



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

جامعة سعد دحلب البليدة-1-
Université Saad Dahlab Blida -1-



Mémoire de fin D'études

En vue de l'obtention du diplôme Master

Faculté des Sciences
Département : Mathématiques

Spécialité : Modélisation stochastique et statistique

THEME

**Estimation de type noyau de l'indice de
pauvreté**

Présenté par :
BOUDJELLAL Nesrine
CHIKHAOUI Fatima

Soutenu le :

Devant le Jury :
Président : O.TAMI MCB **Université de Blida 1**

Promoteur: A. RASSOUL Prof. **ENSH de Blida**

Examineur : R.FRIHI MAA **Université de Blida 1**

Promotion : 2020/2021



REMERCIEMENTS

Je tiens à remercier **ALLAH** pour la volonté, la santé, et la patience.

Je tien à dresser mes remerciement à mon encadreur **Abdelaziz Rassoul** de m'avoir appris le sens de la responsabilité et de travail toujours bien fait et sur tout pour les connaissances, ces conseils judicieux et son apport précieux tout au long de la réalisation de ce mémoire et pour leur disponibilité.

Nous adressons également nos remerciements à tous les membres de jury,

Nous sommes très reconnaissant à leurs remarques et commentaires qui nous aidions beaucoup pour mieux présenter ce document.

Grand remerciement aux mes professeurs **Redouane Frihi** et **Tami Omar** pour son aide.

Mes remerciements vont aussi à tous mes professeurs, enseignants et toutes les personnes qui m.ont soutenu jusqu.au bout, et qui n'ont pas cessé de me donner des Conseils très importants en signe de reconnaissance.



DEDICACE

je dédie ce modeste travail

A mes parents, mon estime pour eux est immense, je vous remercie pour tout ce

que vous avez fait pour moi. Que Dieu vous préserve une longue vie heureuse.

Avec toute ma tendresse.

Pour vous, mon chère fiancé : Hamza

Pour vous, mes chères frères : Bilal, Ibrahim, Youcef

*Pour vous, mes chères soeurs : Chahira, Meriem,
Nora, Aïcha*

Pour vous, mon beau frère : Amrane

Pour vous : Grand-père

*Aux personnes que je n'oublierais jamais mes amies :
Khadidja, Nesrine, Sara, Hanadi*

pour son soutien moral et sa présence dans bons et les mauvais moments.

A tous les membres de ma promotion.

A tous mes professeurs

FATIMA



DEDICACE

*A la lumière de ma vie, mes très chers parents qui
ont toujours été à mes côtés pour tous leurs sacrifices
leur amour, leur tendresse, leur soutien et leurs
prières tout au long de mes études,
A tous ceux qui m'ont enseigné et surtout à monsieur*

A. Rassoul

A mes adorables sœurs Asma, Assia, Sara

A mes frères Oussama, Mohamed

A ma grande famille, mes amis et collègues

A mon binôme Fatima, ainsi que sa famille

A tous mes enseignants du primaire à l'université

A tous ce que j'aime et que je respecte

Nesrine

الملخص:

هذا العمل مخصص لدراسة توصيف وتقدير مؤشر الفقر الذي نهتم به في التقدير غير المعياري و البارامترية للكثافة الاحتمالية نقترح مقدرًا لمؤشر الفقر في فستر وجريير وثوربيك ، تم إنشاء المقدر باستخدام نواة Parzen-Rosenblatt التكيفية. التقارب المنتظم شبه مؤكد والتقارب المنتظم في متوسط الخطأ التريبيعي. ثم نقدم دراسة أداء لهذا المقدر، على بيانات محاكاة، مقارنة بالمقدر من النواة غير قابل للتكيف وللمقدر التجريبي. أظهرت الدراسة أن المقدر يوصى باستخدام نواة تكيفيه. لتقدير الكثافة الاحتمالية للبيانات الموجبة بالطريقة بالنسبة للنواة ، يعد اختيار النواة ومعامل التنعيم h أمرًا مهمًا. ثم نقوم بإجراء عمليات المحاكاة باستخدام برنامج R الذي يسمح بحساب مؤشر الفقر وجذر متوسط الخطأ التريبيعي بواسطة القانون العادي وقانون جاما.

Résumé:

Ce travail est consacré à une étude de la caractérisation et l'estimation de l'indice de pauvreté nous nous intéressons à l'estimation non paramétrique et paramétrique d'une densité de probabilité Nous proposons un estimateur de l'indice de pauvreté de Foster, Greer et Thorbecke, L'estimateur est construit à l'aide du noyau adaptatif de Parzen-Rosenblatt. La convergence uniforme presque sure et la convergence uniforme en erreur quadratique moyenne sont établies. Nous fournissons ensuite une étude de performance de cet estimateur, sur des données simulées, comparativement à l'estimateur provenant du noyau non adaptatif et à l'estimateur empirique. L'étude montre que estimateur à noyau adaptatif est recommandé. Pour estimer la densité de probabilité des données positives par la méthode du noyau, le choix du noyau et du paramètre de lissage h est important. Ensuite, nous réalisons des simulations par utilisant le logiciel R qui nous permet de calculé l'indice de pauvreté et l'erreur quadratique moyenne par la loi normal et la loi gamma.

Summary:

This work is devoted to a study of the characterization and estimation of the poverty index we are interested in the non-parametric estimate and parametric of a probability density We propose an estimator of the poverty index of Foster, Greer and Thorbecke, The estimator is constructed using the Parzen-Rosenblatt adaptive kernel. Uniform convergence almost sure and the uniform convergence in mean squared error are We then provide a performance study of this estimator, on simulated data, compared to the estimator from the kernel non-adaptive and to the empirical estimator. The study shows that estimator at adaptive kernel is recommended. To estimate the probability density of positive data by the method of the kernel, the choice of the kernel and the smoothing parameter h is important. Then, we carry out simulations by using the R software which we allows to calculate the poverty index and the root mean square error by the normal law and gamma law.

TABLE DES MATIÈRES

Introduction Générale	3
1 Généralités sur les indices de pauvreté	5
1.1 Introduction	5
1.2 Seuil de pauvreté	6
1.2.1 Le choix du seuil de pauvreté ou l'identification des pauvres	7
1.2.2 Pauvreté absolue et pauvreté relative	7
1.2.2.1 Seuils de pauvreté absolue	8
1.2.2.2 Seuil de pauvreté relative	8
1.3 Agrégation de la pauvreté	9
1.3.1 Axiome de monotonie	10
1.3.2 Axiome de transfert	10
1.4 Exemples de mesures agrégées	11
1.4.1 Le choix de la mesure de la pauvreté ou l'agrégation des pauvres	13
1.5 L'incidence de la pauvreté (l'indice H)	13
1.6 Le déficit moyen du revenu des pauvres (l'indice I)	14
1.7 L'indice de Sen (1976)	14
1.8 L'indice de Kakwani (1980) :	15
1.9 L'indice de Foster, Greer et Thorbecke (1984)	16
2 Estimateur non paramétrique par la méthode de noyau	19
2.1 Introduction	19
2.2 Estimation par histogramme	19
2.2.1 Propriétés de l'histogramme	20

2.3	Estimation non-paramétrique d'une densité de probabilité à noyau . . .	21
2.3.1	Quelques définitions	21
2.3.2	Expression de la variance et du le biais d'un estimateurs à noyau	23
2.3.2.1	L'espérance	23
2.3.2.2	Le biais	24
2.3.2.3	La variance	25
2.3.2.4	Choix du paramètre de lissage	26
2.3.3	Estimation de la densité de probabilité par la méthode du noyau	27
2.3.4	Noyaux usuels	27
2.3.5	Propriétés d'un estimateur à noyau	27
2.3.6	Expression asymptotique du biais et de la variance	28
2.3.7	Choix théoriques optimaux du paramètre de lissage	29
2.3.8	La construction d'un estimateur à noyau	30
2.3.8.1	Propriété :	31
2.3.8.2	Les noyaux usuel :	31
3	Estimation non paramétrique de la répartition des revenus de l'indice de pauvreté	33
3.1	Introduction et définition de l'estimateur	33
3.2	Estimation empirique de l'indice FGT	33
3.3	Estimation non paramétrique de FGT	34
3.4	Estimation non paramétrique avec noyau adapté	34
3.5	Estimation non paramétrique sans biais	35
3.6	Principaux résultats	36
3.6.1	Cohérence et comportement uniformes presque sûrs du biais . .	36
3.6.2	Consistance quadratique moyenne Uniforme	38
3.7	Détails des preuves	39
3.7.1	Construction de l'estimateur	39
3.7.2	Preuves des principaux résultats	40
3.7.2.1	Preuve du lemme 3.1	40
3.7.2.2	Preuve du Lemme 3.2	44
3.7.2.3	Preuve du Théorème 3.1 et du Théorème 3.2	45
3.7.3	La consistance moyenne quadratique uniforme	46
3.7.3.1	Preuve Lemme 3.3	46
3.7.3.2	Preuve du Théorème 3.5	50

3.7.3.3	Preuve du théorème 3.3	54
3.8	Simulations et Applications	56
3.8.1	Simulations	56
Conclusion		59

TABLE DES FIGURES

2.1 Allures des noyaux : Triangulaire, Biweight, Gaussien, Epanechnikov. . .	32
--	----

LISTE DES TABLEAUX

2.1	Quelques noyaux usuels	27
3.1	Résultats de la simulation et estimation de l'indice de pauvreté pour loi normal	57
3.2	Résultats de la simulation et estimation de l'indice de pauvreté pour loi Gamma	58

12 juillet 2021

ABRÉVIATIONS ET NOTATIONS

- $K(u)$: noyau
- α : Proportionnel à.
- MSE : L'erreur quadratique moyenne :
- $MISE$: L'erreur quadratique moyenne intégrée.
- $E(.)$: espérance mathématique.
- Var : variance.
- X_1, \dots, X_n : échantillanant de taille n
- F : fonction de répartition.
- f : densité de probabilité.
- $v.a$: variable aléatoire
- $i.i.d$: indépendantes et identiquement distribuées.
- FGT : Foster, Greer et THorebecke
- z : le seuil de pauvreté

La théorie de l'estimation est une des préoccupations majeures des statisticiens. On trouve dans la littérature deux types d'approches d'estimations de la densité de probabilité : l'approche paramétrique et l'approche non-paramétrique. L'approche paramétrique a comme inconvénient principal de nécessiter une connaissance préalable du phénomène aléatoire considéré. L'approche non paramétrique estime la densité de probabilité directement à partir de l'information disponible sur l'ensemble des observations.

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

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

et 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.

Dia (2008) a également étudié les propriétés asymptotiques de cet estimateur à noyau, notamment les convergences en moyenne quadratique et presque sure uniformément en z sur des intervalles compacts.

L'estimation à noyau (ou encore méthode de Parzen, Rosenblatt) est une méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire.. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. Cet estimateur sont des fonctions de deux paramètres K , appelé noyau, et h dit paramètre de lissage (largeur de fenêtre). Si le choix du noyau n'est pas un problème dans l'estimation de la densité, il n'en est pas de même pour le choix de paramètre de lissage qui ne dépend que de la taille n de l'échantillon. Plusieurs travaux ont montré que les estimateurs peuvent changer dramatiquement pour de petites variations du paramètre de lissage. Actuellement il n'existe pas de choix optimal pour ce paramètre, le choix optimal qui minimise l'erreur quadratique

moyenne intégrée dépend de la dérivée seconde de la densité inconnue.

L'objectif de ce travail est d'utiliser l'approche Bayésienne pour le choix du paramètre de lissage h dans l'estimation de la fonction densité de probabilité par la méthode des noyaux de l'indice de pauvreté.

Ce mémoire est composé d'une introduction, de trois chapitres et d'une conclusion. 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 mesure de pauvreté discrètes ainsi que quelques axiomes fondamentaux qui les régissent.

Dans le deuxième chapitre, nous présentons l'estimation non paramétrique de la densité de probabilité. Après la présentation de la définition de l'estimateur à noyau de cette fonction de densité et ses propriétés et expression (biais, variances,...), nous avons abordé le problème du choix du noyau et du paramètre de lissage par les techniques classiques.

Le troisième chapitre est consacré à l'estimation empirique et non paramétrique de l'indice de FGT. Nous exposerons notre premier résultat qui est la consistance forte uniforme d'un estimateur général à noyau pour l'indice de pauvreté. Cet estimateur sera construit avec le noyau de Parzen-Rosenblatt et le noyau adapté, à la fin de ce chapitre nous regrouperons les résultats de simulation des différentes méthodes de sélection du paramètre de lissage et de l'indice de pauvreté. Tous les résultats numériques sont effectués à l'aide du logiciel R.

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 agrégés 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 multidimensionnelle 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 unidimensionnelle de la pauvreté.

De prime abord, l'étude unidimensionnelle peut paraître plus limitée, moins complète et donc moins pertinente qu'une étude multidimensionnelle ; 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 multidimensionnelle. 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 ce mémoire, nous intéressons aux mesures de pauvreté unidimensionnelles basées sur le revenu ou les dépenses de consommation ; mais nous utiliserons le terme générique de revenu pour faire allusion aux deux caractères. Cette approche quantitative basée sur le revenu, notons-le, est la plus fréquente dans la littérature et utilise 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 nourrir, 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.

1.2.1 Le choix du seuil de pauvreté ou l'identification des pauvres

L'identification des pauvres revient à fixer un niveau de vie minimum en dessous duquel un individu est considéré comme pauvre. Plusieurs méthodes de construction des lignes de pauvreté sont possibles. Le choix de la méthode dépend fondamentalement du concept de pauvreté utilisé. Ravallion (1996) présente un aperçu critique des différentes approches de construction des lignes de pauvreté. Cependant, nous ne pouvons parler de la construction des seuils de pauvreté sans évoquer les deux conceptions de la pauvreté les plus utilisées dans les travaux empiriques à savoir la définition absolue et la définition relative de la pauvreté.

1.2.2 Pauvreté absolue et pauvreté relative

La pauvreté absolue renvoie à un concept absolu selon lequel la pauvreté d'un individu ne dépend que de sa situation et non de celle du reste de la société.

Sous cet angle, un individu est dit pauvre s'il n'est pas en mesure de satisfaire les besoins élémentaires de base tels que le fait de se nourrir, 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 très adéquate pour les pays sous-développés, où la pauvreté peut impliquer des pénuries alimentaires. 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.

Concernant la pauvreté relative, l'individu est jugé par rapport au reste de la société. Ainsi, un individu est considéré comme pauvre, non par ce qu'il n'a pas un certain niveau de vie donné, mais par ce que son niveau de vie est très bas si on le compare à ceux des autres membres de la société. Cette caractérisation de la pauvreté est surtout utilisée pour les pays développés, puisque les individus 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 ou médian du pays. Le seuil de pauvreté est choisi comme un fractile de la distribution des revenus. Généralement, on prend un pourcentage du revenu médian ou du revenu moyen de la distribution.

1.2.2.1 Seuils de pauvreté absolue

La méthode la plus utilisée pour définir un seuil de pauvreté absolu consiste à identifier certains besoins de consommation de base, jugés pertinents dans le domaine des comparaisons de la pauvreté. Le besoin fondamental le plus important renvoie naturellement aux dépenses alimentaires nécessaires pour assurer la consommation d'énergie nutritive recommandée viennent ensuite les dépenses au titre des biens non alimentaires. Dans cette perspective, il existe essentiellement deux méthodes pour fixer les seuils de pauvreté : la méthode de l'énergie nutritive et la méthode du coût des besoins de base. On présente ici un aperçu des deux méthodes qui ont suscité l'intérêt dans plusieurs travaux.¹

1.2.2.2 Seuil de pauvreté relative

La notion de pauvreté absolue se retrouve à travers les travaux consacrés aux pays en développement tandis que les études portant sur les pays développés s'intéressent à la pauvreté relative et considèrent parfois que la pauvreté est même un phénomène entièrement relatif. Cela s'explique par le fait que, dans les pays développés, il s'agit davantage d'un problème d'inégalité dans la mesure où implicitement il est admis que les besoins essentiels fondamentaux sont supposés satisfaits.

Pour déterminer un seuil de pauvreté relative, la méthode la plus fréquente consiste à fixer celui-ci à une proportion de la moyenne arithmétique ou de la médiane de la distribution de la consommation ou du revenu. De nombreux travaux se réfèrent au seuil de Fuchs (1967) qui fixe un seuil de pauvreté qui correspond à environ 50% de la médiane nationale. Il est possible de se référer à d'autres indicateurs : nous pouvons prendre la demi-moyenne plutôt que la demi-médiane. Choisir la moyenne permet de comparer les ménages pauvres à tout le reste de la distribution, alors que la médiane

1. Voir par exemple Ravallion (1994)

renvoie à une comparaison avec seulement les ménages dont la position dans l'échelle des revenus est intermédiaire. En effet, rajouter à la distribution quelques ménages aux revenus très élevés, ne change pas la valeur de la médiane, ni par conséquent le seuil de pauvreté. En revanche, la moyenne serait sensible à l'introduction de quelques ménages très riches dans la population. Qui découle du fait qu'un groupe de la population a un niveau de revenu inférieur au seuil de pauvreté.

Plusieurs travaux se sont intéressés à ce sujet et qui ont mis à la disposition des utilisateurs une large gamme de mesures présentant des caractéristiques intéressantes pour la détermination d'un profil de pauvreté. Nous présentons dans ce paragraphe les principales mesures de pauvreté qui ont été proposées dans la littérature économique.

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, \dots, y_n , où $n \geq 1$ est un entier-naturel. Cette distribution peut être représentée par un vecteur $y = (y_1, \dots, y_n)$, avec $y_j \in \mathbb{R}_+ = [0, \infty)$, $\forall j = 1, \dots, n$. y_j est le revenu d'un individu j qui est classé comme pauvre si $y_j < z$. Ainsi, $Y = \{y \in \mathbb{R}_+^n, n \geq 1\}$ est l'ensemble des distributions de revenu définies sur \mathbb{R}_+ . Pour toute distribution $y \in Y$, 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. $y_1 \leq y_2 \leq \dots \leq y_n$.

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

$$P : Y \times \mathbb{R}_+ \mapsto [0, 1]$$

$$(y, z) \mapsto p(y, z)$$

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 célèbre 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

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

Ce type d'indice est également critiqué car il est 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 $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ deux distributions de revenu, et $z > 0$ le seuil de pauvreté commun aux deux distributions.

1.3.1 Axiome de monotonie

Une réduction du revenu d'un individu pauvre doit accroître la mesure de pauvreté i.e.

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

1.3.2 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

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

Fort de ces trois axiomes et d'autres encore 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.

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

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

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

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 $\mu_p = \sum_{j=1}^q y_j$ 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 parmi les quelles, on peut citer la célèbre famille d'indices de Foster-Greer-Thorbecke (FGT) (1984). Ces indices sont définis pour $\alpha \geq 0$, par

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

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

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

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

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

Ces mesures non-pondérées sont généralement de la forme

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

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

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

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 :

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

Tandis que Shorrocks (1995) introduisit la mesure

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

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

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

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 considère les revenus (y_1, \dots, y_n) comme une réalisation d'un échantillon aléatoire (Y_1, \dots, 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.

1.4.1 Le choix de la mesure de la pauvreté ou l'agrégation des pauvres

Après l'étape d'identification reposant sur le choix de la ligne de pauvreté permettant ainsi d'identifier qui est pauvre, il y a l'étape d'agrégation laquelle consiste à agréger l'information et à obtenir une mesure qui va refléter des considérations éthiques de l'évaluateur social. Atkinson (1987) suggère que cette mesure peut être interprétée comme étant la perte sociale qui découle du fait qu'un groupe de la population a un niveau de revenu inférieur au seuil de pauvreté. Plusieurs travaux se sont intéressés à ce sujet et qui ont mis à la disposition des utilisateurs une large gamme de mesures présentant des caractéristiques intéressantes pour la détermination d'un profil de pauvreté. Nous présentons dans ce paragraphe les principales mesures de pauvreté qui ont été proposées dans la littérature économique.

1.5 L'incidence de la pauvreté (l'indice H)

C'est l'indice le plus simple qui correspond à la proportion de pauvres par rapport à la population totale. Le terme d'incidence est le plus fréquemment retenu pour désigner cette mesure :

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

Avec q qui correspond au nombre de pauvres, n est la population totale, z représente la ligne de pauvreté et y est le revenu.

Cet indice considère que tous les individus sont dans la même situation, dès lors qu'ils ont un revenu inférieur à la ligne de pauvreté. Il attribue la même importance à une personne ayant un revenu juste inférieur à la ligne de pauvreté, qu'à une personne dont le revenu est presque nul. L'indicateur H ne s'intéresse qu'au nombre de pauvres. Supposons qu'un individu pauvre voit son revenu diminuer, l'indicateur H ne se modifiera pas. En d'autres termes, la mesure d'incidence n'apporte aucune information

sur l'intensité de la pauvreté. Pour en rendre compte, il est usuel de mesurer l'écart relatif moyen au seuil de pauvreté.

1.6 Le déficit moyen du revenu des pauvres (l'indice I)

Cet indice représente la moyenne des déficits de revenu des pauvres, exprimée en pourcentage de la ligne de pauvreté. Pour l'estimer, il nous faut d'abord, calculer les déficits individuels de revenu (g_i), qui correspond à la différence entre la ligne de pauvreté et leur revenus, soit :

$$g_i = z - y_i$$

Où y_j est le revenu individuel. On suppose que les q pauvres sont ordonnés par ordre croissant de revenu, ainsi

$$y_1 \leq y_2 \leq y_3 \dots$$

L'indice I est défini à partir de la relation suivante :

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

Le déficit moyen du revenu des pauvres ne dépend que du revenu moyen des pauvres et du seuil de pauvreté. A une même valeur de I , il peut donc correspondre différentes dispersions du revenu parmi les pauvres. H est donc utile de retenir un indice qui tient compte de l'inégalité de revenu parmi les pauvres. Sur ce dernier aspect, l'indice de Sen (1976) permet de répondre à cette préoccupation.

1.7 L'indice de Sen (1976)

L'apport de l'indice de Sen (1976) est de tenir compte simultanément de la proportion des pauvres, de l'intensité de la pauvreté et de l'inégalité de répartition des revenus parmi les pauvres.

Ayant mis en lumière les limites des indices précédents, Sen (1976) se propose de développer une approche axiomatique sur la mesure de la pauvreté et de définir un certain nombre d'axiomes qu'un bon indice doit satisfaire. Ces axiomes traduisent des principes éthiques et moraux que devrait avoir la société vis-à-vis des pauvres. Ensuite, un indice de pauvreté est établi sur la base de ces conditions.

Selon Sen (1976), les deux principaux axiomes que devrait satisfaire tout indice de pauvreté sont les suivants :

- *Axiome de monotonie* : Toutes choses égales par ailleurs, une réduction du revenu d'une personne qui se trouve en dessous de la ligne de pauvreté doit aug-

menter la valeur de l'indice de pauvreté» Cet axiome peut nous paraître évident et incontestable ; si une personne devient plus pauvre, la pauvreté agrégée sera plus grande. Mais il est à noter, que l'incidence H ne satisfait pas cet axiome.»

- *Axiome de transfert* : Toutes choses égales par ailleurs, un transfert de revenu d'une personne pauvre vers une autre personne plus riche (transfert régressif) ne doit pas diminuer la valeur de l'indice de pauvreté.» Ce second axiome est très important. Il établit que les personnes plus pauvres devraient avoir une plus grande importance dans l'agrégation de la pauvreté. En effet, un transfert de revenu d'une personne pauvre vers une autre moins pauvre modifie seulement la répartition du revenu et non pas le revenu total du pays.

Sur la base de cet axiome, Sen introduit l'idée que les personnes se trouvant dans une situation de pauvreté plus grave devraient être aidées en priorité. Sen a montré que la seule mesure compatible avec les deux axiomes et donnée :

$$s = H(I + \frac{q}{q+1}(1-I)G) = \frac{2}{(q+1)nz} \sum_{i=1}^q (q+1-i)g_i$$

Où H et I sont, respectivement l'incidence de la pauvreté et le déficit moyen du revenu des pauvres, G est le coefficient de Gini, correspondant à l'inégalité dans la distribution du revenu parmi les pauvres, q est le nombre de pauvres classés par ordre croissant de revenu et g_i est les déficits individuels du revenu des pauvres

L'indice de Sen tient compte à la fois de la proportion des personnes en pauvreté, de leur déficit de revenu et des inégalités qui existent parmi eux. Cependant, cette mesure n'échappe pas à certaines faiblesses. En effet, l'un des inconvénients que l'on peut reprocher à l'indice de Sen est qu'il n'est pas décomposable².

1.8 L'indice de Kakwani (1980) :

Dans son approche, Kakwani (1980) propose une généralisation de l'idée de Sen et complète les deux axiomes de Sen par l'axiome de sensibilité aux transferts

« *Axiome de sensibilité aux transferts* : Toutes choses égales par ailleurs, un transfert régressif d'un montant w du i -ème individu pauvre vers le j -ème moins pauvre provoquera une plus grande augmentation de la mesure de pauvreté qu'un transfert régressif du même montant du k -ème individu pauvre vers le l -ème moins pauvre si $y_j - y_i = y_l - y_k \geq 0$ et si $y_k > y_i$.

Cet axiome implique que la pauvreté augmente avec un transfert régressif et que, plus les individus intervenant dans ce transfert sont pauvres, plus l'augmentation de niveau de pauvreté sera forte. La famille des indices vérifiant cet axiome est définie

2. Des critiques et avantages de l'indice de Sen peuvent être consultés dans Zheng (1997)

comme suit :

$$K = \left(\frac{q}{nz \sum_{i=1}^q i^k} \right) \sum_{i=1}^q (z - y_i)(q + 1 - i)$$

Nous remarquons que si K est égal à 1, nous obtenons l'indice de Sen. Lorsque K est supérieur à 1, alors l'indice devient sensible aux transferts.

1.9 L'indice de Foster, Greer et Thorbecke (1984)

Foster, Greer et Thorbecke (FGT) ont proposé en 1984 un indicateur de la pauvreté qui satisfait les deux axiomes de Sen, l'axiome de Kakwani, ainsi que l'axiome de décomposition additive qui est défini comme suit :

Axiome de décomposition : Soit une population composée de m groupes, chaque groupe contenant n_j individus ($j = 1, \dots, m$ et $n = \sum n_j$.)

Si on note P , la mesure de la pauvreté totale calculée sur l'ensemble de la population et P_j le niveau de pauvreté du j -ème groupe, alors :

$$P = \sum \left(\frac{n_j}{n} \right) P_j$$

Ce qui signifie que la pauvreté agrégée de toute la population correspond à la somme des niveaux de pauvreté de chaque groupe, pondérée par la part de chaque groupe dans la population totale.

Ainsi, Foster, Greer et Thorbecke ont défini une famille de mesures qui permet selon les différentes valeurs prises par le paramètre (α) de tenir compte de la situation des pauvres. Ce paramètre représente le degré d'aversion à la pauvreté. Cette famille de mesure s'exprime comme suit :

$$P_\alpha = \frac{1}{n} \sum_{i=1}^q \left(\frac{z - y_i}{z} \right)^\alpha$$

Selon les valeurs de α , cette famille d'indices permet de mesurer l'étendue ou incidence de la pauvreté ($\alpha = 0$), sa profondeur ($\alpha = 1$) et sa sévérité ($\alpha = 2$). Il présente aussi l'avantage d'être décomposable.

Généralement, nous utilisons pour α , des valeurs de 0, 1 et 2 pour mesurer la pauvreté.

- L'incidence de la pauvreté ($\alpha = 0$), L'étendue ou l'incidence de la pauvreté dénommée P_0 , est le ratio de pauvreté. C'est la proportion de pauvres dans l'ensemble de la population étudiée.
- La profondeur de la pauvreté ($\alpha = 1$), La profondeur moyenne de la pauvreté

correspond à $\alpha = 1$. Ainsi, nous avons :

$$P_1 = \frac{1}{n} \sum_{i=1}^q \left(\frac{z - y_j}{z} \right)$$

Appelé profondeur de la pauvreté, P_1 donne une idée de l'intensité de celle-ci.

- La sévérité de la pauvreté ou l'inégalité parmi les pauvres ($\alpha = 2$), Cet indice de sévérité, P_2 permet de mieux comprendre l'extrême pauvreté. Il est représenté par la formule suivante :

$$P_2 = \frac{1}{n} \sum_{i=1}^q \left(\frac{z - y_j}{z} \right)^2$$

L'indicateur P_2 pondère les écarts de pauvreté par eux-mêmes ce qui donne plus de poids aux revenus les plus faibles. C'est un indicateur de ciblage approprié pour réduire la pauvreté .

- Ces indices FGT sont décomposables pour toute valeur de α . Cela signifie que si la population est subdivisée en K sous-groupes, nous pouvons calculer les indices de pauvreté et la contribution à celle-ci de chaque sous-groupe.

En effet, supposons une partition de la population en j groupes (j variant de 1 à n) $P_{j\alpha}$ la valeur de P_α pour le groupe j et K_j est la proportion du groupe j dans la population totale telle que $\sum K_j = 1$, nous avons :

$$p_\alpha = \sum K_j P_{j\alpha}$$

et la contribution de chaque groupe à la pauvreté globale est donnée par la formule

$$c_j = \frac{K_j P_{j\alpha}}{p_\alpha}$$

Le calcul des contributions c_j permet notamment d'identifier les zones ou groupes socioéconomiques caractérisés par une forte concentration de la pauvreté et de fournir un moyen supplémentaire de ciblage des zones de pauvreté. En outre, la propriété d'additivité des indices FGT peut être exploitée pour analyser l'évolution de la pauvreté dans le temps entre deux dates ($t = 1, 2$) et dans l'espace en explorant les modifications de la pauvreté à l'intérieur des secteurs (variations intrasectorielles) et par les déplacements de population entre les secteurs (variations inter sectorielles). En vue d'étoffer l'analyse de la pauvreté, il est possible, d'expliquer la variation de la pauvreté par les modifications de pauvreté à l'intérieur des secteurs (variations intrasectorielles) et par les déplacements de population entre les secteurs (variations intersectorielles) et par les déplacements de population entre les secteurs (variations intersectorielles) [Ravallion, Huppi (1991)]. Etant donné P_α^t un indice FGT ($\alpha \geq 0$) à la date t et $v_j^t = n_j^t/n^t$, la

part dans la population totale du secteur j ($j = 1 \dots k$), il est possible d'exprimer la variation de la pauvreté entre les deux dates par :

$$P_{\alpha}^2 - P_{\alpha}^1 = \sum_{j=1}^k (P_{\alpha j}^2 - P_{\alpha j}^1) v_j^1 + \sum_{j=1}^k (v_j^2 - v_j^1) P_{\alpha j}^1 + \sum_{j=1}^k (P_{\alpha j}^2 - P_{\alpha j}^1) (v_j^2 - v_j^1)$$

- La première composante correspond aux **effets intra-sectoriels**. Elle traduit la contribution des changements de la pauvreté à l'intérieur de chaque secteur, lorsque les proportions de population des différents secteurs sont fixées à leur niveau initial (en $t = 1$).
- La deuxième composante, **l'effet des déplacements de population**, mesure l'impact des changements de la répartition de la population entre secteurs sur la pauvreté initiale.
- La troisième composante mesure **les effets d'interaction**, c'est-à-dire l'éventuelle corrélation entre les variations sectorielles de pauvreté et les mouvements de population, et leur signe nous indique si la population a tendance ou non à se déplacer vers les secteurs dans lesquels la pauvreté est en baisse.

Les indices de pauvreté que nous venons de présenter, dépendent des distributions de revenus et de considérations éthiques. Ces indices peuvent donner lieu à un ordre de préférences. Certaines mesures de pauvreté sont jugées plus complètes que d'autres car elles permettent d'exprimer des préférences entre des distributions que certaines mesures ne permettent pas.

Les indices d'incidence H vont proposer un classement très partiel des différentes distributions que l'on pourrait vouloir comparer, alors qu'un indice sensible à la distribution des pauvres permettrait d'obtenir un classement plus complet. Cependant, l'usage d'indices plus forts (plus complets) ne se fait pas sans accepter certains jugements de valeur.

Afin de se focaliser sur le problème d'agrégation de la pauvreté, nous avons posé que les comparaisons entre distributions s'effectuaient avec un seuil de pauvreté fixé arbitrairement mais aussi sur la base d'une mesure qui comporte une dose d'arbitraire dans la mesure où elle reflète des considérations éthiques. Lever cette hypothèse en considérant un ensemble de seuils qui englobe la totalité des revenus de la distribution, revient à étudier les différents niveaux de pauvreté de différentes distributions pour différents seuils de pauvreté. Cette démarche est en fait analogue à celle suivie dans le cadre de la dominance stochastique.

CHAPITRE 2

ESTIMATEUR NON PARAMÉTRIQUE PAR LA MÉTHODE DE NOYAU

2.1 Introduction

Dans de nombreuses applications, la densité f est inconnue et on dispose d'un n échantillon X_1, X_2, \dots, X_n de variables aléatoires indépendantes et identiquement distribuées admettant f comme densité. Le problème du statisticien consiste alors à utiliser cet échantillon pour construire un estimateur qui soit le plus proche possible de la densité f . Les premiers articles consacrés à ce sujet sont dus au biométricien Karl Pearson [Pearson], Rosenblatt [Rosenblatt(1956)] et Cencov [Cencov, 1962].

Plusieurs estimateurs de la densité de probabilité ont été proposés depuis les travaux de Rosenblatt [Rosenblatt(1956)], Cencov [Cencov, 1962] et Parzen [Parzen(1962)].

La grande majorité d'entre eux se rangent dans une classe très importante contribué par des estimateurs bâtis à partir d'un noyau (méthode de noyau et l'estimation par histogramme). Une autre classe d'estimateurs basée sur la notion de développement en séries de fonctions orthogonales, fut proposée par Cencov [Cencov, 1962]. Bien que moins étudiée que la précédente, elle semble actuellement connaître un nouvel essor (méthode des fonctions orthogonales).

Dans cette partie nous étudions en détail l'estimation de la densité de probabilité par des séries orthogonales.

2.2 Estimation par histogramme

Elle consiste à estimer la densité de la variable aléatoire X en x par n_i le nombre d'occurrences de réalisations x_i appartenant à la i ème classe associée à la valeur x .

Etant données des observations (x_1, x_2, \dots, x_n) qui sont les réalisations des variables

aléatoires réelles indépendantes et identiquement distribuées X_1, X_2, \dots, X_n de densité f inconnue sur un intervalles $[a; b]$. Pour le faire de manière non paramétrique, il est naturel de se donner k intervalle (classe) $(I_j)_{j=0,1,\dots,k-1}$ avec $I_j = [a_{j-1}; a_j]$.

Pour construire l'histogramme nous devons choisir une origine x_0 et une largeur d'intervalle h . la largeur contrôle principalement la qualité de lissage. L'histogramme peut être généralisé en autorisant la largeur d'intervalle à varier. Supposons que la droite est coupée en intervalles, la densité estimée est :

$$\widehat{f}_h(x) = \frac{1}{n} \frac{\text{card} \{x_i \text{ dans le meme intervalle que } x\}}{\text{largeur de l'intervalle contenant } x} \quad (2.1)$$

l'estimateur de f sur $[a_j; a_{j+1}]$ est définie par :

$$\begin{aligned} \widehat{f}_h(x) &= \frac{\text{card} \{x_i \leq a_{j+1}\} - \text{card} \{x_i \leq a_j\}}{nh} \\ &= \frac{n_j}{nh}, x \in]a_j, a_{j+1}]. \end{aligned} \quad (2.2)$$

la construction de $\widehat{f}_h(x)$ peut être faite même si les données ne proviennent pas d'une loi continue.

2.2.1 Propriétés de l'histogramme

Pour construire un bon histogramme, il faut trouver un juste équilibre entre le nombre d'observations n et le nombre de classes déterminé par k . Pour mesurer la précision de \widehat{f}_h en un point x fixé, on peut utiliser l'erreur quadratique moyenne

$$MSE(\widehat{f}_h(x)) = E[\widehat{f}_h(x) - f(x)]^2 \quad (2.3)$$

Il est cependant préférable de mesurer la précision de façon globale en calculant l'erreur quadratique moyenne intégrale :

$$MSE(\widehat{f}_h(x)) = \int MSE(\widehat{f}_h(x)) dx \quad (2.4)$$

d'après [Hall et al(1999)], on a :

$$n_j \rightarrow_{loi} \mathcal{B}(n, p_j)$$

où \mathcal{B} est une loi binomiale de paramètre n et $p_j = P(a_j < X \leq a_{j+1})$ pour $x \in [a_j, a_{j+1}]$, on déduit que

$$E[\widehat{f}_h(x)] = \frac{E(n_j)}{nh} = \frac{p_j}{h} \quad (2.5)$$

et

$$\text{Var}[\widehat{f}_h(x)] = \frac{p_j(1-p_j)}{nh^2} \quad (2.6)$$

L'expression de la variance montre que plus le paramètre h est petit, plus $\widehat{f}_h(x)$ est variable inversement, plus h est grand, moins $\widehat{f}_h(x)$ est variable.

En faisant le développement de Taylor à l'ordre 1, on obtient les propriétés suivantes [Bosq et Lecoutre, 1967] :

1. Biais :

$$\text{biais}(\widehat{f}_h(x)) = \frac{1}{2}f'(x)[h - 2(x - a_j)] + O(h^2). \quad (2.7)$$

2. Variance

$$\text{Var}(\widehat{f}_h(x)) = \frac{f(x)}{nh} - O(n^{-1}). \quad (2.8)$$

D'après la relation (2.7), on peut constater que si $h \rightarrow 0$ entraîne que le biais $(\widehat{f}_h(x)) \rightarrow 0$.

D'après la relation (2.8), pour que $\widehat{f}_h(x)$ soit peu variable, il faut que $nh \rightarrow \infty$.

$$\text{MSE}(\widehat{f}_h(x)) = \frac{f(x)}{nh} + \frac{f''(x)}{4} [h - 2(x - a_j)]^2 + O(h^3) + O(n^{-1}).$$

$$\text{MISE}(f_h(x)) = \frac{1}{nh} + \frac{h^2 \int f''^2(x) dt}{12} + O(h^3) + O(n^{-1}). \quad (2.9)$$

Dans la relation (2.9), on voit que le paramètre de lissage h est relié directement au terme $\frac{h^2 \int f''^2(x) dt}{12}$, provenant du carré de biais intégrée et que ce paramètre est inversement proportionnel à la variance intégrée.

Un petit h donne un histogramme peu biaisé, tandis qu'un grand h et un grand n déterminent un histogramme moins variable.

2.3 Estimation non-paramétrique d'une densité de probabilité à noyau

2.3.1 Quelques définitions

Dans cette section nous allons nous intéresser à l'estimateur à noyau symétrique. Avant d'exposer ses différentes composantes nous présenterons d'abord sa construction d'origine qui se fait à travers de l'estimateur empirique d'une fonction de répartition..

Définition 2.1 On observe X_1, X_2, \dots, X_n variables aléatoires réelles indépendantes et identi-

quement distribué (i.i.d) de fonction de répartition

$$F : x \rightarrow F(x) = p(X_1 \leq x)$$

Définition 2.2 L'estimateur de la fonction de répartition F est la fonction de répartition empirique notée F_n et définie par

$$F_n(x) = \widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_n \leq x\}} = \begin{cases} 0, & \text{si } x \leq X_1 \\ \frac{k}{n}, & \text{si } X_k \leq x < X_{k+1} \\ 1, & \text{si } x \geq X_n \end{cases} .$$

C'est un estimateur non paramétrique de fonction de répartition F :

Définition 2.3 Soit X_1, X_2, \dots, X_n un échantillon de variables aléatoires (i.i.d) de fonction de répartition F on a :

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$

alors, l'estimation non paramétrique de la densité est donné par :

$$f_n(x) = \lim_{h \rightarrow 0} \frac{F_n(x+h) - F_n(x-h)}{2h}$$

L'estimation à noyau de Rosenblatt en 1956 notée $\widehat{f}_n(x)$ et définie par :

$$\begin{aligned} \widehat{f}_n(x) &= \lim_{h \rightarrow 0} \frac{\widehat{F}_n(x+h) - \widehat{F}_n(x-h)}{2h} \\ &= \frac{1}{2h} \left[\frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x+h\}} - \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x-h\}} \right] \\ &= \frac{1}{2nh} \left[\sum_{i=1}^n I_{\{\frac{X_i-x}{h} \leq 1\}} - \sum_{i=1}^n I_{\{\frac{X_i-x}{h} \leq -1\}} \right] \\ &= \frac{1}{2nh} \left[\sum_{i=1}^n I_{\{|\frac{X_i-x}{h}| \leq 1\}} \right] \end{aligned}$$

le premier exemple d'estimation à noyau construit à l'aide du noyau :

$$K(z) = \frac{1}{2} I_{\{|z| \leq 1\}}$$

on plus généralement [Parzen(1962)]

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

où K est le noyau de cet estimateur telle que :

- $K(z) \geq 0$
- $\int_{\mathbb{R}} K(z) dz = 1$
- $\int_{\mathbb{R}} tK(t) dt = 0, \int_{\mathbb{R}} t^2 |K(t)| dt < \infty, \int_{\mathbb{R}} K(t)^2 dt < \infty$
- $\sup_t |K(t)| < +\infty$
- $K(\cdot) \in l^1(\mathbb{R})$ c'est à dire $\int_{\mathbb{R}} |K(t)| dt < \infty$
- $\lim_{|t| \rightarrow \infty} |t|K(t) = 0$

$h > 0$ est une paramètre de lissage il dépend de n et il vérifie $h(n) \rightarrow 0$ lorsque $n \rightarrow \infty$.

$$\lim_{n \rightarrow \infty} \frac{nh_n}{\ln h} = \infty$$

2.3.2 Expression de la variance et du le biais d'un estimateurs à noyau

Pour que $\widehat{f}(x)$ soit une densité, on suppose que $K(u) \geq 0$ et $\int_{\mathbb{R}} K(u) = 1$.
 La fonction noyau est supposée être symétrique autour de zéro, c.à.d., $\int_{\mathbb{R}} uK(u)du = 0$
 et possède un moment d'ordre 2 fini, c.à.d., $\int_{\mathbb{R}} u^2K(u)du < \infty$:

Nous avons établi que :

2.3.2.1 L'espérance

On a :

$$E(\widehat{f}_n(x)) = f(x) + \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2) \text{ avec}$$

$$\mu_2(K) = \int_{-\infty}^{+\infty} t^2 K(t) dt$$

En effet, on a :

$$E[\widehat{f}_n(x)] = \frac{1}{nh} E\left(\sum_{i=1}^n K\left(\frac{|X_i - x|}{h}\right)\right)$$

$$= \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{y-x}{h}\right) f(y) dy$$

on posant $t = \frac{y-x}{h} \Rightarrow dy = hdt$

$$E[\widehat{f}_n(x)] = \int_{-\infty}^{+\infty} K(t) f(x+ht) dy$$

en faisant le développement de Taylor à l'ordre 2 de $f(x+ht)$ au voisinage de x on a

alors

$$f(x + ht) = f(x) + (ht)f'(x) + \frac{(ht)^2}{2}f''(x) + o(h^2)$$

il vient

$$\begin{aligned} E[\widehat{f}_n(x)] &= \int_{-\infty}^{+\infty} K(t)f(x + ht)dt \\ &= \int_{-\infty}^{+\infty} K(t)\left[f(x) + (ht)f'(x) + \frac{(ht)^2}{2}f''(x)\right]dt + o(h^2) \\ &= f(x) \int_{-\infty}^{+\infty} K(t)dt + hf'(x) \int_{-\infty}^{+\infty} tK(t)dt \\ &\quad + \frac{h^2}{2}f''(x) \int_{-\infty}^{+\infty} t^2K(t)dt + o(h^2) \end{aligned}$$

d'après les conditions $\int_{-\infty}^{+\infty} K(t)dt = 1$ et $\int_{-\infty}^{+\infty} tK(t)dt = 0$

$$E[\widehat{f}_n(x)] = f(x) + \frac{h^2}{2}f''(x) \int_{-\infty}^{+\infty} t^2K(t)dt + o(h^2)$$

Alors l'expression finale est donnée par

$$\begin{aligned} E(\widehat{f}_n(x)) &= f(x) + \frac{h^2}{2}f''(x)\mu_2(K) + o(h^2) \quad \text{avec} \\ \mu_2(K) &= \int_{-\infty}^{+\infty} t^2K(t)dt \end{aligned}$$

2.3.2.2 Le biais

$$\begin{aligned} \text{biais}(\widehat{f}_n(x)) &= E(\widehat{f}_n(x)) - f(x) \\ &= f(x) + \frac{h^2}{2}f''(x)\mu_2(K) + o(h^2) - f(x) \\ &= \frac{h^2}{2}f''(x)\mu_2(K) + o(h^2). \end{aligned}$$

2.3.2.3 La variance

$$\begin{aligned}
 \text{var}[\widehat{f}_n(x)] &= \left(\frac{1}{nh}\right)^2 E\left(\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)\right) \\
 &= \frac{1}{nh^2} \text{var}\left(K\left(\frac{X_i-x}{h}\right)\right) \\
 &= \frac{1}{nh^2} E\left[K^2\left(\frac{X_i-x}{h}\right)\right] - \frac{1}{nh^2} \left(E\left[K\left(\frac{X_i-x}{h}\right)\right]\right)^2 \\
 &= \frac{1}{nh^2} \int_{-\infty}^{+\infty} K\left(\frac{y-x}{h}\right)^2 f(y) dy \\
 &\quad - \frac{1}{nh^2} \left(\int_{-\infty}^{+\infty} K\left(\frac{y-x}{h}\right) f(y) dy\right)^2.
 \end{aligned}$$

en posant $t = \frac{y-x}{h} \Rightarrow dy = hdt$.

$$\text{var}(\widehat{f}_n(x)) = \frac{1}{nh} \int_{-\infty}^{+\infty} K^2(t) f(x+ht) dt - \frac{1}{n} \left(\int_{-\infty}^{+\infty} K(t) f(x+ht) dt\right)^2$$

En faisant le développement de Taylor à l'ordre 0 au voisinage de x est alors :

$f(x+ht) = f(t) + o(1)$. Il vient

$$\begin{aligned}
 \text{var}(\widehat{f}_n(x)) &= \frac{1}{nh} \int_{-\infty}^{+\infty} K^2(t) (f(t) + o(1)) dt - \frac{1}{n} \left(\int_{-\infty}^{+\infty} K(t) (f(t) + o(1)) dt\right)^2 \\
 &= \frac{1}{nh} f(x) R(k) + o\left(\frac{1}{nh}\right) \quad \text{avec } R(k) = \int_{-\infty}^{+\infty} K^2(t) dt
 \end{aligned}$$

Les expressions asymptotiques du biais et de la variance nous permettent de trouver des expressions pour l'erreur quadratique moyenne (MSE) et l'erreur quadratique moyenne intégrée (MISE) et par définition le MSE et le MISE d'un estimateur sont donnée respectivement par :

$$\begin{aligned}
 \text{MSE}(f(x), \widehat{f}_n(x)) &= E\left[(\widehat{f}_n(x) - f(x))^2\right] \\
 &= E\left[(\widehat{f}_n(x) - E(\widehat{f}_n(x)))^2\right] + 2E[\widehat{f}_n(x) - E(\widehat{f}_n(x))] \\
 &= \left[E(\widehat{f}_n(x)) - f(x)\right] + \left[E(\widehat{f}_n(x)) - f(x)\right]^2 \\
 &= \text{var}[\widehat{f}_n(x)] + \text{biais}^2((\widehat{f}_n(x)))
 \end{aligned}$$

$$\begin{aligned} MISE(f, \widehat{f}) &= \int_{\mathbb{R}} MSE(f(x), \widehat{f}_n(x)) dx \\ &= \int_{\mathbb{R}} var[\widehat{f}_n(x)] + \int_{\mathbb{R}} biais^2(\widehat{f}_n(x)) \end{aligned}$$

En reprenant l'expression du biais de l'estimateur f_n et la variance et en remplaçant dans l'expression MSE on obtient alors :

$$MSE(f(x), \widehat{f}_n(x)) = \frac{h^4}{4} \delta_K^4 (f^{(2)}(x)) + \frac{f(x)}{nh} \int K^2(y) dy + o(h^4) + \frac{1}{nh}$$

$$MISE(f, \widehat{f}) = \frac{h^4}{4} \delta_K^4 \int (f^{(2)}(x))^2 dx + \frac{1}{nh} \int K^2(y) dx + o(h^5 + \frac{1}{nh})$$

avec

$$\delta_K^4 = \left(\int y^2 K(y) dy \right)^2$$

2.3.2.4 Choix du paramètre de lissage

En pratique, on choisit h en fonction de n . Les règles les plus utilisées sont :

Règle de Sturges [Saporta(1990)] Prendre le nombre k de classes égal à $1 + \log_2 n$. En pratique, cela revient à prendre

$$h = \frac{x_k - x_1}{k}$$

Les $x_{(i)}$ sont les valeurs d'observées d'un échantillon ordonné par ordre croissant. la règle de Sturges a tendance à produire des histogrammes trop lisses.

Règle de Scott [Scott et al(1977)] -La valeur qui minimise l'erreur quadratique moyenne intégrée, MISE est donnée par :

$$h_{opt} = \left[\frac{6}{\int f'(x)^2 dt} \right]$$

En prenant pour f la densité de loi normale $N(\mu, \sigma^2)$ d'après Scott [Scott et al(1977)] on peut alors montrer que :

$$h_{opt} = 3.491 \sigma n^{-\frac{1}{3}}$$

En estimant σ par l'écart type S de l'échantillon, on aura :

$$h_{opt} = 3.491 S n^{-\frac{1}{3}}$$

TABLE 2.1 – Quelques noyaux usuels

Noyaux	$K(u)$
Uniforme	$\frac{1}{2}, u \leq 1$
normal	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) u \in \mathbb{R}$
Triangulaire	$(1 - x), u \leq 1$
Epanechnov	$\frac{3}{4\sqrt{5}} \left(1 - \left(\frac{4}{5}\right)^2\right), u \leq \sqrt{5}$

Remarque 2.1 *L'histogramme convient bien pour des analyses relativement grossières. Néanmoins, ses discontinuités n'apparaissent pas très naturelles et ce qui est plus grave, les points tombant près du bords d'une classe et ceux tombant près du milieu ne sont pas différenciés, ceci explique la variabilité des interprétations statistiques que l'on peut faire d'un histogramme suivant le choix de l'origine et des classes. Pour des densités raisonnablement lisses, l'histogramme apparat donc comme un estimateur sévèrement limité. Dans ce chapitre nous allons présenter une méthode plus robuste que la méthode d'histogramme qui est la méthode de noyau.*

2.3.3 Estimation de la densité de probabilité par la méthode du noyau

Il s'agit de l'estimateur le plus populaire. Il est adapté aux variables aléatoires continues. Soit X_1, X_2, \dots, X_n un échantillon de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F et d'une densité f .

L'estimateur à noyau de densité, notée $\widehat{f}(x)$ est définie par

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

Où K est appelé fonction de poids ou noyau, et h est appelé paramètre de lissage ou fenêtre.

2.3.4 Noyaux usuels

2.3.5 Propriétés d'un estimateur à noyau

—

$$\int_{\mathbb{R}} K(u) du = 1 \text{ et } K(u) \geq 0$$

— L'estimateur à noyau est une fonction de densité.

— \widehat{f} a les mêmes propriétés de continuité et de différentiabilité que K :

Si K est continue, \widehat{f} sera une fonction continue

Si K est différentiable, \widehat{f} sera une fonction différentiable. Si K peut prendre des valeurs ngatives, alors \widehat{f} pourra aussi prendre des valeurs négatives.

2.3.6 Expression asymptotique du biais et de la variance

Une approximation asymptotique de l'espérance de l'estimateur $\widehat{f}(x)$ est donné sous les conditions suivantes sur f , h et K

1. La dérivé seconde $f''(x)$ est continue, de carré intégrable et monotone sur $(-\infty, -M)$ et $(M, +\infty)$ pour $M > 0$
2. $\lim_{n \rightarrow +\infty} h = 0$ et $\lim_{n \rightarrow +\infty} nh = 0$
3. Pour que $\widehat{f}(x)$ soit une densité, on suppose que $K(u) \geq 0$ et $\int K(u)du = 1$. La fonction noyau est supposée être symétrique autour de zéro, c.à.d $\int uK(u)du = 0$ et possède un moment d'ordre 2 fini, c.à.d, $\int u^2K(u)du < \infty$.

Nous avons établi que

$$\text{biais}(\widehat{f}(x)) = \frac{h^2}{2} f''(x) \mu_2 + O(h^2) \quad (2.10)$$

$$\text{var}(\widehat{f}_h(x)) = \frac{f(x)}{n} R(k) + O\left(\frac{1}{nh}\right) \quad (2.11)$$

où $\mu_2 = \int K(u)u^2 du$ et $R(g) = \int g^2(u)du$ pour une fonction g de carré intégrable. Il faut donc essayer de choisir un h qui fasse un compromis entre le biais et la variance.

Les expressions asymptotiques du biais et de la variance nous permettent de trouver des expressions asymptotiques pour le MSE et le MISE

Ces expressions ont été obtenues sous la condition (3) sur K et en supposant que la densité de probabilité f avait toutes les dérivées (continues) nécessaires. On peut obtenir facilement les approximations asymptotiques suivantes pour le MSE et le MISE

$$\text{MSE}(\widehat{f}_h(x)) = \frac{h^4}{4} (f''(x))^2 \mu_2^2 + \frac{f(x)}{nh} R(K) + O\left(\frac{1}{nh}\right) + O(h^4) \quad (2.12)$$

$$\text{MISE}(\widehat{f}_h(x)) = \frac{h^4}{4} \mu_2^2 \int_{\mathbb{R}} (f''(x))^2 dx + \frac{1}{nh} R(K) + O(h^4 + \frac{1}{nh}) \quad (2.13)$$

Sous des conditions appropriées d'intégrabilité de f et ses dérivées

On note l'approximation asymptotique de le MSE par

$$\text{AMSE}(\widehat{f}_h(x)) = \frac{h^4}{4} (f''(x))^2 \mu_2^2 + \frac{f(x)}{nh} R(K) \quad (2.14)$$

et l'approximation asymptotique de le MISE par

$$AIMSE(\widehat{f}_h(x)) = \frac{h^4}{4} \mu_2^2 \int_{\mathbb{R}} (f''(x))^2 dx + \frac{1}{nh} R(K) \quad (2.15)$$

2.3.7 Choix théoriques optimaux du paramètre de lissage

Pour le paramètre de lissage on fait la distinction entre h le paramètre de lissage constant (ou global), et $h(x)$ le paramètre de lissage variable (local).

Ces choix différents du paramètre de lissage résultent les estimateurs à noyau suivants :

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

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

Nous allons décrire des choix théoriques optimaux des paramètres de lissage h et $h(x)$. Un critère approprié pour sélectionner un paramètre de lissage constant h et le MISE. Le paramètre de lissage optimal est la valeur de h qui minimise le MISE. Notons cette valeur par h_{MISE} .

Une approximation asymptotique de h_{MISE} est donnée par h_{AMISE} , la valeur de h qui minimise $AMISE\{\widehat{f}_n(\cdot)\}$. Il est facile de vérifier que

$$h_{AMISE} = \left\{ \frac{R(K)}{\mu^2 R(f'')} \right\}^{\frac{1}{5}} n^{-\frac{1}{5}}$$

et

$$h_{MISE} \approx \left\{ \frac{R(K)}{\mu^2 R(f'')} \right\}^{\frac{1}{5}} n^{-\frac{1}{5}}$$

c'est à-dire

$$\lim_{n \rightarrow +\infty} \frac{h_{MISE}}{h_{AMISE}} = 1$$

Un critère approprié pour sélectionner un paramètre de lissage variable (local) $h(x)$ est la mesure de performance locale $MSE(f_{n,L(x)})$. Nous introduisons les notations suivantes :

$$h_{MSE} = \arg \min_h MSE(\widehat{f}_{n,L(x)})$$

et

$$h_{AMISE} = \arg \min_h AMISE(\widehat{f}_{n,L(x)})$$

sous condition que $f''(x) \neq 0$. Les choix h_{AMISE} et $h_{MSE}(x)$ sont des choix théoriques, qui ne sont pas utilisables en pratique car il dépendent des quantités inconnues f et

f'' . En substituant h_{AMSE} dans l'expression de l'AMISE, on montre que pour l'estimateur à noyau

$$n^{\frac{4}{5}} AMISE \widehat{f}_{h_{AMSE}} = O(1)$$

2.3.8 La construction d'un estimateur à noyau

L'estimateur par noyau donnée par [Parzen(1962)] et [Rosenblatt(1956)], est une méthode non paramétrique d'estimation de la densité de probabilité d'une variable aléatoire, elle est basée sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. En ce sens, cette méthode généralise astucieusement la méthode d'estimation par histogramme, en effet, la fonction indicatrice utilisée pour histogramme est ici remplacée par une fonction continue, l'estimateur à noyau est une fonction de deux paramètres : le noyau K et le paramètre de lissage h . Soient X_1, X_2, \dots, X_n de variables aléatoires réelles indépendantes et identiquement distribuées de fonction de répartition F et de densité f . On peut justifier la construction de l'estimateur à noyau de deux façons. Une première idée (développée par [Rosenblatt(1956)]) va être de le construire à partir de l'estimateur

$$F(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}$$

de la fonction de répartition F , puisque

$$\begin{aligned} \forall h > 0 \quad F(x+h) - F(x-h) &= P(x-h < X_1 \leq x+h) \\ &= \int_{]x-h, x+h[} f(y) dy \end{aligned}$$

si h est petit (et par conséquent, pour $y \in]x-h, x+h[$ on peut espérer que $f(y)$ soit proche de $f(x)$) on a l'approximation suivante :

$$f(x) \approx \frac{F(x+h) - F(x-h)}{2h}$$

on présente l'estimateur suivant :

$$\begin{aligned} f_n(x) &= \frac{F_n(x+h) - F_n(x-h)}{2h} = \frac{1}{2nh} \sum_{i=1}^n (1_{\{X_i \leq x+h\}} - 1_{\{X_i \leq x-h\}}) \\ &= \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} 1_{\{x-h < X_i \leq x+h\}} \\ &= \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} 1_{\{-1 < \frac{X_i - x}{h} \leq 1\}} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \end{aligned}$$

en ayant pose

$$K(t) = \frac{1}{2} 1_{\ell\{-1 < t \leq 1\}}; \forall t \in \mathbb{R}$$

Ce cas particulier nous amène à la généralisation suivante [Parzen(1962)], qui consiste à prendre pour estimateur de la densité :

$$f_n(x) = \frac{1}{nh} \sum K\left(\frac{X_i - x}{h}\right)$$

Qui dépend donc de deux choix : le réel h que l'on appellera la fenêtre et la fonction d'une variable réelle à valeurs réelles $K, K : \mathbb{R} \rightarrow \mathbb{R}$; intégrable et telle que $\int K(t) dt = 1$, que l'on appelle le noyau. D'où le nom d'estimateur à noyau.

2.3.8.1 Propriété :

Si K est positive et $\int_{-\infty}^{+\infty} K(t) dt = 1$, alors $f_n(x)$ est une densité de probabilité. De plus, $f_n(x)$ est continue si K est continue .

2.3.8.1.1 preuve : L'estimateur à noyau est positive et continue car la somme des fonctions positives et continues est elle même une fonction positive et continue. Il faut donc vérifier que l'intégrale de $f_n(x)$ vaut un. En effet,

$$\begin{aligned} \int_{-\infty}^{+\infty} f_n(x) dx &= \int_{-\infty}^{+\infty} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) dx \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{+\infty} K\left(\frac{X_i - x}{h}\right) dx \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(t)(-h) dt, \left(t = \frac{X_i - x}{h}\right) \\ &= - \int_{+\infty}^{-\infty} K(t) dt = \int_{-\infty}^{+\infty} K(t) dt = 1 \end{aligned}$$

2.3.8.2 Les noyaux usuel :

Voici quelques exemples de noyaux les plus communément utilisés :

- $K_0(t) = \frac{1}{2}$ si $t \in [-1, 1]$ noyau Uniforme (rectangulaire)
- $K_1(t) = \frac{1}{2} - |t|$ si $t \in [-1, 1]$ noyau de triangulaire
- $K_2(t) = \frac{3}{4}(1 - t^2)$ si $t \in [-1, 1]$ noyau d'Epanechnikov
- $K_3(t) = \frac{15}{16}(1 - t^2)^2$ si $t \in [-1, 1]$ noyau de biweight
- $K_4(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right)$ si $t \in \mathbb{R}$ noyau gaussien
- $K_5(t) = \frac{35}{32}(1 - t^2)^3$ si $t \in [-1, 1]$ noyau de Triweight

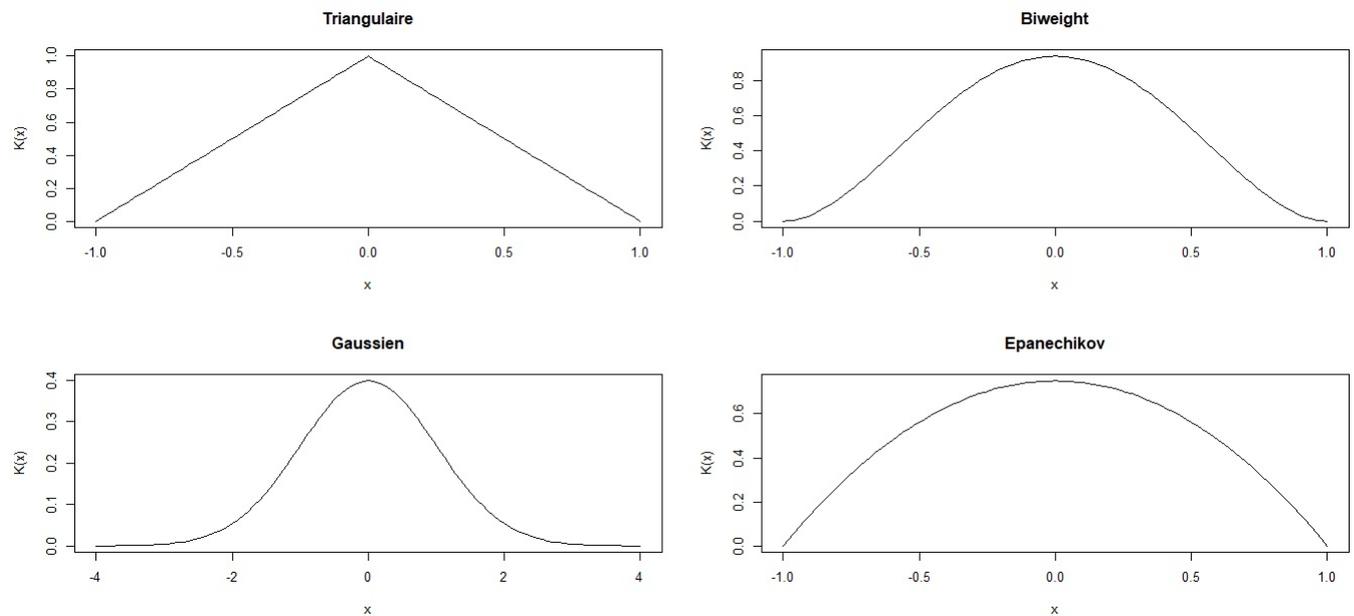


FIGURE 2.1 – Allures des noyaux : Triangulaire, Biweight, Gaussien, Epanechnikov.

2.3.8.2.1 Code R utilisé :

```

K1=function(t){(1-abs(t))*ifelse(abs(t)<=1,1,0)}
K2=function(t){(15/16)*((1-t^2)^2)*ifelse(abs(t)<=1,1,0)}
K3=function(t){dnorm(t)}
K4=function(t){ifelse(abs(t)<=1,(3/4)*(1-t^2),0)}
op=par(mfrow=c(2,2))
curve(K1(x),-1,1,ylab="K(x)",main="Triangulaire")
curve(K2(x),-1,1,ylab="K(x)",main="Biweight")
curve(K3(x),-4,4,ylab="K(x)",main="Gaussien")
curve(K4(x),-1,1,ylab="K(x)",main="Epanechnikov")
    
```

CHAPITRE 3

ESTIMATION NON PARAMÉTRIQUE DE LA RÉPARTITION DES REVENUS DE L'INDICE DE PAUVRETÉ

3.1 Introduction et définition de l'estimateur

Certaines améliorations de l'estimation non paramétrique de la répartition des revenus et de l'indice de pauvreté

Définition 3.1 Soit F la fonction de distribution cumulative de la variable de revenu X d'une population avec une fonction de densité continue f . La classe FGT (Foster, Greer, Thorbecke) [Foster et al. (1984)] de mesures de l'indice de pauvreté par le réel $\alpha \geq 0$ est définie par :

$$\bar{p}(z, \alpha) = \begin{cases} \int_0^z \left(\frac{z-x}{z}\right)^\alpha f(x) dx & \text{si } z > 0 \\ 0 & \text{sinon} \end{cases} \quad (3.1)$$

où z est le seuil de pauvreté.

3.2 Estimation empirique de l'indice FGT

Soit X_1, \dots, X_n un échantillon aléatoire de taille n à partir de variables aléatoires de revenu (v.a) avec la fonction de distribution F . Seidl (1988) [Seidl (1988)] a introduit Un estimateur empirique de l'indice de pauvreté FGT $P(z, \alpha)$ comme suit :

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

Cet estimateur à été pleinement utile dans un large éventail d'applications en économie (largement utilisé dans la pratique en économétrie et actuariat). Il s'agit d'un estimateur cohérent sans biais et est asymptotiquement normal. Lo et al. [Lo et al (2009)] a utilisé des processus empiriques et une méthodologie des valeurs extrêmes pour étudier cet estimateur. Seck [Seck(2011)] , Seck et Lo [Seck et al (2009)] ont utilisé des mesures de pauvreté non pondérées, considérées comme des processus stochastiques et indexées par des nombres ou des fonctions monotones, pour suivre l'évolution de la pauvreté entre deux périodes.

3.3 Estimation non paramétrique de FGT

Dia [Dia (2008)] et aussi Ciss et coll. [Ciss et al. (2015)] ont proposé de nouveaux estimateurs noyaux, basés respectivement sur la somme de Riemann pour $\alpha = 0$ et $\alpha \geq 1$ et $\alpha \in [0, 1]$ Dia [Dia (2008)] et Ciss et al. [Ciss et al. (2015)] ont également considéré l'estimateur non paramétrique classique de la densité f (Parzen – Rosenblatt) :

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

où h est une fonction de n qui tend vers zéro lorsque n tend vers l'infini et K vérifie les hypothèses suivantes :

$$\begin{aligned} (H_1) \quad \sup_{-\infty < x < +\infty} |K(x)| &< +\infty, \\ (H_2) \quad \int_{-\infty}^{+\infty} K(x) dx &= 1 \\ (H_3) \quad \lim_{x \rightarrow \pm\infty} |xK(x)| &= 0 \end{aligned} \tag{3.2}$$

et proposé comme estimateur de l'indice de pauvreté FGT, le suivant :

$$\widetilde{p}_n(z, \alpha) = \frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{\lfloor z/h \rfloor} \left(\frac{z - ih}{z}\right)^\alpha K\left(\frac{ih - X_j}{h}\right). \tag{3.3}$$

3.4 Estimation non paramétrique avec noyau adapté

Récemment, [Zakaria et al (2018)] ont considéré l'estimateur adaptatif par noyau suivant de la densité f

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

et proposé comme estimateur de l'indice de pauvreté FGT, l'estimateur suivant :

$$p_n^\lambda(z, \alpha) = \frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{\lfloor z/h\lambda_j \rfloor} \left(\frac{z - ih\lambda_j}{z} \right)^\alpha K\left(\frac{ih\lambda_j - X_j}{h\lambda_j} \right) \quad \text{pour } \alpha = 0 \quad \text{ou } \alpha \geq 1. \quad (3.4)$$

où λ_j , est un paramètre qui varie en fonction de la concentration locale des données.

3.5 Estimation non paramétrique sans biais

[Parzen(1962)]et [Silverman (1986)] ont souligné que si $\int uk(u) du = 0$ et f est deux fois continûment dérivable dans un voisinage de x , alors

$$\begin{aligned} \text{Biais}(\widehat{f}(x)) &= E(\widehat{f}(x) - f(x)) \\ &= \frac{1}{2} f''(x) h^2 \int u^2 K(u) du + O(h^3). \end{aligned} \quad (3.5)$$

Afin de réduire le biais de l'estimateur noyau classique \widehat{f} , nous considérons l'estimateur suivant introduit dans[Xie et al (2014)]

$$f(x) = \widehat{f}(x) - \widehat{\text{bias}}(\widehat{f}(x))$$

qui peut être écrit sous la forme suivante

$$\begin{aligned} \widetilde{f}(x) &= \widehat{f}(x) - \frac{h^2}{2} \widehat{f}''(x) \int u^2 K(u) du \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) - \frac{h}{2n} \int u^2 K(u) du \sum_{i=1}^n K''\left(\frac{x - X_i}{h}\right) \end{aligned} \quad (3.6)$$

De plus, nous supposons que les hypothèses H_1, H_2, H_3 sont valables pour K et K'' , ce sont intégrable au sens de Riemann et deux fois continûment différentiable, et que f est bornée avec le support inclus dans \mathbb{R}_+ . On note x_0 le minimum de ce support. Soit substitué dans 3.1 f par \widetilde{f} On obtient

$$\begin{aligned} \widetilde{J}_n(z, \alpha) &= \int_0^z \left(\frac{z-x}{z} \right)^\alpha (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \\ &\quad - \frac{h}{2n} \int u^2 K(u) du \int_0^z \left(\frac{z-x}{z} \right)^\alpha \sum_{i=1}^n K''\left(\frac{x - X_i}{h}\right) \end{aligned} \quad (3.7)$$

soit $\Delta_{hi} = [hi, h(i+1)]$, $0 \leq i < \lfloor \frac{z}{h} \rfloor$ soit une partition de $[0, z]$ et en utilisant la définition de la somme de Riemann de l'intégrale, on établit qu'elle correspond à l'intégrale

$\widetilde{J}_n(z, \alpha)$ de la somme suivante :

$$\begin{aligned} \widetilde{p}_n(z, \alpha) = & \frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{\lfloor z/h \rfloor} \left(\frac{z-ih}{z} \right)^\alpha \left[K \left(\frac{ih-X_j}{h} \right) \right. \\ & \left. - \frac{h^2}{n} \int u^2 K(u) du K'' \left(\frac{hi-X_j}{h} \right) \right] + V_{n,b}(z) \end{aligned} \quad (3.8)$$

où $\lfloor \frac{\cdot}{h} \rfloor$ est la partie entière de $\frac{\cdot}{h}$ et $v_{n,b}(z) \rightarrow 0$ en probabilité comme $n \rightarrow +\infty$ (pour plus de détails, on peut voir la section 3.7.1. Enfin, nous proposons comme estimateur de l'indice de pauvreté FGT, le nouvel estimateur suivant :

$$\begin{aligned} \widehat{P}_{n,b}(z, \alpha) = & \frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{\lfloor z/h \rfloor} \left(\frac{z-ih}{z} \right)^\alpha \left[K \left(\frac{ih-X_j}{h} \right) \right. \\ & \left. - \frac{h^2}{n} \int u^2 K(u) du K'' \left(\frac{hi-X_j}{h} \right) \right] \end{aligned} \quad (3.9)$$

pour $\alpha = 0$ ou $\alpha \geq 1$.

Des hypothèses supplémentaires sont faites à la fois sur K et K'' , c'est-à-dire :

- (H_4) K est la fonction de variation bornée $V_{-\infty}^u K$ sur \mathbb{R} et soit $V(\mathbb{R})$ sa variation totale.
- (H_5) $\int_{\mathbb{R}} |uK(u)| du < +\infty$ et $\int u^2 K(u) du < +\infty$.
- (H_6) Il existe une fonction λ non croissante telle que $\lambda(\frac{u}{h}) = O(h)$ sur des intervalles bornés.

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

3.6 Principaux résultats

Nos principaux résultats sont relatifs aux éléments supplémentaires suivants concernant la fonction de densité f :

C_1 : f est uniformément continue

C_2 : f est deux fois presque partout différentiable $\dot{f}, f'' \in L_1(\mathbb{R})$.

3.6.1 Cohérence et comportement uniformes presque sûrs du biais

Théorème 3.1 *Supposons que les hypothèses H_4 et C_1 soient vérifiées. Alors pour tout $M > 0$, l'estimateur $P_{n,b}(z, \alpha)$ converge uniformément presque sûrement sur $[0, M]$ vers $p(z, \alpha)$*

comme $n \rightarrow +\infty$ i.e.

$$P\left(\lim_{n \rightarrow +\infty} \sup_{z \in [0, M]} |p_{n,b}(z, \alpha) - p(z, \alpha)| = 0\right) = 1$$

avec

$$nh^2(\log \log n)^{-1} \rightarrow +\infty \text{ lorsque } n \rightarrow +\infty.$$

Théorème 3.2 *Supposons que les hypothèses H_4 , H_5 et C_2 soient vérifiées. Alors pour tout $M > 0$, l'estimateur $P_{n,b}(z, \alpha)$ converge uniformément presque sûrement sur $[0, M]$ vers $p(z, \alpha)$ comme $n \rightarrow +\infty$ i.e.*

$$P\left(\lim_{n \rightarrow +\infty} \sup_{z \in [0, M]} |p_{n,b}(z, \alpha) - p(z, \alpha)| = 0\right) = 1$$

avec

$$nh^2(\log \log n)^{-1} \rightarrow +\infty \text{ comme } n \rightarrow +\infty.$$

Pour la démonstration des théorèmes, nous utilisons le théorème 3.2 de Kiefer [?] et les lemmes suivants montrant que $P_{n,b}(z, \alpha)$ est uniformément asymptotique sans biais sur tout intervalle

Lemme 3.1 *Si l'hypothèse C_1 est vraie, alors $\forall M > 0$, on a*

$$\lim_{n \rightarrow +\infty} \sup_{z \in [0, M]} |E(p_{n,b}(z, \alpha)) - p(z, \alpha)| = 0$$

Lemme 3.2 *Si les hypothèses H_5 et C_2 sont vérifiées, alors :*

$$\begin{aligned} \sup_{z \in \mathbb{R}} |E(p_{n,b}(z, \alpha)) - p(z, \alpha)| \leq h \left(\left(\int_{\mathbb{R}} |f(x)| dx \right) \left(\int_{\mathbb{R}} (|u| + 1) |K(u)| du \right) \right. \\ \left. + 2(\alpha D + Ah) \int_{-\infty}^{+\infty} |K(u)| du \right) + h^3 \int u^2 K(u) du. \end{aligned} \quad (3.10)$$

où

$$D = \sup_{x \in \mathbb{R}} F(x) \text{ et } A = \sup_{x \in \mathbb{R}} f(x)$$

Remarque 3.1 *Si K satisfait l'hypothèse H_5 , alors en utilisant H_1 , le noyau $\widehat{K} = \frac{K^2}{\int_{\mathbb{R}} K^2(y) dy}$ le satisfaire également. Des deux lemmes précédents, nous obtenons les corollaires suivants :*

Corollaire 3.1 *Sous les hypothèses du lemme 3.1, on a sur $[0, M]$ (resp \mathbb{R})*

$$\begin{aligned} & \lim_{n \rightarrow +\infty} E \left(\sum_{i=1}^{+\infty} \left[1 - \frac{ih}{z} \right]^{2\alpha} \left[K \left(\frac{ih - X_j}{h} \right) - \frac{1}{2} \int u^2 K(u) dK'' \left(1 - \frac{ih}{z} \right) \right]^2 \right) \\ &= \left(\int_{\mathbb{R}} k(y) dy \right) p(z, 2\alpha) \end{aligned}$$

Corollaire 3.2 *Si les hypothèses du théorème 3.2 sont vérifiées et si $h = O(n^{-1} \log \log n)^{1/4}$, alors pour tout $M > 0$, on a presque sûrement :*

$$\sup_{z \in [0, M]} |E(p_{n,b}(z, \alpha)) - p(z, \alpha)| = O(n^{-1} \log \log n)^{3/4}$$

3.6.2 Consistance quadratique moyenne Uniforme

Théorème 3.3 *Si H_6 et C_1 sont vérifiées, on a :*

1.

$$\lim_{n \rightarrow +\infty} n \text{Var}(p_{n,b}(z, \alpha)) = \left(\int_{\mathbb{R}} K^2(y) dy \right) p(z, 2\alpha) - (p(z, \alpha))^2$$

2. Et Pour tout $M > 0$

$$\lim_{n \rightarrow +\infty} \sup_{z \in [0, M]} E(p_{n,b}(z, \alpha) - p(z, \alpha))^2 = 0.$$

Théorème 3.4 *Supposons que H_6 et C_2 sont satisfaites. Alors :*

1.

$$\lim_{n \rightarrow +\infty} n \text{Var}(p_{n,b}(z, \alpha)) = \left(\int_{\mathbb{R}} K^2(y) dy \right) p(z, 2\alpha) - (p(z, \alpha))^2$$

2. De plus, si H_5 tient, on a pour tout $M > 0$,

$$\lim_{n \rightarrow +\infty} \sup_{z \in [0, M]} E(p_{n,b}(z, \alpha) - p(z, \alpha))^2 = 0.$$

Pour la démonstration de ce théorème, nous supposons que l'hypothèse C_1 ou C_2 est vraie et avant cela, nous prouvons le théorème 3.5 ci-dessous en utilisant le lemme 3.3 suivant.

Lemme 3.3 *Let $0 \leq \theta_i \leq 1$, $i = 1, 2$. $\forall x, y$ et $x \neq y$ nous avons*

$$\lim_{n \rightarrow +\infty} \left((h)^{-2} \int_{-\infty}^{+\infty} \left(\left| \left[K \left(\frac{u - x + \theta_1}{h} \right) - \frac{1}{2} \int s^2 K(s) ds K'' \left(\frac{u - x + \theta_1}{h} \right) \right] \right|^2 \right) \right)$$

$$\times \left[K\left(\frac{u-y+\theta_2}{h}\right) - \frac{1}{2} \int s^2 K(s) ds K''\left(\frac{u-y+\theta_2}{h}\right) \right] f(u) du = 0$$

Théorème 3.5 *Supposons que l'hypothèse H_6 soit vraie. Alors pour tout $M > 0$.*

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \sup_{z \in [0, M]} \sum_{0 \leq i \neq j \leq \lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z}\right)^\alpha \left(1 - \frac{jh}{z}\right)^\alpha \\ & \times \int_{\mathbb{R}} \left[K\left(\frac{u-ih}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-ih}{h}\right) \right] \\ & \times \left[K\left(\frac{u-jh}{h}\right) - \frac{h^2}{2} \int_{\mathbb{R}} s^2 K(s) ds K''\left(\frac{u-jh}{h}\right) \right] f(u) du = 0 \end{aligned}$$

Remarque 3.2 *L'estimateur $p_{n,-b}(z, \alpha)$ a une efficacité asymptotique par rapport à $\widehat{p}(z, \alpha)$,*

$$e(z, \alpha) = \frac{\left(\int_{\mathbb{R}} K^2(y) dy \right) p(z, 2\alpha) - (p(z, \alpha))^2}{p(z, 2\alpha) - (p(z, \alpha))^2}.$$

L'intégrale $\int_{\mathbb{R}} K^2(y) dy$ est strictement inférieur à 1 pour les noyaux conventionnels [Parzen(1962)] p.1068. En suite, nous avons dans ce cas $e(z, \alpha) < 1$. Dans le 3.4, la vitesse de convergence quadratique moyenne est d'ordre $O(\frac{1}{n^2})$ si h est d'ordre de $O(\frac{1}{\sqrt{n}})$.

3.7 Détails des preuves

3.7.1 Construction de l'estimateur

Pour $z > 0$ et $\Delta_{h,i} = [hi, h(i+1)[$, $i = 0, \dots, \lfloor \frac{z}{h} \rfloor$ Nous avons la somme de Riemann suivante sur l'intervalle $[0, z]$:

$$\begin{aligned} S_n(z, \alpha) &= \frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{\lfloor \frac{z}{h} \rfloor - 1} \left(1 - \frac{ih}{z}\right)^\alpha \left[K\left(\frac{ih - X_j}{h}\right) - \frac{h^2}{2} \int u^2 K(u) du K''\left(\frac{ih - X_j}{h}\right) \right] \\ &+ \left(z - h \lfloor \frac{z}{h} \rfloor \right) \sum_{j=1}^n \left(1 - \frac{h \lfloor \frac{z}{h} \rfloor}{z}\right)^\alpha \frac{1}{n} \frac{1}{h} \\ &\times \left[K\left(\frac{h \lfloor \frac{z}{h} \rfloor - X_j}{h}\right) - \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) du K''\left(\frac{h \lfloor \frac{z}{h} \rfloor - X_j}{h}\right) \right] \end{aligned}$$

correspond à l'intégrale

$$J_n(z, \alpha) = \int_0^z \left(\frac{z-x}{z}\right)^\alpha \frac{1}{n} \sum_{j=1}^n \left[K\left(\frac{x - X_j}{h}\right) - \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) du K''\left(\frac{x - X_j}{h}\right) \right] dx.$$

La somme S_n peut être réécrite comme

$$S_n = \frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z}\right)^\alpha \left[K\left(\frac{ih - X_j}{h}\right) - \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) du K''\left(\frac{ih - X_j}{h}\right) \right] + V_{n,b}(z)$$

$$\begin{aligned} V_{n,b}(z) &= \frac{1}{n} \sum_{j=1}^n \frac{(z - h\lfloor \frac{z}{h} \rfloor) - h}{h} \left(1 - \frac{h\lfloor \frac{z}{h} \rfloor}{z}\right)^\alpha \\ &\quad \times \left[K\left(\frac{h\lfloor \frac{z}{h} \rfloor - X_j}{h}\right) - \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) du K''\left(\frac{h\lfloor \frac{z}{h} \rfloor - X_j}{h}\right) \right] \end{aligned}$$

Maintenant, nous devons montrer que $V_{n,b} \rightarrow 0$ en probabilité lorsque $n \rightarrow \infty$. Depuis $\lfloor \frac{z}{h} \rfloor \leq \frac{z}{h} < \lfloor \frac{z}{h} \rfloor + 1$ on obtient pour $\alpha \geq 0$:

$$|V_{n,b}(z)| \leq \frac{1}{n} \sum_{j=1}^n \left| K\left(\frac{h\lfloor \frac{z}{h} \rfloor - X_j}{h}\right) - \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) du K''\left(\frac{h\lfloor \frac{z}{h} \rfloor - X_j}{h}\right) \right| \quad (3.11)$$

Notez que, pour tout $x \in \mathbb{R}$, $hx + h\lfloor \frac{z}{h} \rfloor \rightarrow z$ lorsque $n \rightarrow \infty$. Ainsi, pour $\alpha = 0$, en utilisant la continuité de f et le fait que $K, K'' \in L_1(\mathbb{R})$ on a pour n assez grand :

$$\begin{aligned} &\left| E \left[\frac{1}{n} \sum_{j=1}^n \left\{ K\left(\frac{h\lfloor \frac{z}{h} \rfloor - X_j}{h}\right) - \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(s) ds K''\left(\frac{h\lfloor \frac{z}{h} \rfloor - X_j}{h}\right) \right\} \right] \right| \quad (3.12) \\ &= \left| h \int \left[K(u) - \frac{h^2}{2} \int s^2 K(s) ds K''(u) \right] f\left(h\lfloor \frac{z}{h} \rfloor - uh\right) du \right| \\ &\leq hf(z)(1 + o(1)) \int |K(u)| du \end{aligned}$$

Par conséquent, $E|V_{n,b}(z)| = O(h)$ comme $n \rightarrow \infty$. L'estimateur tient en utilisant l'inégalité de Markov.

3.7.2 Preuves des principaux résultats

3.7.2.1 Preuve du lemme 3.1

Définissons $\bar{\Delta}_{h,i} = \Delta_{h,i} \cap [0, M]$; et χ_B la fonction d'indicatrice de l'ensemble B .

$$\varphi_h(u) = \frac{h^2}{2} \int s^2 K(s) ds K''(u)$$

Par un changement de variables, nous avons

$$\begin{aligned}
 E(p_n(z, \alpha) - p(z, \alpha)) &= \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z}\right)^\alpha \int (K(u) - \varphi_h(u)) f(ih - uh) du - p(z, \alpha) \\
 &= \int_0^z \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \chi_{\bar{\Delta}_{h,i}}(x) \left(1 - \frac{ih}{z}\right)^\alpha \int (K(u) - \varphi_h(u)) f(ih - uh) du dx - p(z, \alpha) \\
 &\quad + (h(\lfloor \frac{z}{h} \rfloor + 1) - z) \left(1 - \frac{h\lfloor \frac{z}{h} \rfloor}{z}\right)^\alpha \int (K(u) - \varphi_h(u)) f(h\lfloor \frac{z}{h} \rfloor - uh) du \\
 &= A_{n,1} + A_{n,2}
 \end{aligned}$$

Nous étudions d'abord le terme $A_{n,2}$. Soit $z \in [0, M]$, puisque $|h(\lfloor \frac{z}{h} \rfloor + 1) - z| \leq h$ et $0 < h\lfloor \frac{z}{h} \rfloor \leq z$, on obtient pour n grand

suffisant :

$$\begin{aligned}
 \sup_{x \in [0, M]} |A_{n,2}| &= \sup_{x \in [0, M]} \left| \left(h(\lfloor \frac{z}{h} \rfloor + 1) - z \right) \left(1 - \frac{h\lfloor \frac{z}{h} \rfloor}{z}\right)^\alpha \int (K(u) - \varphi_h(u)) f(h\lfloor \frac{z}{h} \rfloor - uh) du \right| \\
 &\leq h(\sup f(z))(1 + o(1)) \int |K(u)| du
 \end{aligned}$$

Maintenant, nous allons étudier le terme $A_{1,n}$. Sous H_2 , on peut réécrire $p(z, \alpha)$ comme :

$$\begin{aligned}
 p(z, \alpha) &= \int_0^z \left(1 - \frac{x}{z}\right)^\alpha \int ((K(u) - \varphi_h(u)) du) f(x) dx \\
 &\quad + \int_0^z \left(1 - \frac{x}{z}\right)^\alpha \int \varphi_h(u) du f(x) dx
 \end{aligned}$$

Soit $x \in \bar{\Delta}_{h,i}, i = 0, \dots, \lfloor \frac{z}{h} \rfloor$. Nous avons

$$\begin{aligned}
 &\left| \left(1 - \frac{x}{z}\right)^\alpha \int ((K(u) - \varphi_h(u)) f(ih - uh) du) - \left(1 - \frac{x}{z}\right)^\alpha \int ((K(u) - \varphi_h(u)) du) f(x) \right| \quad (3.13) \\
 &= \left| \int \left[\left(1 - \frac{ih}{z}\right)^\alpha f(ih - uh) - \left(1 - \frac{x}{z}\right)^\alpha f(x) \right] ((K(u) - \varphi_h(u)) du) \right| \\
 &\leq \int \left| \left(1 - \frac{ih}{z}\right)^\alpha - \left(1 - \frac{x}{z}\right)^\alpha \right| \times |K(u) - \varphi_h(u)| f(x) du \\
 &\quad + \int \left(1 - \frac{ih}{z}\right)^\alpha |f(ih - uh) - f(x)| \times |K(u) - \varphi_h(u)| du f(x)
 \end{aligned}$$

Pour $\alpha = 0$, nous avons $\left| \left(1 - \frac{ih}{z}\right)^\alpha - \left(1 - \frac{x}{z}\right)^\alpha \right| = 0$. En suite, pour $\alpha \geq 1$, en utilisant la

formule de Taylor du premier ordre sur $[ih, x]$ pour la fonction $g(t) = (1 - \frac{t}{z})^\alpha$, il existe $c_i \in]ih, x[$, tels que $g(ih) - g(x) = \dot{g}(c_i)(ih - x)$. C'est-à-dire

$$\left| \left(1 - \frac{ih}{z}\right)^\alpha - \left(1 - \frac{x}{z}\right)^\alpha \right| = \frac{\alpha}{z} \left|1 - \frac{c_i}{z}\right|^{\alpha-1} |ih - x| \leq \frac{\alpha h}{z}$$

Désigné par $I_1^i(x)$ (respectivement $I_2^i(x)$), le premier terme (respectivement le second terme) du côté droit de l'inégalité 3.13. Pour simplifier les notations, laissez

$$I_1(x) = \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \chi_{\Delta_{h,1}^-(x)} I_1^i(x) \text{ et } I_2(x) = \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \chi_{\Delta_{h,1}^-(x)} I_2^i(x)$$

on a

$$|A_{1,n}| \leq \int_0^z I_1(x) dx + \int_0^z I_2(x) dx + \int_0^z \left(1 - \frac{x}{z}\right)^\alpha \int |\varphi_h(u)| du f(x) dx.$$

Nous remarquons par définition que $\int |\varphi_h(u)| du = O(h^2)$. Cela mène à

$$|A_{1,n}| \leq \int_0^z I_1(x) dx + \int_0^z I_2(x) dx + O(h^2)$$

Maintenant, nous allons étudier les termes $\int_0^z I_1(x) dx$ et $\int_0^z I_2(x) dx$. Premièrement, nous avons

$$\begin{aligned} \int_0^z I_1(x) dx &\leq \frac{\alpha h}{z} \int_0^z \left(\int f(x) |k(u) - \varphi_h(u)| du \right) dx \\ &\leq \frac{\alpha h}{z} F(z) \int (|k(u)| + \varphi_h(u)) du \end{aligned} \quad (3.14)$$

Remarquant également que $\int |\varphi_h(u)| du = O(h^2)$, on obtient par hypothèses

$$\int_0^z I_1(x) dx = O(h)$$

Ensuite, notez que

$$\begin{aligned} I_2(x) &\leq \left(\int |f(ih - uh) - f(ih)| \times |K(u) - \varphi_h(u)| du \right. \\ &\quad \left. + \int |f(ih) - f(x)| \times |K(u) - \varphi_h(u)| du \right) \end{aligned} \quad (3.15)$$

Soit $\varepsilon > 0$, puisque f est uniformément continue, il existe $\eta_0 = \eta_0(z) > 0$ tel que $|ih - x| \leq \eta_0$ donc si $h \leq \eta_0$, on ont $|f(ih) - f(x)| < \frac{\varepsilon}{M}$. Par conséquent,

$$\begin{aligned} \int_0^z I_2(x) dx &= \leq \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int |f(ih - uh) - f(ih)| \\ &\quad \times |K(u) - \varphi_h(u)| du + \varepsilon \int |K(u) - \varphi_h(u)| du \end{aligned}$$

Par la continuité uniforme de f nous avons

$$\exists \eta_1 = \eta_1(z) > 0, |uh| < \eta_1 \Rightarrow |f(ih - uh) - f(ih)| < \frac{\varepsilon}{M}$$

Par conséquent

$$\begin{aligned} &\sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int |f(ih - uh) - f(ih)| \times |K(u) - \varphi_h(u)| du \\ &\leq \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int_{|uh| < \eta_1} \frac{\varepsilon}{M} |K(u) - \varphi_h(u)| du \\ &\quad + \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int_{|uh| \geq \eta_1} |f(ih - uh) - f(ih)| \times |K(u) - \varphi_h(u)| du \\ &\leq \varepsilon \int |K(u) - \varphi_h(u)| du + \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int_{|uh| \geq \eta_1} |f(ih - uh)| \times |K(u) - \varphi_h(u)| du \\ &\quad + \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int_{|uh| \geq \eta_1} |f(ih)| \times |K(u) - \varphi_h(u)| du \end{aligned}$$

Puisque f est continue, elle est Riemann-intégrable et en remarquant que

$$(h(\lfloor \frac{z}{h} \rfloor + 1) - z)f(h\lfloor \frac{z}{h} \rfloor) \rightarrow 0, n \rightarrow +\infty$$

f nous avoir la somme

$$\sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} hf(ih) \rightarrow \int_0^z f(x) dx = F(z)$$

est borné. Soit A ce dernier. Par le changement des variables $v = uh$, on obtient

$$\begin{aligned} & \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int_{|uh| \geq \eta_1} f(ih - uh) |K(u) - \varphi_h(u)| du + \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int_{|uh| \geq \eta_1} f(ih) |K(u) - \varphi_h(u)| du \\ & \leq \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int_{|v| \geq \eta_1} f(ih - v) \frac{1}{h} \left| K\left(\frac{v}{h}\right) - \varphi_h\left(\frac{v}{h}\right) \right| dv + A \int_{|uh| \geq \eta_1} |K(u) - \varphi_h(u)| du \end{aligned}$$

Sous (H_3) , il existe $C > 0$ fixe, tel que $|\frac{v}{h}| \geq C$ nous avons

$$\frac{v}{h} \left| K\left(\frac{v}{h}\right) - \varphi_h\left(\frac{v}{h}\right) \right| \leq \frac{\varepsilon \eta_1}{M}$$

Soit $\eta = \inf(\eta_1, Ch)$, h étant assez petit, alors

$$\begin{aligned} \frac{1}{\eta_1} \int_{|v| \geq \eta_1} f(ih - v) \frac{v}{h} \left| K\left(\frac{v}{h}\right) - \varphi_h\left(\frac{v}{h}\right) \right| dv & \leq \frac{\varepsilon}{M} \int_{|v| \geq \eta_1} f(ih - v) dv \\ & \leq \frac{\varepsilon}{M} \int_{|v| \geq \eta_1} f(x) dx \leq \frac{\varepsilon}{M} \end{aligned}$$

D'où,

$$\begin{aligned} \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} h \int_{|v| \geq \eta_1} f(ih - uh) \left| K\left(\frac{v}{h}\right) - \varphi_h\left(\frac{v}{h}\right) \right| dv & \leq \frac{\varepsilon z}{M} \\ & \leq \varepsilon \end{aligned}$$

puisque $A \int_{|uh| \geq \eta_1} |K(u) - \varphi_h(u)| du \rightarrow 0, n \rightarrow +\infty$, nous avons ensemble avec 3.14

$$\lim_{n \rightarrow +\infty} \sup_{z \in [0, M]} |E(p_n(z, \alpha)) - p(z, \alpha)| \leq 2\varepsilon \int |K(u)| du + \varepsilon$$

La preuve du lemme 3.2 est complète.

3.7.2.2 Preuve du Lemme 3.2

Pour $x \in \bar{\Delta}_{h,i}, i = 0, \dots, \lfloor \frac{z}{h} \rfloor$ On a

$$\begin{aligned} & \int (1 - \frac{ih}{z})^\alpha |f(ih - uh) - f(x)| \times |K(u) - \varphi_h(u)| du \\ & \leq \left(\int \int_x^{x+h(|u|+1)} |f(t)| \times |K(u) - \varphi_h(u)| dt du \right) \end{aligned}$$

Donc, en utilisant à nouveau l'expression de $I_2(x)$ définie dans la preuve du lemme(3.1),

on a

$$\int_0^z I_2(x)dx \leq \int_0^z \left(\int_x^{x+h(|u|+1)} |\hat{f}(t)| \times |K(u) - \varphi_h(u)| dt du \right) dx \quad (3.16)$$

Par le changement de variable avec $t = x + h(|u| + 1)v$ et en utilisant le théorème de Fubini, on obtient

$$\begin{aligned} \int_0^z I_2(x)dx &\leq h \int (|u| + 1) |K(u) - \varphi_h(u)| du \\ &\quad \times \left(\int |\hat{f}(x + h(|u| + 1)v)| \times |K(u) - \varphi_h(u)| du \right) dx du \int_0^1 dv \\ &= \int_{\mathbb{R}} |\hat{f}(x)| dx \end{aligned}$$

Cette inégalité ainsi que 3.13 et 3.14 conduisent à compléter la preuve.

3.7.2.3 Preuve du Théorème 3.1 et du Théorème 3.2

Preuve.. Soit \hat{F}_n la distribution empirique de l'échantillon (X_1, X_2, \dots, X_n) définie par

$$F_n = n^{-1} \sum_{i=1}^n (\chi_{X_i} < l)$$

où χ_A représente la fonction indicatrice A . On peut écrire

$$p_n(z, \alpha) = \int_{\mathbb{R}} \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z}\right)^\alpha \left[K\left(\frac{l-ih}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{l-ih}{h}\right) \right] d\hat{F}_n(l)$$

et

$$E(p_n(z, \alpha)) = \int_{\mathbb{R}} \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z}\right)^\alpha \left[K\left(\frac{l-ih}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{l-ih}{h}\right) \right] dF(l)$$

on a

$$|p_n(z, \alpha) - E(p_n(z, \alpha))| = \left| \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \int_{\mathbb{R}} \left(1 - \frac{ih}{z}\right)^\alpha \left[K\left(\frac{l-ih}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{l-ih}{h}\right) \right] (d\hat{F}_n(l) - dF(l)) \right|$$

l'intégration par parties donne

$$\begin{aligned}
 |p_n(z, \alpha) - E(p_n(z, \alpha))| &\leq \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \int_{\mathbb{R}} \left| d \left[K \left(\frac{l-ih}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{l-ih}{h} \right) \right] \right| \\
 &\quad \sup_{l \in \mathbb{R}} |\hat{F}_n(l) - F(l)| \\
 &\leq \sum_{i=1}^{\lfloor \frac{z}{h} \rfloor} \int_{\mathbb{R}} d \left[K_{-\infty}^{\frac{l-ih}{h}} - \frac{h}{2} \int s^2 K(s) ds K''_{-\infty}^{\frac{l-ih}{h}} \right] \sup_{l \in \mathbb{R}} |\hat{F}_n(l) - F(l)| \\
 &\leq \left[\frac{z}{h} \right] V(\mathbb{R}) \left(1 - \frac{h^2}{2} \int s^2 K(s) ds \right) \sup_{l \in \mathbb{R}} |\hat{F}_n(l) - F(l)|
 \end{aligned}$$

Remarquant que

$$\begin{aligned}
 |p_n(z, \alpha) - (p_n(z, \alpha))| &= |p_n(z, \alpha) - E(p_n(z, \alpha)) + E(p_n(z, \alpha)) - p(z, \alpha)| \\
 &\leq |p_n(z, \alpha) - E(p_n(z, \alpha))| + |E(p_n(z, \alpha)) - p(z, \alpha)|
 \end{aligned}$$

Par le Lemme 3.1 nous avons

$$|E(p_n(z, \alpha)) - p(z, \alpha)| \rightarrow 0 \quad n \rightarrow +\infty$$

et les résultats précédents que nous avons

$$|p_n(z, \alpha) - E(p_n(z, \alpha))| \rightarrow 0 \quad n \rightarrow +\infty$$

■

3.7.3 La consistance moyenne quadratique uniforme

3.7.3.1 Preuve Lemme 3.3

Nous supposons que C_1 est vrai. Soit $\delta > 0$.

Définir

$$\begin{aligned}
 I_n(x, y) &= (h)^{-2} \int_{-\infty}^{+\infty} \left| \left[K\left(\frac{u-x+\theta_1 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-x+\theta_1 h}{h}\right) \right] \right| \\
 &\quad \times \left| \left[K\left(\frac{u-x+\theta_2 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-x+\theta_2 h}{h}\right) \right] | f(u) | du \\
 &= \int_{-\infty}^{+\infty} \left((h)^{-1} \left[K\left(\frac{v}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{v}{h}\right) \right] \right) \\
 &\quad \times \left| (h)^{-1} \left[K\left(\frac{v+x-\theta_1-y+\theta_2 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{v+x-\theta_1-y+\theta_2 h}{h}\right) \right] \right| \\
 &\quad f(v+x-\theta_1 h) du \\
 &= \int_{|v-\theta_1 h| \leq \delta} + \int_{|v-\theta_1 h| > \delta}
 \end{aligned}$$

Puisque $f(x)$ est continue, elle est bornée sur $I = [x - \delta, x + \delta]$. On suppose n assez grand pour que $x + v \pm \theta_1 h \in I$.

Par conséquent

$$\begin{aligned}
 \int_{|v-\theta_1 h| \leq \delta} &\leq \sup_{|v-\theta_1 h| \leq \delta} f(u-x+\theta_1 h) \int_{-\frac{\delta}{h}+\theta_1 \leq u \leq \frac{\delta}{h}+\theta_1} \left| K(u) - \frac{h^2}{2} \int s^2 K(s) ds K''(u) \right| \quad (3.17) \\
 &\quad \times \left| \left[K\left(\frac{x-\theta_1 h-y+\theta_2 h}{h} + u\right) - \frac{h^2}{2} \int s^2 K(s) ds k''\left(\frac{x-\theta_1 h-y+\theta_2 h}{h} + u\right) \right] (h)^{-1} du \right| \\
 &= \sup_{|v-\theta_1 h| \leq \delta} f(u-x+\theta_1 h) \int_{-\infty}^{+\infty} \chi_{-\frac{\delta}{h}+\theta_1 \leq u \leq \frac{\delta}{h}+\theta_1}(u) \left| k(u) - \frac{h^2}{2} \int s^2 K(s) ds K''(u) \right| \\
 &\quad \times \left| \left[K\left(\frac{x-\theta_1 h-y+\theta_2 h}{h} + u\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{x-\theta_1 h-y+\theta_2 h}{h} + u\right) \right] (h)^{-1} \right| du
 \end{aligned}$$

Pour chaque u

$$\lim_{n \rightarrow +\infty} \left| \left[k\left(\frac{x-\theta_1 h-y+\theta_2 h}{h} + u\right) - \frac{h^2}{2} \int s^2 k(s) ds k''\left(\frac{x-\theta_1 h-y+\theta_2 h}{h} + u\right) \right] (h)^{-1} \right| = 0$$

écrivez

$$\begin{aligned}
 & \left| \left[k\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) - \frac{h^2}{2} \int s^2 k(s) ds k''\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) \right] (h)^{-1} \right| \\
 &= \left| \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) \left[k\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) - \frac{h^2}{2} \int s^2 k(s) ds k''\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) \right] (h)^{-1} \right| \\
 &= \left| \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) \left[k\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) - \frac{h^2}{2} \int s^2 k(s) ds k''\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) \right] \right| \times \left| \frac{1}{x - \theta_1 h - y + \theta_2 h + hu} \right|
 \end{aligned}$$

on a

$$\left| \frac{1}{x - \theta_1 h - y + \theta_2 h + hu} \right| = \frac{1}{|x - y| \left| 1 - \frac{\theta_1 - \theta_2 - u}{x - y} h \right|}$$

Puisque $|u| \leq \frac{\delta}{h} + \theta_1$ on peut choisir δ assez petit pour que pour $n \geq n_0$ on ait

$$\left| \frac{\theta_1 - \theta_2 - u}{x - y} h \right| \leq \frac{3h + \delta}{|x - y|} = \eta < 1$$

Par conséquent

$$\left| \frac{1}{x - \theta_1 h - y + \theta_2 h + hu} \right| \leq \frac{1}{|x - y|(1 - \eta)} \tag{3.18}$$

puisque H_3 implique qu'il existe B tel que

$$\begin{aligned}
 & \left| \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) \left[K\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) \right] \right| \\
 & \leq B
 \end{aligned}$$

Ensuite nous avons

$$\begin{aligned}
 & \left| \left[K\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u\right) \right] (h)^{-1} \right| \\
 & \leq \frac{B}{|x - y|(1 - \eta)}
 \end{aligned}$$

$|k(u)|$ étant intégrable, par convergence dominée

$$\int_{|v-\theta_1 h| \leq \delta} \rightarrow 0 \text{ comme } n \rightarrow +\infty.$$

Laisser $\int_{|v-\theta_1 h| \leq \delta}$ écrivez-le sous la forme

$$\begin{aligned} \int_{|v-\theta_1 h| \leq \delta} &= \int_{|v-\theta_1 h| \leq \delta} \left| v(h)^{-1} \left[K\left(\frac{v}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{v}{h}\right) \right] \right. \\ &\quad \times (h)^{-1} \left[K\left(\frac{v+x-\theta_1-y+\theta_2 h}{h}\right) \right. \\ &\quad \left. \left. - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{v+x-\theta_1-y+\theta_2 h}{h}\right) \right] \right. \\ &\quad \left. f\left(\frac{x+v-\theta_1 h}{v}\right) \right| dv \end{aligned}$$

On a

$$\begin{aligned} \int_{|v-\theta_1 h| > \delta} &\leq \frac{2}{\delta - \theta_1 h} \sup_{|v-\theta_1 h| > \delta} \left| \frac{v}{h} \left[K\left(\frac{v}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{v}{h}\right) \right] \right| \quad (3.19) \\ &\quad \times \int_{|v-\theta_1 h| \leq \delta} (h)^{-1} \left[K\left(\frac{v+x-\theta_1-y+\theta_2 h}{h}\right) \right. \\ &\quad \left. - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{v+x-\theta_1-y+\theta_2 h}{h}\right) \right] \\ &\quad \times |f(x+v-\theta_1 h)| dv \end{aligned}$$

Soit le changement de variable défini par

$$v+x-\theta_1 h-y+\theta_2 h = u$$

Puis

$$\begin{aligned} \int_{|v-\theta_1 h| > \delta} &\leq \frac{2}{\delta - \theta_1 h} \sup_{|v-\theta_1 h| > \delta} \left| \frac{v}{h} \left[K\left(\frac{v}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{v}{h}\right) \right] \right| \quad (3.20) \\ &\quad \times \int_{\mathbb{R}} \left| (h)^{-1} \left[k\left(\frac{u}{h}\right) - \frac{h^2}{2} \int s^2 k(s) ds k''\left(\frac{u}{h}\right) \right] \right| f(u+y-\theta_2 h) du \end{aligned}$$

Lemme 3.1 (replaçant $K - \frac{h^2}{2} \int s^2 K(s) ds K''$ by $\left| K - \frac{h^2}{2} \int s^2 K(s) ds K'' \right|$) et (H_3) nous avons

$$\left| \int_{|v-\theta_1 h| > \delta} \right| \rightarrow 0 \text{ comme } n \rightarrow +\infty$$

et la convergence est uniforme. Ainsi, la preuve du Lemme 3.3.

Remarque 3.3 Si la condition C_2 est vérifiée, alors l'intégrale du côté droit de 3.20 devient

$$\begin{aligned}
 & \int_{\mathbb{R}} \left| (h)^{-1} \left[K\left(\frac{u}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u}{h}\right) \right] \right| \left| f(u+y-\theta_2 h) - f\left(\frac{u}{h}\right) + f\left(\frac{u}{h}\right) \right| du \\
 & \leq \int_{\mathbb{R}} \left| (h)^{-1} \left[K\left(\frac{u}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u}{h}\right) \right] \right| \int_{\frac{u}{h}}^{u+y-\theta_2 h} |f'(t)| dt du \\
 & \quad + \int_{\mathbb{R}} f\left(\frac{u}{h}\right) \left| (h)^{-1} \left[K\left(\frac{u}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u}{h}\right) \right] \right| du \\
 & \leq \int_{\mathbb{R}} \left| (h\lambda)^{-1} \left[K\left(\frac{u}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u}{h}\right) \right] \right| du \int_{\mathbb{R}} |f(x)| dx \\
 & \quad + \int_{\mathbb{R}} \left| K(u) - \frac{h^2}{2} \int s^2 K(s) ds K''(u) \right| f(u) du
 \end{aligned}$$

L'intégrale du côté droit de cette dernière inégalité d'où le Théorème est valide sous l'hypothèse C_2

3.7.3.2 Preuve du Théorème 3.5

Preuve.. Nous supposons que la condition C_1 est vérifiée. Soit $\Delta = [0, z] \times [0, z]$. Nous pouvons écrire

$$\begin{aligned}
 & \sum_{0 \leq i \neq j \leq \lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z}\right)^\alpha \left(1 - \frac{jh}{z}\right)^\alpha \int_{\mathbb{R}} \left| \left[K\left(\frac{u-ih}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-ih}{h}\right) \right] \right| \\
 & \quad \times \left| \left[K\left(\frac{jh-u}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{jh-u}{h}\right) \right] f(u) \right| du \\
 & = \int_{\{(x,y) \in \Delta : |x-y| > 0\}} \Phi_{n(x,y)} dx dy
 \end{aligned}$$

où

$$\begin{aligned}
 \Phi_n(x, y) & = \frac{1}{(h)^2} \sum_{0 \leq i \neq j \leq \lfloor \frac{z}{h} \rfloor} \chi_{\Delta_{h,i} \times \Delta_{h,j}}(x, y) \left(1 - \frac{ih}{z}\right)^\alpha \left(1 - \frac{jh}{z}\right)^\alpha \\
 & \quad \times \int_{\mathbb{R}} \left| \left[K\left(\frac{u-ih}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-ih}{h}\right) \right] \right| \\
 & \quad \times \left| \left[K\left(\frac{jh-u}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{jh-u}{h}\right) \right] \right| f(u) du
 \end{aligned}$$

si $(x, y) \in \Delta_{h,i} \times \Delta_{h,j}$ $i \neq j$ avec la représentation

$$x = hi + \theta_1 h \quad y = hj + \theta_2 h \quad 0 \leq \theta_l < 1, \quad l = 1, 2$$

On a

$$\begin{aligned}
 & \frac{1}{(h)^2} \left(1 - \frac{ih}{z}\right)^\alpha \left(1 - \frac{jh}{z}\right)^\alpha \int_{\mathbb{R}} \left[\left| K\left(\frac{u-x+\theta_1 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-x+\theta_1 h}{h}\right) \right| \right] \cdot 21) \\
 & \times \left[\left| K\left(\frac{u-y+\theta_2 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-y+\theta_2 h}{h}\right) \right| \right] (u) du \\
 & \leq \frac{1}{(h)^2} \int_{\mathbb{R}} \left[\left| K\left(\frac{u-x+\theta_1 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-x+\theta_1 h}{h}\right) \right| \right] \\
 & \times \left[\left| K\left(\frac{u-y+\theta_2 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-y+\theta_2 h}{h}\right) \right| \right] f(u) du
 \end{aligned}$$

Le côté droit de 3.21 tend vers zéro lorsque $n \rightarrow +\infty$ par le lemme 3.3.

Soit $\delta = \frac{z}{2}$. Écrivez

$$\begin{aligned}
 & \frac{1}{(h)^2} \int_{\mathbb{R}} \left[\left| K\left(\frac{u-x+\theta_1 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-x+\theta_1 h}{h}\right) \right| \right] \\
 & \times \left[\left| K\left(\frac{u-y+\theta_2 h}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u-y+\theta_2 h}{h}\right) \right| \right] f(u) du \\
 & = \int_{|v| \leq \delta} + \int_{|v| > \delta} .
 \end{aligned}$$

Ensuite nous avons

$$\begin{aligned}
 & \int_{\{(x,y) \in \Delta : |x-y| > 0\}} \Phi_n(x,y) dx dy \\
 & \leq \int_{\{(x,y) \in \Delta : |x-y| > 0\}} \sum_{0 \leq i \neq j \leq [\frac{z}{h}]} \chi_{\Delta_{h,i}} \times \chi_{\Delta_{h,j}}(x,y) \left(\int_{|v| \leq \delta} + \int_{|v| > \delta} \right)
 \end{aligned}$$

La preuve du reste est effectuée comme suit :

Considérons d'abord

$$\int_{\{(x,y) \in \Delta : |x-y| > 0\}} \int_{|v| \leq \delta}$$

Soit $A = \sup_{x \in [0,z]} f(x)$. Les notations étant comme dans la démonstration du Lemme 3.2 avec $\delta = \frac{z}{2}$, nous avons conformément à l'inégalité 3.17

$$\begin{aligned}
 \int_{|v| \leq \delta} & \leq A \int_{-\infty}^{+\infty} \chi_{-\frac{\delta}{h} \leq u \leq \frac{\delta}{h}} \int_{\mathbb{R}} \left| K(u) - \frac{h^2}{2} \int s^2 K(s) ds K''(u) \right| \\
 & \times \left[\left| K\left(\frac{x-\theta_1 h - y + \theta_2 h}{h} + u\right) \right. \right. \\
 & \left. \left. - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{x-\theta_1 h - y + \theta_2 h}{h} + u\right) \right] (h)^{-1} du.
 \end{aligned}$$

Pour chaque u

$$\lim_{n \rightarrow +\infty} \left| \left[K \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) \right] (h)^{-1} \right| = 0$$

on a

$$\begin{aligned} & \left| \left[K \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) \right] (h)^{-1} \right| \\ = & \left| \left[K \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{x - \theta_1 h - y + \theta_2 h}{h} + u \right) \right] \right. \\ & - \left[K \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right] \\ & \left. + \left[K \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right] \right| (h)^{-1} \\ \leq & \left(\lambda \left(\frac{2z}{h} \right) + \left| K \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right. \right. \\ & \left. \left. - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right| \right) (h)^{-1}. \end{aligned}$$

En outre,

$$\begin{aligned} & \left| K \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right| (h)^{-1} \\ = & \left| \frac{2z + x - \theta_1 - y + hu}{h} \right| \left| K \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right. \\ & \left. - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right| \frac{1}{|2z + x - y + hu|} \end{aligned}$$

Soit $B = \sup_{y \in \mathbb{R}} |y| |K(y) - \frac{h^2}{2} \int s^2 K(s) ds K''(y)|$ et $C = \sup_{y \in \mathbb{R}} |K(y) - \frac{h^2}{2} \int s^2 K(s) ds K''(y)|$, alors on a

$$\begin{aligned} & \left| \frac{2z + x - y + hu}{h} \right| \left| K \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right. \\ & \left. - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{2z + x - \theta_1 - y + \theta_2 h}{h} + u \right) \right| \\ \leq & B + 2hC. \end{aligned}$$

Par conséquent, nous avons

$$\begin{aligned} & \left| K\left(\frac{2z+x-\theta_1-y+\theta_2h}{h}+u\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{2z+x-\theta_1-y+\theta_2h}{h}+u\right) \right| (h)^{-1} \\ & \leq \frac{B+2hC}{|2z+x-y+hu|} \end{aligned}$$

Par conséquent

$$\begin{aligned} & \left| \left[K\left(\frac{x-\theta_1h-y+\theta_2h}{h}+u\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{x-\theta_1h-y+\theta_2h}{h}+u\right) \right] (h)^{-1} \right| \\ & \leq \lambda\left(\frac{2z}{h}\right) \frac{B+2hC}{|2z+x-y+hu|} \end{aligned}$$

Nous concluons que pour h assez petit

$$\begin{aligned} \int_{|v|\leq\delta} \frac{A}{|2z+x-y+hu|} \int_{\mathbb{R}} \left| K(u) - \frac{h^2}{2} \int s^2 K(s) ds K''(u) \right| (B+2hC) du & < \frac{AD}{|2z+x-y+hu|} \\ & \leq \frac{AD}{|2z+x-y+hu|} \\ & \leq \frac{AD}{(2z+x-y+hu)} \end{aligned}$$

Détant la borne finie de

$$\int_{\mathbb{R}} \left| K(u) - \frac{h^2}{2} \int s^2 K(s) ds K''(u) \right| (B+2hC) du$$

Enfin, nous avons

$$\int_{|v|\leq\delta} \leq \frac{AD}{(2z+x-y+hu)} + O(h)$$

Depuis

$$-\delta \leq hu \leq \delta \text{ on a } \frac{z}{2} \leq 2z+x-y+hu \leq \frac{7z}{2}.$$

D'où

$$\int_{|v|\leq\delta} \leq \frac{2AD}{Z} + O(h).$$

Par conséquent, par convergence dominée par Lebesgue, nous avons

$$\lim_{n \rightarrow +\infty} \int \int_{\Delta} \left(\int_{|v|\leq\delta} \right) dx dy = \int \int_{\Delta} \lim_{n \rightarrow +\infty} \left(\int_{|v|\leq\delta} \right) dx dy = 0 \quad (3.22)$$

Considérez alors $\int_{|v|\leq\delta}$.

Nous utilisons la seconde partie, par un raisonnement analogue, de la preuve du

lemme 3.2.

$$\begin{aligned} \int_{|v|>\delta} &\leq \frac{2}{\delta} \sup_{|v|>\delta} \left| \frac{v}{h} \right| \left| K\left(\frac{v}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{v}{h}\right) \right| \\ &\times \int_{\mathbb{R}} \left| (h)^{-1} \left[K\left(\frac{u}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{u}{h}\right) \right] \right| \\ &\times f(u + y - \theta_2 h) du. \end{aligned} \quad (3.23)$$

On a

$$\int_{|v|\leq\delta} \rightarrow 0, \quad n \rightarrow +\infty \text{ uniformément}$$

D'où

$$\lim_{n \rightarrow +\infty} \int_{\Delta} \int_{\mathbb{R}} \rightarrow 0, \quad n \rightarrow +\infty$$

puisque Δ est borné. La preuve du lemme est complète. ■

Remarque 3.4 Si C_2 est vérifié, alors le théorème est un gain valide, En effet, il suffit d'appliquer la remarque 3.3 à l'inégalité 3.23 .

3.7.3.3 Preuve du théorème 3.3

Preuve..Nous supposons que la condition (C_1) est satisfaite

$$\begin{aligned} nVar(p_n(z, \alpha)) &= E \left(\sum_{i=0}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z}\right)^\alpha \left[K\left(\frac{ih - X_j}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{ih - X_j}{h}\right) \right] \right)^2 \\ &- E^2 \left(\sum_{i=0}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z}\right)^\alpha \left[K\left(\frac{ih - X_j}{h}\right) - \frac{h^2}{2} \int s^2 K(s) ds K''\left(\frac{ih - X_j}{h}\right) \right] \right) \end{aligned} \quad (3.24)$$

puisque

$$\begin{aligned}
 & E \left[\sum_{i=0}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z} \right)^\alpha \left[K \left(\frac{ih - X_j}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{ih - X_j}{h} \right) \right] \right]^2 \quad (3.25) \\
 &= E \left[\left[\sum_{i=0}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z} \right)^\alpha \left[K \left(\frac{ih - X_j}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{ih - X_j}{h} \right) \right] \right]^2 \right] \\
 &= E \left[\left\{ \sum_{i=0}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z} \right)^{2\alpha} \left[K \left(\frac{X_k - ih}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{X_k - ih}{h} \right) \right]^2 \right. \right. \\
 &\quad \left. \left. + \sum_{i \neq j}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z} \right)^\alpha \left(1 - \frac{jh}{z} \right)^\alpha \left[K \left(\frac{X_k - ih}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{X_k - ih}{h} \right) \right] \right. \right. \\
 &\quad \left. \left. \times \left[K \left(\frac{X_k - ih}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{X_k - ih}{h} \right) \right] \right\} \right] \\
 &= \sum_{i=0}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z} \right)^{2\alpha} \int_{\mathbb{R}} \left[K \left(\frac{u - ih}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{u - ih}{h} \right) \right]^2 f(u) du \\
 &\quad + \sum_{i \neq j}^{\lfloor \frac{z}{h} \rfloor} \left(1 - \frac{ih}{z} \right)^\alpha \left(1 - \frac{jh}{z} \right)^\alpha \int_{\mathbb{R}} \left[K \left(\frac{u - ih}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{u - ih}{h} \right) \right] \\
 &\quad \times \left[K \left(\frac{jh - u}{h} \right) - \frac{h^2}{2} \int s^2 K(s) ds K'' \left(\frac{jh - u}{h} \right) \right] f(u) du.
 \end{aligned}$$

Il s'ensuit par corollaire (3.1), du lemme (3.2) et du théorème (3.5) que, comme $n \rightarrow +\infty$, le premier terme de la partie droite de (??) tend à $\left(\int_{\mathbb{R}} K^2(y) p(z, 2\alpha) \right)$ et le deuxième terme tend à zéro uniformément sur $[0, b]$.

en suite

$$n \text{Var}(p_n(z, \alpha)) \rightarrow \left(\int_{\mathbb{R}} K^2(y) p(z, 2\alpha) - p^2(z, \alpha) \right)$$

D'après le Lemme (3.1), nous avons

$$E(p_n(z, \alpha)) \rightarrow p(z, \alpha), \quad n \rightarrow +\infty$$

donc

$$E^2(p_n(z, \alpha)) \rightarrow p^2(z, \alpha), \quad n \rightarrow +\infty$$

Définir

$$\text{biais}(p_n(z, \alpha)) = E(p_n(z, \alpha)) - \text{Var}(p_n(z, \alpha))$$

nous avons

$$E(p_n(z, \alpha) - p(z, \alpha))^2 = \text{biais}^2(p_n(z, \alpha)) + \text{Var}(p_n(z, \alpha))$$

et

$$\left| \left(\int_{\mathbb{R}} K^2(y) p(z, 2\alpha) - p^2(z, \alpha) \right) \right| \leq \int_{\mathbb{R}} K^2(y) + 1$$

D'où

$$\text{Var}(p_n(z, \alpha)) = O\left(\frac{1}{n}\right)$$

Par le lemme (3.1), nous avons

$$|E(p_n(z, \alpha) - p(z, \alpha))| \rightarrow 0, \quad n \rightarrow +\infty$$

donc

$$\text{biais}^2(p_n(z, \alpha)) \rightarrow 0, \quad n \rightarrow +\infty$$

Par conséquent

$$E(p_n(z, \alpha) - p(z, \alpha))^2 \rightarrow 0, \quad n \rightarrow +\infty$$

Si la condition C_2 est satisfaite, le théorème est à nouveau valide, par le corollaire (3.1), du lemme (3.2) et en utilisant la remarque (3.4) du théorème (3.5) et théorème (3.2). ■

3.8 Simulations et Applications

3.8.1 Simulations

Dans cette section, nous effectuons une étude de simulation donnant l'erreur quadratique moyenne et la variance empirique de 50 échantillons de taille $n = 1000$ des trois estimateurs que nous avons comparés. Notre estimateur à noyau de réduction de biais et l'estimateur classique sont évalués par un noyau gaussien vérifiant les hypothèses $H_i, i = 1, \dots, 6$, en prenant $h = (n \log n)^{-1/2}$, nous avons calculé l'erreur quadratique moyenne mse_1 de $(p_{n,b,1}(z, \alpha), \dots, p_{n,b,50}(z, \alpha))$, mse_2 de $(p_{n,1}(z, \alpha), \dots, p_{n,50}(z, \alpha))$ et mse_3 de $(p_{n,1}^\lambda(z, \alpha), \dots, p_{n,50}^\lambda(z, \alpha))$ et les variances empiriques respectives σ_1^2, σ_2^2 et σ_3^2 pour différentes valeurs de (z, α) par les statistiques suivantes :

$$\widehat{p}_{n,b}(z, \alpha) = \frac{1}{50} \sum_{i=1}^{50} p_{n,b,i}(z, \alpha)$$

$$mse_1 = \frac{1}{50} \sum_{i=1}^{50} (p_{n,b,i}(z, \alpha) - p(z, \alpha))^2$$

TABLE 3.1 – Résultats de la simulation et estimation de l'indice de pauvreté pour loi normal

z	0.1	0.2	0.3	0.7
	$\alpha = 0$			
$\tilde{p}(z, \alpha)$	0.04783471	0.07536943	0.1039409	0.2554683
$p_n^\lambda(z, \alpha)$	0.05457776	0.08398905	0.121543	0.3003079
$\widehat{p}_{n,b}(z, \alpha)$	0.0478348	0.07536951	0.1039409	0.2554683
$\bar{p}(z, \alpha)$	0.03982784	0.07925971	0.1179114	0.2580363
e_1	0.008006876	-0.003890281	-0.01397055	-0.00256809
e_2	0.01474992	0.00472934	0.003631598	0.04227155
e_3	0.008006962	-0.0038902	-0.01397054	-0.002568077
	$\alpha = 1$			
$\tilde{p}(z, \alpha)$	0.02216492	0.03484911	0.06083344	0.1464611
$p_n^\lambda(z, \alpha)$	0.02791531	0.04363642	0.07555525	0.1820963
$\widehat{p}_{n,b}(z, \alpha)$	0.017132	0.03029619	0.05610549	0.1415566
$\bar{p}(z, \alpha)$	0.01993051	0.03976178	0.05939654	0.1341959
e_1	0.017132	-0.004912672	0.0014369	0.01226522
e_2	0.0079848	0.003874643	0.01615871	0.04790047
e_3	-0.00279851	-0.009465592	-0.003291053	0.007360727
	$\alpha = 2$			
$\tilde{p}(z, \alpha)$	0.01503913	0.02699854	0.03448113	0.09795609
$p_n^\lambda(z, \alpha)$	0.01882744	0.03446934	0.04314724	0.1228461
$\widehat{p}_{n,b}(z, \alpha)$	0.01049205	0.022468	0.03046886	0.09316492
$\bar{p}(z, \alpha)$	0.01329143	0.02654311	0.03971585	0.0908831
e_1	0.001747702	0.000455425	-0.005234724	0.007072992
e_2	0.005536004	0.007926231	0.003431385	0.03196297
e_3	-0.00279938	-0.004075108	-0.00924699	0.002281815

et

$$\sigma_1^2 = \frac{1}{50} \sum_{i=1}^{50} (p_{n,b,i}(z, \alpha) - \widehat{p}_{n,b}(z, \alpha))^2$$

De même, $\tilde{p}_n(z, \alpha)$ et $p_n^\lambda(z, \alpha)$, (mse_2, σ_2^2) et (mse_3, σ_3^2) sont respectivement calculés pour l'estimateur $p_n(z, \alpha)$ et $p_n^\lambda(z, \alpha)$ Les résultats sont présentés dans les tableaux suivants :

Pour une loi normal standard, les résultats sont présentés dans les tableaux 3.1 et 3.2.

loi de gamma Les cas d'études $p(z, 0), p(z, 1), p(z, 2)$ sont communément et respectivement appelés taux de pauvreté ou indice numérique, la profondeur de la pauvreté ou l'indice d'écart de pauvreté et la gravité de la pauvreté [foster et al (1984)]. Une comparaison de simulation les résultats montrent que pour de petits échantillons, chaque point z , notre estimateur à noyau de réduction de biais fournit une erreur beaucoup plus faible et variance pour les trois valeurs de α considérées. Ainsi, nous pouvons conclure que notre estimateur est recommandé.

TABLE 3.2 – Résultats de la simulation et estimation de l'indice de pauvreté pour loi Gamma

z	0.1	0.2	0.3	0.7
	$\alpha = 0$			
$\tilde{p}(z, \alpha)$	0.03804349	0.0558999	0.09818228	0.2155924
$p_n^\lambda(z, \alpha)$	0.03956232	0.05880223	0.1090826	0.2396744
$\widehat{p}_{n,b}(z, \alpha)$	0.03804363	0.05589997	0.09818248	0.2155924
$\bar{p}(z, \alpha)$	0.0327839	0.06449301	0.09516258	0.2081104
e_1	0.2010808	0.2010808	0.2010808	0.2010808
e_2	0.003778424	-0.005690789	0.01391998	0.03156394
e_3	0.00525973	-0.008593049	0.003019893	0.00748192
	$\alpha = 1$			
$\tilde{p}(z, \alpha)$	0.01811887	0.04208204	0.05771322	0.09497297
$p_n^\lambda(z, \alpha)$	0.02094148	0.04994295	0.06699422	0.1099849
$\widehat{p}_{n,b}(z, \alpha)$	0.01370974	0.03780112	0.05341871	0.09155327
$\bar{p}(z, \alpha)$	0.01648301	0.03260478	0.04837418	0.1080981
e_1	0.2010808	0.2010808	0.2010808	0.2010808
e_2	0.00445847	0.01733818	0.01862004	0.001886786
e_3	-0.002773275	0.00519634	0.005044528	-0.01654487
	$\alpha = 2$			
$\tilde{p}(z, \alpha)$	0.01106843	0.02425144	0.02584541	0.07329075
$p_n^\lambda(z, \alpha)$	0.01281576	0.02812273	0.0307454	0.08796304
$\widehat{p}_{n,b}(z, \alpha)$	0.007603438	0.02020249	0.02265838	0.06967952
$\bar{p}(z, \alpha)$	0.01101913	0.02185674	0.03251639	0.0734445
e_1	0.2010808	0.2010808	0.2010808	0.2010808
e_2	0.001796623	0.006265994	-0.001770992	0.01451854
e_3	-0.003415695	-0.001654243	-0.009858014	-0.003764984

CONCLUSION

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 étudié 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 Nous avons présenté l'estimateur de la densité de probabilité et la fonction de répartition.

L'estimation à noyau est une méthode non paramétrique basée sur l'utilisation d'une fonction appelée noyau et d'un paramètre de lissage h .

On remarque que le choix sur le noyau qui n'a pas une grande influence pour cette estimation, par contre le choix du paramètre de lissage a un impact important, et qui est en effet, beaucoup plus déterminant pour l'obtention des bons estimateurs.

Afin d'étudier les qualités statistiques des estimateurs obtenus, on les compare numériquement par simulation. Les résultats numériques montrent que : pour des petits échantillons, chaque point z , notre estimateur à noyau adaptatif fournit une erreur beaucoup plus faible pour les trois valeurs considérées. Ainsi, nous pouvons conclure que notre estimateur est recommandé pour les petits échantillons.

BIBLIOGRAPHIE

- [Ciss et al. (2015)] Ciss, Y. et, Dia, G. et, Diakhaby, A. (2015), Estimation non paramétrique de la distribution des revenus. et indice de pauvreté dans le contexte unidimensionnel avec $\alpha \in]0, 1[$. C.R. Acad.Sci. Paris, Ser. I., 353 (10), pp. 947-952
- [Dia (2008)] Dia, G. (2008), Estimation non paramétrique de la distribution des revenus et de l'indice de pauvreté. C. R. Acad.Sci. Paris, Ser. I 346, pp. 907-912
- [foster et al (1984)] Foster, J. E. et Greer, J. et, Thorbecke, E.(1984) Une classe de mesures de pauvreté décomposables. *Econometrica*, 52, p. 761-776
- [Kiefer (1961)] Kiefer, J.(1961) Sur les grands écarts de la distribution empirique de la variable de hasard vectorielle et de la loi du logarithme itéré. *Pacific J. Math.* 11, pp. 143-154
- [Lo et al (2009)] Lo, G.S., Sall, S.T., et Seck, C.T. (2009) Une théorie asymptotique générale des mesures de pauvreté. *C.R. Math. Rep. Acad. Sci. Canada.* 31 (2), p. 45-52
- [Parzen(1962)] Parzen, E.(1962) Sur l'estimation d'une fonction et d'un mode de densité de probabilité. *Annales de mathématiques Statistiques*, volume 33, numéro 3, p. 1065-1076
- [Seck(2011)] Seck, C. T.(2011) Estimation Non-paramétrique et Convergence Faible des Mesures de Pauvreté. université Pierre et Marie Curie.
- [Seck et al (2009)] Seck, C. T. et, Lo, G. S.(2009) Convergence uniforme des mesures de pauvreté non pondérées. *Communications en statistique Théorie et méthodes.* 38 (20), pp, 3697-3704

- [Seidl (1988)] Seidl, C.(1988) Mesure de la pauvreté : une enquête, dans : D. Bos, C. Seidl (Eds.),Welfare and Efficiency in Économie publique. Springer- Verlag, Heidelberg. pp. 71-147
- [Silverman (1986)] Silverman, B. W.(1986) Estimation de la densité pour les statistiques et l'analyse des données. Londres, Chapman & Salle
- [Xie et al (2014)] Xie, X. R. et Wu, J.J.(2014) Quelques améliorations sur les taux de convergence de l'estimateur de densité de noyau. Mathématiques appliquées, 5, 1684-1696.
- [Zakaria et al (2018)] Zakaria, Baradine; Ciss, Youssou, Diakhaby, Aboubakary. (2018) Estimation adaptative de la densité du noyau de la distribution des revenus et de l'indice de pauvreté.
- [Bosq et Lecoutre, 1967] D.Bosq, J.Lecoutre. Théorie de l'estimation fonctionnelle. Economicaedition, 1967.
- [Cencov, 1962] N. Cencov. Evaluation of unknow distribution density from observation. SovMaths,(3) :1559-1562, 1962
- [Parzen(1962)] E.Parzen. *On the estimation of a probability density function and mode.* Annals of Mathematical Statistics, 33 :1065-1027, 1962.
- [Pearson] K. Pearson. *On the systematic fitting of curvesto observations and measurements.*Biometrika, 1,265-303. 2, 1-23.
- [Rosenblatt(1956)] M.Rosenblatt. *Remarks on some non parametric estimates of a density function.* Annals of Mathematical Statistics, 27 :8332-837, 1956.
- [Hall et al(1999)] P.hall,R.Wol and Q.Yao. *Methods for estimating a conditional distrib function.*Journal of Americain statistical Association 94,p.154-163.(1999).
- [Saporta(1990)] G. Saporta. *Probabilités, Analyse des données et Statistiques.*Technip, Paris, 1990
- [Scott et el(1977)] D. Scott R. A. Tapia and J.R. Thompson. *Kernel density estimation revisited.* Nonlinear Analysis, Theory, Method and Applications, 1 :339-372, 1977.
- [Sen, A.K. (1976).] Poverty : An ordinal approach to measurement. Econometrica 44, 219-231.
- [Zheng,(1997).] Aggregate Poverty Measures. Journal of Economic Surveys 11(2), 123-162.

BIBLIOGRAPHIE

- [Chakravarty,(1983).] A New Index of Poverty. *Mathematical Social Sciences*, 6, 307-313.
- [Atkinson, (1987).] On the Measurement of Poverty. *Econometrica*, 55, 749-764
- [Foster et al. (1984)] A class of decomposable Poverty Measures,*Econometrica*, vol.52, N°3, 761-766.