to \mathbb{R}$ *bornées, admettant une fonction enveloppe* F *telle qu'il existe des constantes* $C, \nu > 1$ *et* $0 < \sigma < \beta$ *satisfaisant aux conditions suivantes :*
i) $\mathbb{E}[F^2(X)] \leq \beta^2$,
ii) $N(\varepsilon, \mathcal{F}) \leq C\varepsilon^{-\nu}, \quad 0 < \varepsilon < 1$
iii) $\sigma_0^2 := \sup_{f\in\mathcal{F}} \mathbb{E}(f^2(X)) \leq \sigma^2$,
iv) $\sup_{f\in\mathcal{F}} \|f\|_\infty \leq \frac{1}{4\sqrt{\nu}}\sqrt{n\sigma^2/\log(C_1\beta/\sigma)}$, *où* $C_1 = C^\nu \vee e$.
Alors

$$\mathbb{E}\|\sum_{i=1}^n \varepsilon_i f(X_i)\|_{\mathcal{F}} \leq A_3\sqrt{\nu n\sigma^2 \log(\beta \vee (1/\sigma))},$$

où $A_3 > 0$ *est une constante universelle.*

Proposition 6.6. *Soit* $\mathcal{F}$ *et* $\mathcal{G}$ *deux classes de fonctions réelles mesurables sur* $\mathcal{X}$ *telles que*

$$|f(x)| \leq F(x), \quad f \in \mathcal{F}, \quad x \in \mathcal{X},$$

où F est une fonction enveloppe finie mesurable sur $\mathcal{X}$ et

$$\|g\|_\infty \leq M, \quad g \in \mathcal{G},$$

où $M > 0$ est une constante finie. Si de plus, pour toute mesure de probabilité Q telle que $0 < Q(F^2) < \infty$, on a

$$N(\varepsilon\|F\|_{2,Q}, \mathcal{F}, L_2(Q)) \leq C_1\varepsilon^{-\nu_1}, \quad 0 < \varepsilon < 1$$

et

$$N(\varepsilon M, \mathcal{G}, L_2(Q)) \leq C_2\varepsilon^{-\nu_2}, \quad 0 < \varepsilon < 1$$

où ν_1, ν_2, C_1, C_2 sont des constantes positives appropriées. Alors, pour toute mesure de probabilité Q telle que $0 < Q(F^2) < \infty$, on a

$$N(\varepsilon M\|F\|_{2,Q}, \mathcal{F}\mathcal{G}, L_2(Q)) \leq C_3\varepsilon^{-\nu_1-\nu_2}, \quad 0 < \varepsilon < 1$$

avec $0 < C_3 < \infty$.

A.4 Delta-méthode

Soient D et E deux espaces topologiques métrisables. Une application $\phi : D_\phi \subset D \mapsto E$ est dite *Hadamard-différentiable* en $\theta \in D_\phi$ s'il existe une application linéaire continue $\phi'_\theta : D \mapsto E$ telle que

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi'_\theta(h),$$

pour toutes suites $t_n \to 0$ et $h_n \to h$ telles que $\theta + t_n h_n \in D_\phi$, pour tout n.
Soit $D_0 \subset D$, on dit que ϕ est *Hadamard-différentiable* en θ tangentiellement à D_0 si la suite h_n converge vers $h \in D_0$.

Théorème 6.3. *Soit $\phi : D_\phi \subset D \mapsto E$ une application Hadamard-différentiable en θ tangentiellement à D_0 ; $X_n : \Omega \mapsto D_\phi$ une suite d'applications telles que $r_n(X_n - \theta) \rightsquigarrow X$, où X est séparable et prend ses valeurs dans D_0. Alors $r_n(\phi(X_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(X)$. Si de plus, ϕ'_θ est définie et continue sur D, alors la suite $r_n(\phi(X_n) - \phi(\theta)) - \phi'_\theta(r_n(X_n - \theta))$ converge en probabilité (extérieure) vers 0.*

Conclusion et Perspectives

L'analyse de la pauvreté est une étape importante pour les pouvoirs publics nationaux et les organismes internationaux dans l'élaboration de leurs politiques de réduction de la pauvreté. Dans ce travail nous avons proposé un indice général de pauvreté qui inclut toutes les mesures de pauvreté basées sur le revenu et disponibles jusqu'ici dans la littérature. Nous avons d'abord établi la consistance, uniforme sur des classes de fonctions appropriées, d'un estimateur à noyau de cet indice général ; puis, un théorème central limite fonctionnel uniforme pour un estimateur de type "plug-in" de ce même indice général. Des études concluantes de simulations de données ont été réalisées pour illustrer ces résultats. Des applications ont également été fournies notamment, l'utilisation de la structure de covariance établie dans le théorème central limite fonctionnel, pour tester la dominance en temes de pauvreté entre deux distributions de revenu. Cette procédure d'inférence pour comparer deux distributions est applicable à toutes les mesures de pauvreté, décomposables comme non-décomposables (du type de Sen). Cependant, elle requiert l'utilisation d'un seuil de pauvreté fixe pour les distributions à comparer ; ce qui est un point de faiblesse pour ces résultats car le seuil de pauvreté peut dépendre de la distribution, en en étant une fonctionnelle par exemple.

Si G représente la distribution étudiée, alors le seuil de pauvreté pourrait s'écrire $z = z(G)$. Un estimateur de z est $z_n = z(G_n)$, où G_n est la distribution empirique associée à G. Donc une perspective intéressante est l'extension de ces résultats en faisant estimer z par une fonctionnelle de la distribution et en reconsidérant les estimateurs de l'indice général de pauvreté introduit dans cette thèse.

Bibliographie

[1] Atkinson, A.B. (1987). On the Measurement of Poverty. *Econometrica*, **55**, 749-764

[2] Barrett, G. F., and Donald, S. G. Consistent tests for stochastic dominance. *Econometrica*, **71**, 1 (2003), 71-104.

[3] Beach, C.M. and J. Richmond (1985). Joint Confidence Intervals for Income Shares and Lorenz Curves. *International Economic Review*, **26**, 439-450.

[4] Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

[5] Bishop, J.A., Formby, J.P. and Thistle, P. (1992). The convergence of South and non-South income distribution, 1969-1979. *American Economic review.* **82**, 262-272.

[6] Bishop, J.A., Formby, J.P. and Zheng, B. (1997). Statistical inference and the Sen index of poverty. *International Economic review.* Vol.38, N°2, pp. 381-387.

[7] Blondin, D. (2004). *Lois limites uniformes et estimation non-paramétrique de la régression.* Thèse, Université Paris 6, France.

[8] Bosq, D. et J.-P. (1987). *Théorie de l'Estimation Fonctionnelle.* Economica, Paris.

[9] Chakravarty, S.R. (1983). A New Index of Poverty. *Mathematical Social Sciences*, **6**, 307-313.

[10] Chakravarty, S.R. (1990). *Ethical Social Index Numbers.* Springer, New York.

[11] Chung, K. L. (1948). On the maximum partial sum of sequences of independent random variables. *Trans. Am. Math. Soc.*, **67**, 36-50.

[12] Davidson, R., and Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, **68**, 6 , 1435-1464.

[13] Deheuvels, P.(2000). Uniform limit laws for Kernel density estimators on possibly unbounded intervals. In *Recent Advances in Reliability Theory : Methodology, Practice and Inference.* (N.Limnios and M.Nikulin,eds.) , 477-492. Birkhäuser, Boston.

[14] Deheuvels, P. et Mason, D. M. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Probab.*, **20**, 1248-1287.

[15] Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.* **7**, 225-277.

[16] Devroye, L. (1987). *A course in Density Estimation.*Birkhäuser-Verlag, Boston.

[17] Dia, G. (2008).Estimation nonparamétrique de la distribution des revenus et de l'indice de pauvreté, *C.R. Acad. Sci. Paris, ser. I 346(2008).*

[18] Dony, J. (2008). *Nonparametric regression estimation : An empirical process approach to the uniform in bandwidth consistency of kernel-type estimators and conditional U-statistics.* Ph.D. Thesis, Vrije Universiteit Brussel, Belgium.

[19] Dony, J., Einmahl, U. and Mason, D.M. (2006). Uniform in bandwidth consistency of local polynomial regression function estimators. *Austr. J. Stat.* **35,** 105-120.

[20] Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Prob.* **13** , 1-37.

[21] Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.* **33,** 1380-1403.

[22] Foster, J.E. and Shorrocks, A.F. (1988). Poverty Orderings and Welfare Dominance. *Social Choice Welfare*, **5**, 179-198.

[23] Foster, J. (1984). *On Economic poverty : a survey of aggregate measures.* In : Basmann, R.L., Rhodes, G.F. (Eds.), Advances in Econometrics, Vol. 3. JAI Press, Connecticut.

[24] Fisher, G., Willson, D. and Xu, K. (1998). An empirical analysis of term premium using significance tests for stochastique dominance. *Economic Letters.* **60**, 195-203.

[25] Foster, J. E., Greer, J. and Thorbecke, E. (1984). A class of decomposable Poverty Measures, *Econometrica*, **vol.52**, N°3, 761-766.

[26] Gouriéroux, CH., Holly, A. and Monfort, A. (1982). Likelihood Ratio Test, Wald Test and Kuhn-Tucker test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica* **50**, 63-80.

[27] Härdle, W., Janssen, P. and Serfing, R. (1988). Strong uniform consistency rates of estimators of conditional functionals. *Ann. Statist.*, **16.4**, 1428-1449.

[28] Howes, S. (1994).Testing for dominance : inferring population rankings from sample data. World Bank discusion paper.

[29] Kodde, D. A. and Palm, F. C. (1986). Wald criteria for jointly testing equaliy and inequality restrictions. *Econometrica* **50**, 1243-1248.

[30] Komlós, I., Major, P., Tusnády, G. (1975). An approximation of partials sums of independent random variable and the sample distribution function. *Z. Wahrsch. Verw. Gebiete*, **32**, 111-131.

[31] Kosorok, M. R. (2006). *Introduction to Empirical Processes and Semiparametric In- ference.* Springer, New York.

[32] Kudô, A. (1963). A Multivariate Analogue of the One-Sided-Test. *Biometrika* **50**, 403-418.

[33] Lo, G.S. (2003). Estimation des lois du revenu et des dépenses dans une base de données de pauvreté. *Publications de l'UFR-SAT, lerstad, N°8, UGB, Sénégal.*

[34] Lo, G.S. and Seck, C.T. (2009). Uniform convergence of the non-weighted poverty measures.*Communications in Statistics - Theory and Methods.* 38, 3697-3704.

[35] Lo, G.S., Sall, S.T. and Seck, C.T. (2009). Une théorie asymptotique générale des mesures de pauvreté. *C.R. Math. Rep. Acad. Sci. Canada.* Vol. 31(2) 2009, pp 45-52.

[36] Maddala, G., and Singh, S. (1976). A function for size distribution of incomes. *Econometrica*, **44**.

[37] Mason, D.M. and Swanepoel, J. : A general result on the uniform in bandwidth consistency of kernel-type function estimators. To appear.

[38] Mason, D.M. : Proving consistency of non-standard kernel estimators. To appear.

[39] McDonald, J.B. (1984). Some generalized functions for the size distribution of income.*Econometrica,* vol.52, pp.647-663.

[40] Nadaraya, E.A. (1964). On estimating regression. *Theor. Probab.Appl.* **9**, 141-142.

[41] Nadaraya, E.A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves.* Kluwer Academic publishers, Dordrecht. (1961), 649-660.

[42] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**, 1065-1076.

[43] Perlman, M. D. (1969). One Sided Problems in Multivariate Analysis. *Annals of Mathematical Statistics*, **40**, 549-567.

[44] Pollard, D. (1984). *Convergence of Sthochastic Process,* Springer Verlag, New York.

[45] Rongve, I. (1997). Statistical Inference for Poverty Indices with fixed Poverty lines. *Applied Economics* **29**, 387-392.

[46] Sen, A.K. (1976). Poverty : An ordinal approach to measurement. *Econometrica* **44**, 219-231.

[47] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[48] Shorrocks, A. (1995) : Revisiting the Sen Poverty Index.*Econometrica* **63**, 1225-1230.

[49] Silverman, B.W. (1978). Weak ande strong uniform consistency of the kernel estimate of a density function and its derivative. *Annal of Statistics* **6**, 177-184.

[50] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

[51] Stute, W. (1982a). The oscillation behavior of empirical processes. *The Annals of Probability*, **Vol.10**, N°1, 414-422.

[52] Stute, W. (1982b). A law of the logaritm for kernel density estimators. *The Annals of Probability*, **Vol.10**, N°1, 86-107.

[53] Talagrand, M. (1994). Sharper bounds for Gaussian process and empirical processes. *Ann. Probab.* **22**, 28-76.

[54] Thon, D. (1979). On Measuring Poverty. *Review of Income and Wealth*, **25**, 429-439.

[55] van de Geer, S. (2000). *Empirical Processes in M-estimation.* Cambridge Univerity Press, New York.

[56] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes,* Springer Verlag, New-York.

[57] van der Vaart, A. W., and Wellner, J.A. (2007). Empirical processes indexed by estimated functions. *Asymptotics : Particules, Processes and Inverser Problems*, **55**, 234-252.

[58] Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

[59] Wolak, F. A. (1989). Testing inequality constraints in linear econometric models. *Journal of Econometrics* **41** 205-235.

[60] Zheng, B. (1997). Aggregate Poverty Measures. *Journal of Economic Surveys* **11**(2), 123-162.

[61] Zheng, B. (2001). Statistical inference for poverty measures with relative poverty lines. *Journal of Econometrics*, **101**, 337-356.

[62] Zitikis, R. (2002). Large sample estimation of a family of economic inequality indices.*Pakistan Journal of Statistics*(Special Issue in Honour of Dr. S.Ejaz Ahmed),**28**, 305-317.