N° d’ordre : 05/2004-M/MT

UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE
U.S.T.H.B

FACULTE DE MATHEMATIQUES

MEMOIRE

Présenté
Pour l’obtention du diplôme de

MAGISTER

EN MATHEMATIQUES
Spécialité : PROBABILITES-STATISTIQUES

Par : RAHMOUNE AHMED

THEME

ESTIMATION NON PARAMETRIQUE DE LA REGRESSION APPLICATION A LA PREVISION

Soutenue le : 10 / 02 / 2004 , devant le jury composé de :

K. BOUKHETALA, Professeur à l’U.S.T.H.B.	Président
M. BENTARZI, Professeur à l’U.S.T.H.B.	Directeur de thèse
A. AISSANI, Professeur à l’U.S.T.H.B.	Examinateur
O. ANES, Maître de conférences à l’I.N.P.S.	Examinateur
T. LARDJANE. T, Chargé de cours à l’U.S.T.H.B.	Examinateur

REMERCIEMENT

Monsieur **BOUKHETALA. K**, Professeur me fait le grand honneur de présider le jury de cette thèse, je tiens à lui exprimer ma profonde reconnaissance.

Monsieur **BENTARZI. M**, Professeur ayant suivi avec attention et intérêt toutes mes recherches, c'est grâce à ses conseils et ses encouragements que j'ai pu mener à bien ce travail, je tiens à lui exprimer ma profond gratitude .

Messieurs **AISSANI. A**, Professeur, **ANES. O**, Maître de conférence et **LARDJANE. T**, Chargé de cours ont accepté de faire partie de ce jury, je les remercie vivement.

DEDICACE

Je dédie ce travail personnel :
Contribution modeste dans le domaine de la statistique inférentielle non paramétrique

à

toute ma famille,
mes parents et mes enfants et à mes collègues enseignants.

.

RESUME

Le présent travail traite un sujet de la statistique inférencielle en utilisant les méthodes non

paramétrique (NP), et plus exactement l'estimation NP de quelques paramètres fonctionnels, et

application à la prévision dans les processus.

Il est successivement présenté:

► Des généralités sur l'estimation fonctionnelle.

► L'estimation NP de l'espérance conditionnelle, en choisissant une approche qui géneralise les méthodes connues du noyau et du régressogramme au travers de l'estimation de la densité conditionnelle.

► L'estimation NP de la médiane dans les deux cas unidimensionnel et multidimensionnel.

► L'estimation NP de la médiane conditionnelle par deux méthodes; celle du médianogramme et celle du noyau pour le cas univarié et le cas multivarié.

Il est à remarquer que les deux paramètres espérance conditionnelle et médiane conditionnelle

se confondent dans les populations statistiques de lois conditionnelles symétriques, d'où l'intéret de les étudier simultanement dans une classe de loi plus large.

Et en guise de conclusion, nous présentons une application de cette étude à la prévision dans

les processus, en utilisant deux prédicteurs (Prédicteur du médianogramme et prédicteur du noyau).

L'étude du présent thème a permis d'obtenir certains résultats concernant:

Les propriétés asymptotiques des estimateurs étudiés, les divers modes de convergences des différents estimateurs utilisés, ainsi que leurs performances et la confirmation de certains résultats théoriques, à l'aide des simulations.

Directeur deThèse:

M^r **Bentarzi Mohamed** Professeur à l'U.S.T.H.B

Chapitre 0

0.1 Introduction

La statistique non paramétrique a connu un développement trés intense ces deux dernières décennies, notamment l'estimation fonctionnelle.

Cette derniére a donné lieu à beaucoup d'articles.

L'évolution des résultats étant continuelle, cependant certains paramètres ont éte moins étudiés que d'autres.

Si l'estimation non paramétrique de la densité a été l'objet de nombreux travaux, depuis l'article de Rosenblatt(1956), la littérature concernant l'estimation non paramétrique des paramètres de régression est beaucoup moins abondante.

Parmi les paramètres les moins étudiées:

L'espérance conditionnelle et la médiane conditionnelle.

On s'est intéressé à l'étude de ces deux paramétres dans le cas des échantillons et dans le cas des processus (cas fortement mélangeant).

0. 2 Présentation de la thèse et Apport de l'étude

Le travail que je presente est une contribution modeste dans la théorie de l'inférence dans la statistique non paramétrique: Estimation des fonctions de régresion (Espérance conditionnelle et Médiane conditionnelle) et application à la prédiction.

Notre choix s'est porté sur ses méthodes non paramétriques car d'une part les méthodes paramétriques sont mal adaptées à l'estimation fonctionnelle (lorsque la loi est connue). En effet:

Les estimateurs sans biais sont rares et les méthodes bayésiennes sont difficiles à utiliser et d'autre part lorsque la loi est inconnu, les méthodes paramétriques ne peuvent pas répondre à ce genre de problème (l'estimation d'un paramétre fonctionnel, i. e: l'existence d'une infinité de paramètres).

Le travail qu'on présente se décompose en deux parties contenant successivement deux et quatre chapitres. La première partie est une introduction.

le premier chapitre traite les généralités sur l'estimation fonctionnelle: Propriétés et classification des paramètres, construction d'estimateurs et étude de la convergence. On a choisit une présentation dans le cadre générale: Les observations sont à valeurs dans un espace de Hilbert.

Le deuxiéme chapitre est réservé à l'estimation non paramétrique de:

$$r_\alpha(x) = E(Y^\alpha / X = x)$$

en particuliér $\alpha = 1$(dans les deux cas Y borné et Y non borné). On a choisit une approche qui généralise la méthode du noyau et la méthode du régressogramme, en passant par l'estimation non paramétrique de la densité conditionnelle.

La deuxième partie: est consacrée à l'estimation non paramétrique de la médiane, et de la médiane conditionnelle- paramètres trés peu étudiés- dans les deux cas univarié, et multivarié par deux méthodes: Méthode du noyau et méthode du médianogramme -cas d'un échantillon et cas d'un processus- et l'application de cette étude à la prédiction, et la comparaison avec le modèle paramétrique avec la déscription de la démarche de Box et Jenkins.

Les résultats obtenus sont résumés dans les points suivants:

♣ Dans l'étude des propriétés des paramètres:

Les paramètres étudiés ne sont pas sans biais il sont asymptotiquement sans biais, on l'a montré par des contres exemples, et par l'application d'une conséquence du **théorème de Bickel. Neimann**.

Les estimateurs utilisés ne sont pas stricts, et c'est l'un des inconvénients de cette méthode: Approcher un paramètre continu, lipschitzien par un etimateur fonction en éscalier.

♣ Dans l'étude de la convergence,on a suivi les étapes suivantes:

Le point de démarrage est le fameux théorème de **Glivenko. Cantelli** qui donne la convergence de μ_n mesure empirique vers la mesure théorique μ, et de $F_n$fonction de répartition empirique vers F fonction de répartition théorique ainsi celle de $F_n(\cdot/x)$fonction de répartition conditionnelle empirique vers $F(\cdot/x)$ puis déduir celle de $\hat{m}_n(x)$ estimateur de la médiane conditionnelle vers $m(x) = F^{-1}(\frac{1}{2}/x)$.

Pour l'étude de la convergence de $\hat{r}_n(x)$ estimateur de l'espérance conditionnelle vers $r(x)$, la réflexion est la suivante : Soit écrire $r(x)$ sous la forme d'un rapport $r(x) = \frac{\varphi(x)}{g(x)}$ où $\varphi(x) = \int yP_{X,Y}(dy)$ et $g(x) = \int P_{X,Y}(dy)$

($g(x)$ est la marginale de$X$), étudier séparément la convergence de $\varphi_n$estimateur du numérateur φ, puis celui de g_n estimateur du dénominateur g et ainsi déduire par composition d'applications continues, les différents modes de convergence de $\hat{r}_n$vers r.

Les propriétés de la convergence de l'estimateur de la densité marginale se déduisent de celle de $\hat{F}_n$.

Ou de mettre r sous la forme:

$$r(x) = \int yP_Y^{X=x}(x)$$

et étudier la convergence de la loi conditionnelle empirique$(P_Y^{X=x})_n$(estimateur de $P_Y^{X=x}$la loi conditionnelle théorique).

Les propriétés de convergence de

$$(P_Y^{X=x})_n \text{ vers } P_Y^{X=x}$$

se déduisent de celles de de$\hat{F}_n(\cdot/x)$ vers $F(\cdot/x)$. Parmi ses avantages elle permets de retrouver les méthodes du noyau et celle du régressogramme. D'autre part afin d'évaluer la vitesse de convergence on a appliqué l'inégalité des grandes déviations (**inégalité de Bernstein. Fréchet**).On a Déduit et confirmer que la vitesse de convergence par la méthode du noyau est plus grande que celle de la méthode empirique: celle de l' histogramme pour la densité, régressogramme pour l'espérance conditionnelle et le médianogramme pour la médiane conditionnelle.

♣ On a appliqué le **Théorème deBochner** pour évaluer:

$$\lim_{n\to\infty} ET_n(x)$$

et

$$\lim_{n\to\infty} VT_n(x)$$

Ainsi on déduit la convergence en $m.\ q$ de T_n vers T où T est le paramètre étudié et T_n est l'estimateur utilisé. Et par l' application du **théorème de Borel Cantelli:**

$$p.co \Rightarrow p.s$$

avec un choix convenable du terme majorant, on a montré la convegence forte**.** La formule des probabilités totales, le developpement limité ainsi que les propriétés de la compacité de l'espace ont été largement utilisés.

♣ L'étude dans le cas des processus a été étudié et particulièrement dans le cas des **mélangeances**.

♣ On a confirmé la convergence des prédicteurs, et la comparaison avec les méthodes paramétriques (déscription de la démarche de **Box et Jenkins** en quatre étapes) nous à permis de déduir et de confirmer que les méthodes non pramétriques sont mieux adaptées que les méthodes paramétriques, par leurs facilités dans la pratique

0. 3 Notations et Abréviations:

$[x]$: partie entère de x

∂A : frontière de A

∇^2 : laplacien

$\circ,\ 0$: notation de Landau petit o, et grand o

$\mathcal{B}_E$: tribu borélienne de E

$L^p_\mu(E,\ F)$: espace des classes d'équivalence des applications de E dans F de $p^{i\grave{e}me}$puissance intégrable, par rapport à la mesureμ

1_A : fonction indicatrice de A

δ_a: mesure de Dirac au point a

$\sigma(X)$: Tribu engendrée par la v. a X

$f_\mu,\ F_\mu$ densité de la loi μ(resp . fonction de répartition de la loi μ)

$E(Y/X=x),\ E^XY$ espérance de Y conditionnelle à X

$V(Y/X=x)$: variance de Y conditionnée par $X=x$

$\mu*\nu,\ f*g$: produit de convolution des mesures μ et ν(resp des applications f et g)

$\hat{\mu},\ \hat{f}$: Transformée de Fourier de la mesure μ(resp. densité f)

Abréviations:

v. a. r: variable aléatoire réelle

e. m. v: estimateur du maximum de vraisemblance

e. s. b: estimateur sans biais

i. i. d: indépendantes et identiquement distribuées

d° : degré de l'estimateur

E. Q. I: erreur quadratique intégrée.

S. O: Statistique d'ordre.

0. 4 Symboles et renvois

Symboles:

$\mathbb{N}^*=\mathbb{N}-\{0\}$

$\overline{\mathbb{R}}=\mathbb{R}\cup\{\infty\}$

$\mathbb{C}$: Esemble des nombres complexes

Renvois:

1. 2. 3: signifie chapitre1 section 2, sous section 3.

TABLE DES MATIERES

Contents

Chapter 1

Généralités sur l'estimation fonctionnelle

1.1 Etude Préliminaire

1.1.1 Introduction et Motivation

Introduction:

On présente les notions de base de la théorie de la statistique non paramétrique et en particulier la théorie de l'estimation fonctionnelle:

On énonce des théorèmes importants, utiles pour la suite et qui vont nous permettre de répondre à certaines questions telles que: L'existence d'estimateurs sans biais, résolue en appliquant un corollaire du théorème de Bickel-Lehmann, amélioré par Bosq (1970). La question de l'existence d'estimateurs convergents. La question de classification des estimateurs et l'optimalité traité parRao-Blackwell et Lehmann-Scheffé.

On a privilégie dans cette présentation le cadre générale:

Les observations sont à valeurs dans $(\mathbb{H},\ \mathcal{B}_{\mathbb{H}})$: Espace de Hilbert séparable muni de son produit scalaire$<,\ >_H$et de sa norme$\|\cdot\|_H$ où $\mathcal{B}_{\mathbb{H}}$désigne sa tribu borélienne. Le cas $\mathbb{L}^2\ (\mathbb{R},\ \mathcal{B}_{\mathbb{R}},\ \lambda)$ espace des classes d'équivalence des applications réelles de carré intégrable muni de son produit scalaire usuel $< f,\ g >= \int_{\mathbb{R}} f\ (x)\ g\ (x)\ dx$ et λ la mesure de Lebesgue, en fait partie.

La notion d'intégration au sens de Bochner-intégration forte- (Intégrale au sens de Pettis-Intégrale faible) des variables aléatoires à valeurs dans un Banach s'impose (Voir annexe).

Remarque: La validité des théorèmes importants tel que les théorèmes de Lebesgue, de Fubini dans les Banach. Ces derniers sont d'une grande utilité lors de la généralisation des propriétes à des espaces de Hilbert. Voir: Hoffmann-Jorgensen (1976) ou (Grenander (1981) pour une étude détaillée des Probabilités dans les Banach.

Motivation:

Cette démarche: Le choix de présenter les notions de base de l'estimation non-paramétrique dans les Banach et en particulier dans les Hilbert, est motivée par :

1) l'identification d'un processus à trajectoire continue à une variable aléatoire à valeur dans un Banach tel que par exemple $[C[0,1], \|\cdot\|_u]$: Espace des fonctions réelles continues sur $[0,1]$ muni de sa norme de la convergence uniforme.

2) La correspondance entre la notion d'estimation d'un paramètre de régression et la prévision.

Ce qui conduit par transitivité à l'étude de la prévision dans les Banach.

Shématiquement:

$$\begin{array}{c} \text{Estimation de la régression} \leftrightarrows \text{Prévision.} \\ \text{Processus} \leftrightarrows \text{v. a. à valeur dans } C[0,1] \\ \Rightarrow \text{Prevision dans } C[0,1] \end{array}$$

L'analyse statistique d'un processus à temps continu observé sur des intervalles de temps sccessifs de même longueur est facilitée par l'interprétation des observations comme des réalisations de vriablles aléatoires à valeurs dans un espace fonctionnel par exemple $C[0,1]$ ou $D[0,1]$ (Espace des fonctions c. a. d, l. a. g: fonctions continues à droite, possédant une limite finie à gauche), cette technique connue est peu utilisé.

Parmi les recherches actuelles dans ce sens: Modélisation d'un processus à trajectoire continue par un $AR[p]$ Hilbertien noté $ARH[p]$ (processus autorégressifs d'order p dans$H$)

Remarque: Beaucoup de questions sont encore ouvertes.

1.1.2 Définitions préliminaires et Exemples

Définitions:

Un modèle statistique est la donnée d'un triplet $(\mathbb{E}, \mathcal{A}, \mathcal{P})$ où:

$\mathbb{E}$ est un ensemble appelé espace des observations, $\mathcal{A}$ une tribu sur $\mathbb{E}$, $\mathcal{P}$ une famille de mesures de probabilité sur $(\mathbb{E}, \mathcal{A})$.

Remarque: $\mathcal{P}$ peut s'écrire sous la forme $(P_\theta, \theta \in \Theta)$, où Θ est l'espace des paramètres.

L'ensemble des variables observées est résumé par une variable aléatoire X à valeurs dans $\mathbb{E}$ et de loi inconnue P appartenant à $\mathcal{P} = (P_{\theta,}\theta \in \Theta)$.

Soit g une application de $\mathcal{P}$ dans un ensemble quelconque Θ' (généralement un espace fonctionnelle, un espace de Hitbert H, ou un espace de Banach séparable).

Estimer $g\ (P)$ c'est chercher à l'évaluer au vu d'une réalisation de X i, e d'un élément de $\mathbb{E}$ tiré au hasard suivant P. Ainsi la règle choisie pour procéder à cette estimation est donc une application de $\mathbb{E}$ dans Θ'; on dit que c'est un estimateur T du paramètre $g\ (P)$.

On appelle souvent paramètre l'application $g\ :\ \mathcal{P} \mapsto \Theta'$ ou même l'application $\theta \mapsto g\ (P_\theta)$ quand $\mathcal{P} =\ (P_\theta,\ \theta \in \Theta)$.

Les Méthodes Non Paramétriques

Modèle non paramétrique:

Lorsque Θ est "vaste" on dit que le modèle $(\mathbb{E},\ \mathcal{A}, (P_\theta)\ _{\theta\in\Theta})$ est non paramétrique, expression discutable mais qui est devenue classique. Il n'y a pas de définition précise d'un modèle non paramétrique car la frontière entre paramétrique et non paramétrique est assez floue. Il y a d'autre modèle p. ex le modèle semi paramétrique dont la définition est moin précise, pour fixer les idées, on peut considérer la définition suivante:

Définition:

Le modèle est non paramétrique si Θ contient un ensemble convexe de dimension infinie, il est paramétrique si Θ est un ouvert de $\mathbb{R}^s$, $(s \geq 1)$.

A ces modèles sont associés les méthodes d'estimation non paramétriques.

On considére essentiellement le cas ou g prend ses valeurs dans un espace vectoriel topologique de dimension infinie, autrement dit un espace de fonctions, on dit alors que $g\ (P)$ est un paramètre fonctionnel.

Exemples:

A) Cas d'un échantillon

Soit $X =\ (X_1,\ .\ .\ .\ ,\ X_n)$ où les X_i , $1 \leq i \leq n$, sont des $v.\ a.$ à valeurs dans $\mathbb{R}^s$ muni de sa tribu borélien

$\mathcal{B} = \mathcal{B}_{\mathbb{R}^s}= \mathcal{B}_{\mathbb{R}}^{\bigotimes s}$ ($\bigotimes$désigne le produit tensoriel) -tribu engendrée par les pavés de $\mathbb{R}^s$-Remarquer la derniére égalité est dû à la séparabilité de $\mathbb{R}^s$-

On suppose en outre que les. $(X_i)\ _{i=1,\ldots,n}$est un échantillon de taille n de loi $(\mu)\ ^n$où $\mu \in \mathcal{P}_o$avec $\mathcal{P}_o$ une famille de probabilités sur $\mathbb{R}^s$ et, par conséquent le modèle statistique est:

$$(\mathbb{E}, \mathcal{A}, \mathcal{P}) = (\mathbb{R}^{sn}, \mathcal{B}^n, (\mu^n)_{\mu\in\mathcal{P}_o})$$

avec: $\mathcal{B}^n =$ tribu produit $\mathcal{B} \bigotimes .\ .\ .\ \bigotimes \mathcal{B}$ (n fois) et $\mu^n = \mu^{\bigotimes\ n}$mesure produit de μ par μ, n fois

Exemple 1:

La fonction de répartition $(f.\ d.\ r)$, la fonction définie par:

$$F_\mu(x_1, .., x_s) = \mu \left\{ \prod_{i=1}^{s}]-\infty, x_i] \right\} ; (x_1, ..., x_s) \in \mathbb{R}^s$$

(pour une réalisation $X_i,\ 1 \leq i \leq n$), le paramètre g est l'application de $\mathcal{P}$ dans l'espace fonctionnelle $\Theta' = D_b\ (\mathbb{R}^s)$ définie par: $\mu \mapsto F_\mu$, où $D_b\ (\mathbb{R}^s)$ est l'espace des fonctions réelles définies sur $\mathbb{R}^s$ bornées, continues par morceaux et n'admettant que des discontinuités de premiére éspèce.

Exemple 2:

la fonction caractéristique $(f.\ c.\ )$, définie par:

$$\mu(t) = E_\mu \exp\left[i < t, X_1 >\right], t \in \mathbb{R}^{s.}$$

(au vu d'une réalisation X_i de $\mathbb{R}^s$, i fixé, soit $i = 1$ car les X_i ont même lois) μ est un paramètre à valeurs dans $C_b\ (\mathbb{R}^s)$: espace des fonctions réelles ou complexes définies sur $\mathbb{R}^s$continues et bornées.

Exemple 3:

Pour $s = 1$, la fonction quantile d'ordre α définie par:

$$F_\mu^{-1}(\alpha) = \inf \{t \in \mathbb{R} : F_\mu(t) \geq \alpha\}, 0 < \alpha < 1$$

F_μ^{-1}est un paramètre à valeurs dans Θ' l'espace des fonctions réelles définies sur $]0.1[$ monotones non décroissantes et continues à gauche (pour $\alpha = \frac{1}{2}$, on retrouve la médiane)

Exemple 4:

La densité et ses dérivées.

si $\mathcal{P}_o$ est dominé par λ, $\frac{d\mu}{d\lambda}$ est un paramètre à valeurs dans L^1.

Lorsque$\frac{d\mu}{d\lambda}$ admet une version f_μ bornée (resp. continue et bornée) on peut la considérer comme un paramètre à valeurs dans L^2 (resp. $C_b\ (\mathbb{R}^s)$). enfin, si f_μ est différentiable, on définit de nouveaux paramètres fonctionnels: les dérivées partielles de la densité.

Exemple 5: les paramètres de régression:

Posons $X_i = (X_i^{(1)},\ X_i^{(2)}),\ 1 \leq i \leq n$ où $X_i^{(1)}$ (resp. $X_i^{(2)}$) prend ses valeurs dans $\mathbb{R}^{s_1}$ (resp. $\mathbb{R}^{s_2}$), $s_1 + s_2 = s$, et soit $(\mu^x_{X_i^{(2)}},\ x \in \mathbb{R}^{s_1})$ une famille de versions spécifiées des lois de $X_i^{(2)}$sachant que $X_i^{(1)} = x$. Toute fonction de la forme:

$$\mathbf{x} \longmapsto \mathbf{r}\ (\mu^x_{X_i^{(2)}})$$

est un paramètre de régression. les plus importants sont la **densité conditionnelle, l'espérance conditionnelle (la "régression"), la médiane conditionnelle** et la fonction frontière. -

La fonction frontière elle est définie pour $s_2 = 1$ par:

$$x \longmapsto M\ (x,\ \mu)$$

où $M\ (x,\ \mu)$ désigne la borne supérieure du support de $\mu^x_{X_i^{(2)}}$, quand $X_i^{(1)}$suit la loi μ, ce support étant supposé compact pour tout $\mu \in \mathcal{P}_o$et tout $x \in \mathbb{R}^{s_1}$.

B). Cas d'un processus

Remarque: Les paramètres précédents sont encore définis quand $X_1, \ldots\ldots, X_n$ sont corrélés mais de même loi. voici d'autres exemples de paramètres fonctionnels associés à l'observation d'un processus:

Exemple 1:

La densité spectrale d'un processus réel $(X_t,\ t \in \mathbb{Z})$ du second ordre, faiblement stationnaire, définie par:

$\varphi(\lambda) = \frac{1}{2\pi} \sum_{t\in\mathbb{Z}} cov(X_o,\ X_t) e^{i\lambda t},\ \lambda \in [-\pi, +\pi]$ dés que $\sum_{t\in\mathbb{Z}} |cov(X_o, X_t)| < +\infty$.

Exemple 2:

Soit $(X_t,\ t \in \mathbb{R}_+)$ un processus de diffusion solution de l'équation différentielle stochastique

$$dX_t = m\ (X_t,\ t)\ dt + \sigma\ (X_t,\ t)\ dB_t$$

avec

$$B_t = W_t - tW_1$$

où W_t est le processus de Wiener et $(B_t,\ t \in \mathbb{R}_+)$ est un pont Brownien. la dérivé de (X_t) est le paramètre fonctionnel défini par

$$(x,\ t)\ \mapsto m\ (x,\ t)$$

Remarque: En combinant les exemples précédents on peut construire de nouveaux paramètres, les paramètres composés, en particulier:

Exemple 3: Le taux de panne défini sur $\mathbb{R}$ par

$$x \mapsto \frac{f_\mu(X)}{1 - F_\mu(X)}$$

Exemple 4: la dérive d'un processus de diffusion vérifiant certaines conditions de régularité permettant de l'écrire:

$$m\ (x)\ = \sigma^2 \frac{f'(x)}{f_\mu(x)}$$

$x \in \mathbb{R}$. où σ est supposé constant.

1.2 Propriétés et Méthodes d'estimations

1.2.1 Classification des paramètres fonctionnels et Etude de l'optimalité

Les exemples précédents montrent que les paramètres fonctionnels peuvent prendre des formes variées. Il est cependant possible de regrouper certain paramètres en tenant compte de leurs propriétés communes. Ceci permet de dégager les méthodes naturelles d'estimation pour chaque classe de paramètres ainsi définie. Dans ce qui suit on présente ces propriétés avec des exemples.

a) Homogénéité et localisation.

Soit $(\mathbb{E},\ \mathcal{B},\ \mathcal{P})$ un modèle statistique où $\mathbb{E}$ est un espace métrique muni de sa tribu borélienne et soit g un paramètre à valeurs dans un espace vectoriel H. pour $p \in \mathcal{P}$, on note:

$P_B\ (A)\ = P\ (A \cap B);\ A \in \mathcal{B},\ B \in \mathcal{B}$

$P^B\ (A)\ = P\ (A \cap B)\ /P\ (B);\ A \in \mathcal{B},\ B \in \mathcal{B},\ P\ (B)\ > 0.$

Définitions

Définition 1: on dit que le paramètre g est homogène sur le fermé F de $\mathbb{E}$ si: Pour tout ouvert U contenant F et pour tout P et Q dans$\mathcal{P}$ on a:

$$g(P) = 0 \text{ si } P(U) = 0 \quad (1)$$

$$\frac{g(P)}{P(U)} = \frac{g(Q)}{Q(U)} \text{ si } P(U)Q(U) > 0 \text{ et } P^U = Q^U. \ (2)$$

Définition 2: On dit que g est localisé par le fermé Fsi, pour tout ouvert U contenant F, (1) est vérifié et

$$g(P) = g(Q) \text{ si } P_U = Q_U \qquad (2^{'})$$

Définition 3: On dit que g est **local** s'il est localisé par un point; g est **global** s'il n'est localisé que par $\mathbb{E}$.

Remarquons qu'un paramètre homogène sur F est localisé par F, la réciproque étant fausse.

Exemples:

Exemple 1: Le paramètre $g\ (P)\ = P\ (F)$ est homogène sur le fermé F. -

Exemple 2: Si une version F_μ de $\frac{d\mu}{d\lambda}$ est continue au voisinage d'un point a de $\mathbb{R}^s$,

le paramètre $g\ (P)\ = f_\mu\ (a)$ est homogène sur le fermé $\{a\}$ et local.

Exemple 3: $g\ (P)\ = F_\mu^{-1}\ (p_o)$ est global $(p_o \in]0.1[)$.

Exemple 4: Pour un choix convenable des $\mu^x_{X_i^{(2)}}$ la fonction frontière en un point a de $\mathbb{R}^{s_1}$est localisée par la droite $x = a$ mais n'est pas homogène.

b) Paramètres intégraux et leurs limites.

Définition:

Soit g un paramètre fonctionnel à valeurs dans un espace $\mathcal{F}\ (\mathbb{E}_1,\ \mathbb{C}) = \mathbb{C}^{\mathbb{E}_1}$: l'espace des applications de $\mathbb{E}_1$dans $\mathbb{C}$

Définition: On dit qu'il est intégral de noyau K s'il existe $K \in \mathcal{F}\ (\mathbb{E} \times \mathbb{E}_1,\ \mathbb{C})$ tel que:

$$(\forall t \in \mathbb{E}_1),\ K\ (.,\ t)\ \in \bigcap_{p \in P} L^1_P\ (\mathbb{E}\times\mathbb{E}_1,\ \mathbb{C}) \text{ et } g\ (P)\ \ (t)\ = \int_{\mathbb{E}} K\ (x,\ t)\ dP\ (x);\ t \in \mathbb{E}_1,\ P \in \mathcal{P}$$

Exemples:

Exemple 1: Soit F_μ La fonction de répartition de la loi μ. F_μ est un paramètre intégral puisque:

$$F_\mu(t_1, .., t_s) = \int_{\mathbb{R}^s} 1_{s_{\prod_{i=1}^s]-\infty,\ t_i]}}(x) d\mu(x)$$

Exemple 2: La fonction caractéristique est un paramètre intégral de noyau

$$(x_1,\ .\ \ .\ \ .\ \ .\ ,\ \ x_n,\ t)\ \ \rightarrow e^{i<x_1,t>}.$$

De nombreux paramètre fonctionnels sont limites d'une suite de paramètres intégraux (le mode de convergence étant à préciser). La densité est l'exemple le plus important d'un tel paramètre. .

Comparaison d'estimateurs

Soit g un paramètre à valeurs dans H espace vectoriel topologique muni de sa tribu borélienne $\mathcal{B}$ et soit $\mathcal{T}$une famille d'estimateurs mesurables de g. Afin de comparer l'efficacité des éléments de $\mathcal{T}$ on utilise une relation de préférence,selon un sens qu'on précisera:

$$\prec : S \prec T$$

se lit "S est préférable à T".

Si g est à valeurs réelles il est usuel de mesurer l'efficacité de l'estimateur T au point de θ de Θ par son erreur quadratique:

$$E_\theta \left|T - g(\theta)\right|^2 \in \mathbb{R}^+ \cup \{+\infty\}$$

on en déduit la relation de préférence:

$$S \prec T \Leftrightarrow (\forall \theta \in \Theta), E_\theta \left|S - g(\theta)\right|^2 \leq E_\theta \left|T - g(\theta)\right|^2 . \ (3)$$

Plus généralement, si H est un espace de Hilbert séparable muni de son produit scalaire $< . , . >_H$ et de sa norme $\|.\|_H$on définit $\prec$ par:

$$S \prec T \Leftrightarrow (\forall y \in H), (\forall_\theta \in \Theta),\ E_\theta \left| < S - g(\theta), y >_H \right|^2 \leq E_\theta \left| < T - g(\theta), y >_H \right|^2 \quad (4)$$

Ce qui signifie que:

$$S \prec T \Leftrightarrow (\forall y \in H, < S,\ y >_H \prec < T,\ y >_H$$

pour estimer le paramètre réel $< g\ (\theta),\ \ y >_H$.

Remarques:

Remarque 1: La comparaison d'estimateurs à valeur dans un espace de Hilbert se ramene à la comparaison dans un espace réel. Quand les éléments de T sont du second ordre i. e:

$$(\forall t \in T), (\forall_\theta \in \Theta), E_\theta \|T\|_H^2 < +\infty)$$

(4)$\Leftrightarrow$ à:

$$S \prec T \Longleftrightarrow (\forall_\theta \in \Theta), c^\theta_{S-g(\theta)} \prec\prec c^\theta_{T-g(\theta)} \quad (5)$$

où c^θ_Y désigne l'opérateur de covariance de Y pour P_θ et $\prec\prec$ la relation d'ordre usuelle entre opérateurs symétriques d'un espace de Hilbert.

Remarque 2: IL y a une relation de préférence moins précise mais plus maniable qui est définie par:

$$S \prec' T \Longleftrightarrow (\forall_\theta \in \Theta),\ E_\theta\ (\|S - g(\theta)\|_H^2)\ \leq E_\theta\ (\|T - g(\theta)\|_H^2)\ (6)$$

Dans le cas particulier important où:

$$H = L^2_\lambda$$

λ la mesure de Lebesgue, la fonction de risque associée à $\prec'$ s'appelle l'erreur quadratique intégrée $(E.\ Q.\ I.\ )$, car le théorème de Fubini permet de l'écrire sous la forme:

$$J\ (T,\ \theta)\ = \int E_\theta \left[T(t) - g(\theta)(t)\right]^2 d\lambda\ (t).\ (7)$$

Remarque 3: Quand H est un $e.\ v.\ t.$ dont le dual topologique H^* est suffisamment riche, on peut encore définir $\prec$ en remplaçant dans (4) " $(\forall y \in H)$ " par " $(\forall y \in H^*)$ ".

3) Que l'on choisisse $\prec$ ou$\prec'$ il n'est pas possible en général d'obtenir un estimateur optimal (i. e. préférable à tout élément de $\mathcal{T}$) si $\mathcal{T}$ est trop vaste. On a cependant les résultats suivants:

Etude de l'optimalité.

Théorème (Rao-Blackwell et Lehmann-Scheffe).

Soit g un paramètre à valeurs dans l'espace de Hilbert séparable H;soit T un estimateur du second ordre et U une statistique exhaustive. Alors:

$$S = E_\theta(T/U) \text{ est un estimateur préférable à } \ddot{}T \text{ pour } \prec \text{ et } \prec'$$

Remarque: S est un fonction de U. Si de plus U est complète, S est alors optimal dans la famille$\mathcal{T}_T$des estimateurs de $g\ (\theta)$ qui sont du second ordre et meme espérance que T pour tout θ.

Preuve: Comme U est exhaustive, il existe une version de la loi conditionnée par U qui est indépendante de θ, donc Sest bien un estimateur, il est clair qu'il suffit d'établir les résultats annoncés pour $\prec$et pour g à valeurs réelles, d'après la définition (4) et en remarquant que:

$$< E\ (T/U),\ y >_H = E\ (< T,\ y >_H /U),\ y \in H.$$

or:

$$(\forall \theta \in \Theta),\ E_\theta \left| T - g(\theta) \right|^2 = E_\theta \left| S - g(\theta) \right|^2 + E_\theta \left| T - S \right|^2,$$

donc $S \prec T$. maintenant si $T_1 \in \mathcal{T}_T,\ S_1 = E_\theta\ (T_1/U)\ \in \mathcal{T}_T$ est préférable à T_1. alors comme $E_\theta\ (S_1 - S)\ = 0,\ \theta \in \Theta$, et que $S_1 - S$ est une fonction de U, la complétude de U entraine $S_1 - S = 0$, donc S est optimal dans $\mathcal{T}_T$. ■

Remarque: En pratique pour montrer qu'une statistique (estimateur$U$) n'est pas compléte il suffit d'éxiber $\varphi \neq 0$: $E\varphi\ (U)\ = 0$, exple: La statistique d'ordre n'est compléte.

On rappelle: U est compléte: $\varphi\ (U)$ réelle, $E_\theta\ (\varphi\ (U)\ )\ = 0 \Rightarrow \varphi\ (U)\ = 0\ p_\theta\ \ p.\ \ s$ pour tous θ.

Conclusion: Le meilleur estimateur sans biais du paramètre $g\ (\theta)$ est celui qui est fonction d'une statistique **exhaustive**, si de plus la statistique est **compléte** elle réalise **l'optimalité.**

Remarque: Dans les modèles non paramètriques les statistiques exhaustives non triviales sont rares. si $X = (X_1, \ldots, X_n)$ où les X_i forment un échantillon de variables réelles, la statistique exhaustive la plus importante est la statistique d'ordre(S. O)

$$U\ (x_1, \ldots, x_n)\ =\ (X_{1,n}, \ldots, X_{n,n})$$

avec $X_{k,n}$, $1 \leq k \leq n$ est la k-iéme S. O.

U engendre la tribu $\mathcal{B}_U$ des boréliens symétriques (invariants par permutation des coordonnées) et les fonctions $\mathcal{B}_U-$ mesurables sont les fonctions mesurables symétriques.

Considérons alors les égalités:

$$\mu^{\bigotimes n}(A \cap B) = \int_B 1_A(x_1, ..., x_n) d\mu(x_1)...d\mu(x_n) =$$

$$\int_B \left[\frac{1}{n!} \sum_{\sigma \in S} 1_A(x_{\sigma(1)}, ..., x_{\sigma(n)}) \right] d\mu(x_1)...d\mu(x_n).$$

$A \in \mathcal{B}_{\mathbb{R}} n,\ B \in \mathcal{B}_U,\ \mu \in \mathcal{P}_o$; où S désigne l'ensemble des permutations des n premiers entiers.

On constate que:

$$\frac{1}{n!} \sum_{\sigma \in S} \delta_{(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}$$

est la loi conditionnée par U pour tout $\mu \in \mathcal{P}_o$: U est bien exhaustive dés que $\mathcal{P}_o$ est suffisamment vaste, U est également complète; c'est le cas si $\mathcal{P}_o$ contient les lois uniformes par morceaux, le théorèmed'optimalité de Rao-Blackwell permet alors d'affirmer qu'il existe un seul estimateur T du second ordre, symétrique et d'espérance $\theta \to E_\theta\ T$ fixée.

Remarque: La plupart des résultats précédents s'étendent au cas où H est un espace de Banach séparable

(espace polonais)

1.2.2 Estimation sans biais et Estimation minimax.

Soit $(\mathbb{E},\ \mathcal{A},\ \mathcal{P})$ un modèle statistique et g un paramètre à valeurs dans l'espace de Banach séparable$H$. soit T un estimateur de 1er ordre (i. e. T est mesurable et $(\forall p \in P),\ \ E_p \|T\|_H < \infty$).

Définition 1: On dit que T est un estimateur sans biais de $g(p)$ si:

$$(\forall p \in \mathcal{P}), E_p T = g(p).$$

Remarque: Si g est à valeurs dans un espace de fonctions concret, on peut donner une définition plus directe d'un estimateur sans biais $(e.\ s.\ b)$.

Soit $(\mathbb{E},\ \mathcal{A},\ \mathcal{P})$ un modèle statistique et g un paramètre à valeurs dans $\mathcal{F}\ (\mathbb{E}',\ \mathbb{C})$.

Définition 2: Un estimateur $x \mapsto T\ (x,\ .\ )$ est dit sans biais de $g(p)$ point par point si pour tout $t \in \mathbb{E}'$ et tout $p \in \mathcal{P},\ T\ (.\ ,\ \ t)$ est p-intégrable et

$$E_p T(X, t) = g(p)(t).(*)$$

Pour $H = C\,[0.1]$ on vérifie aisément qu'un $e.\ s.\ b.$ est sans biais point par point. Il en est de même si H est un espace de Hilbert séparable à noyau reproduisant, la réciproque étant fausse sans hypothèses supplémentaire sur T. Enfin il est clair que(*) signifie simplement que g est un paramètre intégral de noyau $T\ (x,\ t)$. en effet:

$$E_pT\ (X,\ t)\ = g\ (p)\ \ (t)\ \Longleftrightarrow g\ (p)\ \ (t)\ = \int_{\mathbb{E}'} T\ (x,\ t)\, dP\ (x); t \in \mathbb{E}',\ P \in \mathcal{P}.$$

Etant donnés une suite

$$(E_o^n,\ A_o^{\bigotimes n},\ \mu^{\bigotimes n},\ \mu \in \mathcal{P}_o),\ n \geq 1$$

de modèles statistiques et un paramètre $g:\ \mathcal{P}_o \rightarrow H,$ où H est un espace de Banach séparable.

Définition 3: On appelle estimateur d'ordre n toute application T_n de E_o^n dans H. On dira que g est estimable s'il existe un entier $n \geq 1$ tel que g admette un estimateur sans biais d'ordre n.

Le degré de g noté $d^o g$ est le plus petit entier n pour lequel g est estimable (en ce sens), par convention, un paramètre constant est de degré zero et un paramètre non estimable est de degré infini.

Exemple: L'espérance mathématique est estimable d'ordre 1, une observation suffit pour évaluer:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Pour la variance, au moins deux observations, D'une maniére générale les moments d'ordre k :

$$m_k = E\ (X^k)$$

sont estimables d'ordre k

Existence d'estimateur sans biais.

Théorème de Bickel-Lehmann (Existence d'un estimateur sans biais).

Si $\mathcal{P}_o$ est convexe et si g est estimable, les conditions suivantes sont équivalentes:

$$(i) \Longleftrightarrow (ii) = a \text{ et } b$$

où:

$$i) g \text{ est de degré } n.$$

(qui signifie: On peut construire un estimateur sans biais à partir de n observations au minimum

iia) $$\forall \mu, \nu \in \mathcal{P}_o,\ \forall \alpha \in [0,1]\,,\ Q_{\mu,\nu}\ (\alpha)\ = g\ (\alpha\mu +\ (1-\alpha)\ \nu)$$

est un polynôme en α à coefficients dans H et de degré inférieur ou égal à α

iib) $\forall \mu \in \mathcal{P}_o,\ \exists \nu \in \mathcal{P}_o$ telque $Q_{\mu,\nu}\ (\alpha)$ soit exactement un polynôme de degré n.

L'essentiel de la démonstration est contenu dans le lemme suivant:

Lemme: Sous les hypothèses du théorème, on a:

$$\text{i)} \qquad d^o g \leq n \Rightarrow \forall \mu,\ \nu \in \mathcal{P}_o,\ Q_{\mu,\nu}\ (\alpha)$$

est un polynôme de degré $\leq n$.

$$\text{ii) Si: } \exists \mu \in \mathcal{P}_o \text{ telque: } \forall \nu \in \mathcal{P}_o,\ Q_{\mu,\nu}\ (\alpha)$$

soit un polynôme de degré strictement inférieur à n, alors

$$d^o g < n$$

.

Preuve: i) Soit T_n un estimateur sans biais d'ordre n que l'on peut toujours supposer symétrique. Alors:

$$Q_{\mu,\nu}\ (\alpha)\ = \int_{E_o^n} T_n\ (x_1, ...,\ x_n)\ d\ (\alpha\mu +\ (1-\alpha)\ \nu)\ ^n\ (x_1, ...,\ x_n)\ =$$

$$\sum_{k=o}^{n} C_n^k \alpha^k\ (1-\alpha)\ ^{n-k} \int_{E_o^n} T_n\ (x_1, ...,\ x_n)\ d\ (\mu^k \nu^{(n-k)})\ \ (x_1, ...,\ x_n)$$

-à cause de la multinéarité du produit tensoriel ce qui montre i.

ii) Il suffit de montrer que s'il existe un estimateur sans biais d'ordre n, on peut trouver un estimateur sans biais d'ordre $n-1$. Soit donc T_n un estimateur symétrique sans biais d'ordre n. la décomposition de $Q_{\mu,\nu}\ (\alpha)$ donnée précédemment est unique. en écrivant que le coefficient de α^n est nul on obtient:

$$\sum_{k=0}^{n} (-1)^{n-k} C_n^k \int_{E^n o} T_n d(\mu^k \bigotimes \nu^{(n-k)}) = 0$$

et par suite:

$$g(\nu) = \sum_{k=1}^{n} (-1)^{2n-k+1} C_n^k \int_{E_o^n} T_n d(\mu^k \bigotimes \nu^{(n-k)}).$$

posons:

$$T'_k(x_1, ..., x_{n-k}) = \int T_n(x_1, ..., x_n) d\mu^{\bigotimes k}(x_{n-k+1}, ..., x_n)$$

quand cette intégrale est définie, sinon on pose $T'_k = 0$.

par application du théorème de Fubini, on a pour tout $\nu \in \mathcal{P}$

$$g(\nu) = \sum_{k=1}^{n} (-1)^{2n-k+1} C_n^k \int T'_k d\nu^{\bigotimes (n-k)} =$$

$$\sum_{k=1}^{n} (-1)^{2n-k+1} C_n^k \int T'_k(x_1, ..., x_{n-k}) d\nu^{\bigotimes (n-1)}(x_1, ..., x_{n-1})$$

ce qui montre

$$(x_1, ...; x_{n-1}) \mapsto \sum_{k=1}^{n} (-1)^{2n-k+1} C_n^k T'_k(x_1, ..., x_{n-k})$$

est un estimateur

sans biais d'ordre $(n-1)$: ii) est établi. ■

Preuve du théorème:

Si $d^o g = n$, le lemme (i) montre que $Q_{\mu,\nu}(\alpha)$ est un polynôme de degré $\leq n$. Si la seconde condition n'est pas vérifiée, le lemme (ii) montre que $d^o g < n$ ce qui contredit l'hypothèse.

Réciproquement le lemme (ii) montre que $d^o g < n+1$. Montrons que $d^o g > n-1$: Si l'on avait $d^o g \leq n-1$ le lemme (i) impliquerait degré de $Q_{\mu,\nu}(\alpha) \leq n-1$ ce qui est contradictoire. Donc $d^o g = n$. ■

Corollaire:

Si $\mathcal{P}$ est convexe et s'il existe μ et v dans $\mathcal{P}$ tel que $Q_{\mu,v}(\alpha)$ ne soit pas un polynôme en α, g n'est pas estimable.

Résultat:

Les paramètres de régression auqels on est intérressé (la densité conditionnelle, l'esperance conditionnelle, la médiane conditionnelle) ne sont pas estimables (dans le sens: Estimation sans biais). Ce résultat a lieu aussi bien globalement que point par point, conséquence directe du corollaire, a titre d'exemple pour l'esperance conditionnelle.

En effet: supposons $(X_i, Y_i)_{i=1,n}$ de même loi que le couple (X, Y) de densité f, on a:

$$r_f(x) = E_f(Y/X = x) = \int y f(y/x) dy = \frac{\int y f(x,y) dy}{\int f(x,y) dy},$$

ainsi on a:

$$r_{\alpha f+(1-\alpha)g}(x)=\frac{\alpha\int yf(x,y)dy+(1-\alpha)\int yg(x,y)dy}{\alpha\int f(x,y)dy+(1-\alpha)\int g(x,y)dy}$$

on a obtenue **une fraction en** α (pourvue que: $\int f\ (x,\ y)\ dy \neq \int g\ (x,\ y)\ dy$) d'où la non existence d'estimateur sans biais.

Théorème: (Degré d'un paramètre homogène).

Soit $(\mathbb{E},\ \mathcal{B},\ \mathcal{P})$ un modèle statistique où $\mathbb{E}$ est un espace métrique muni de sa tribu borélienne $\mathcal{B}$.

Soit g un paramètre réel, non constant homogène sur un fermé F de P-mesure nulle. $\forall\ P\in\mathcal{P}$ alors si:

$\forall\ U$ouvert contenant F, $P\in\mathcal{P},\ P(U)\ >0\Rightarrow P^U\in\mathcal{P}$(où P^Udésigne la loi conditionnelle)

Alors: g est de degré strictement supérieur à 1.

Applications: Si

$$(\mathbb{E},\ \mathcal{B},\ \mathcal{P})=(\mathbb{R}^{sn},\ B_{\mathbb{R}^{sn}},\ \mu^{\bigotimes n};\mu\in\mathcal{P}_o)$$

où $\mathcal{P}_o$ est constitué des lois de $\mathbb{R}^s$ à densité continue dans un voisinage de $a\in\mathbb{R}^s$, les théorèmes précédents impliquent que pour ce modèle statistique, la densité au point a n'est pas estimable.

Existence d'estimateur minimax

Estimation minimax.

Remarque: Il n'existe pas d'estimateur optimal pour un paramètre g lorsque la classe d'estimateurs envisagée est trop vaste. Cependant on va voir qu'il y a souvent un estimateur minimax.

On reprend le modèle $(\mathbb{E}_o^n,\ \mathcal{B}_o^n,\ \mu^n;\mu\in\mathcal{P}_o)$ où $\mathcal{P}_o$ est dominé par une mesure λ $\sigma-$ finie.

Soit g un paramètre défini sur $\mathcal{P}_o$ et à valeurs dans L^2_ν supposé séparable. où ν est une mesure $\sigma-$ finie sur $(\mathbb{E}_o,\ \mathcal{B}_o)$. pour $h\in\mathcal{M}\ (\mathbb{E}_o^{n+1},\ \mathbb{R})$: Ensemble des application mesurables de $\mathbb{E}_o^{n+1}$dans $\mathbb{R}$, on pose:

$$\|h\|_{\mathcal{P}_o}=\sup_{\mu\in\mathcal{P}_o}\left[\int_{\mathbb{E}_o^{n+1}}h^2d(\mu^n\nu)\right]^{\frac{1}{2}}$$

Si l'on identifie h à sa classe d'équivalence pour la relation:

$$h_1R\ h_2\Longleftrightarrow\left\{h_1=h_2\ \mu^n\bigotimes\nu\ p.p \text{ pour tout } \mu\in\mathcal{P}_o\right\}$$

alors $\Lambda_2=\left\{h:\|h\|_{\mathcal{P}_o}<+\infty\right\}$ muni de $\|.\|_{\mathcal{P}_o}$est un espace de Banach. Par ailleurs, tout estimateur mesurable possède une version dans $\mathcal{M}\ (E_o^{n+1},\ \mathbb{R})$:

on peut donc identifier Λ_2 à une classe d'estimateurs de g. Maintenant pour mesurer l'efficacité de $h \in \Lambda_2$ on pose:

$$\delta(h) = \sup_{\mu \in \mathcal{P}_o} \left[\int_{\mathbb{E}_o^{n+1}} \left[h(x^{(n)}, x) - g(\mu)(x) \right]^2 d\mu^{\bigotimes n}(x^{(n)}) d\nu(x) \right]^{\frac{1}{2}}$$

où $x^{(n)} = (x_1, \ . \ . \ . \ , \ x_n)$ décrit $\mathbb{E}_o^n$. On dit que $h_o \in \Lambda_2$ est minimax si $\delta \ (h_\circ) \ = \min_{h \in \Lambda_2} \delta \ (h)$.

Existence d'un estimateur minimax.

Théorème: Si

$\sup\limits_{\mu \in \mathcal{P}_o} \displaystyle\int [g(\mu)(x)]^2 \, d\nu(x) < +\infty$, alors g admet un estimateur minimax

Exemples: 1) la densité: si

$$g \ (\mu) \ = \frac{d\mu}{d\lambda} \in L^2_\lambda$$

et si

$$\sup_{\mu \in P_o} \int \ (\frac{d\mu}{d\lambda}) \ ^2 d\lambda < +\infty$$

Alors: g admet un estimateur minimax.

2) l'espérance conditionnelle: Si$\mathcal{P}_o$ est dominée par la mesure de Lebesgue sur $\mathbb{R}^s$, et

si le support de la loi de $X_i^{(2)}$ est borné indépendamment de μ, alors la régression admet un estimateur

minimax pourvu que:

$$\sup_{\mu \in \mathcal{P}_o} \int \left[E_\mu (X_i^{(2)} / X_i^{(1)} = x) \right]^2 dx < +\infty$$

ν est ici la mesure de Lebesgue sur $\mathbb{R}^{s_1} \times C$, où C est une boule contenant le support de la loi de $X_i^{(2)}$.

1.3 Estimation Fonctionnelle Asymptotique

La théorie de l'estimation asymptotique est l'etude des suites d'estimateurs lorsque le nombre d'observations tend vers l'infini d'un point de vue purement théorique le modèle asymptotique est caractérisé par la donnée d'un modèle statistique:

$(\mathbb{E}_\infty; \mathcal{A}_\infty; (P_{\theta,\infty})_{\theta \in \Theta})$et d'une suite croissante $(\mathcal{A}_n, n \geq 1)$des sous-tribus de $\mathcal{A}_\infty$

Etant donnée un paramètre g à valeurs dans l'espace mesurable $(H, \mathcal{T})$, suite d'estimateurs mesurables (ou par abus de langage, un estimateur) est une suite $(T_n, \ n \geq 1)$ de $\mathcal{M}\ (\mathbb{E}_\infty, \ H)$ qui est $(\mathcal{A}_n)$ adaptée (i. e. $\forall n \geq 1$, T_n est $\mathcal{A}_n-$ mesurable). Dans le cas où $X = \ (X_n, \ n \geq 1)$ est un échantillon de P_θ probabilité sur $(\mathbb{E}, \ \mathcal{A})$ on a: Le modèle statistique:

$$(\mathbb{E}_\infty; \mathcal{A}_\infty; (P_{\theta,\infty})_{\theta\in\Theta}) = (\mathbb{E}^{\mathbb{N}}; \mathcal{A}^{\mathbb{N}}; (P_\theta^{\mathbb{N}})_{\theta\in\Theta})$$

et $\mathcal{A}_n = \sigma(X_1, \ . \ . \ . \ , \ X_n), \ n \geq 1$

1.3.1 Modes de convergence.

On rappelle maintenant les définitions des principaux modes de convergence stochastique.

Soit $(\Omega, \ \mathcal{A}, \ \mathcal{P})$ un espace probabilisé, H un espace métrique séparable muni de sa distance d et de sa tribu borélienne $\mathcal{B}$, $(\xi_n, \ n \in \mathbb{N})$ une suite de $v. \ a. \ \mathcal{A} - \mathcal{B}$ mesurables

Remarque: La séparabilité de l'espace assure alors que $d\ (\xi_n, \ \xi_o)$ est une $v. \ a. \ r.$

Notations.

On note $\rightarrow_L$la convergence en loi, $\rightarrow_P$la convergence en probabilité $\rightarrow_{P.s.}$la convergence presque sûre $\rightarrow_{P.co.}$la convergence presque complète et $\rightarrow_{N_P}$la convergence en moyenne d'ordre p $(p \geq 1)$.

Ces convergences sont définies de la façon suivante:

Définitions.

$\xi_n \rightarrow_L \xi_o \Longleftrightarrow \forall B \in \mathcal{B} : \ P\ (\xi_o \in \partial B) \ = 0, \ P\ (\xi_n \in B) \ \rightarrow P\ (\xi_o \in B)$ où ∂B désigne la frontière de B.

$\xi_n \rightarrow_P \xi_o \Longleftrightarrow \forall \varepsilon > 0, \ P\left[d(\xi_n, \xi_o) \geq \varepsilon\right] \rightarrow 0.$

$\xi_n \rightarrow_{P.s.} \xi_o \Longleftrightarrow P\left[d(\xi_n, \xi_o) \rightarrow 0\right] = 1.$

$\xi_n \rightarrow_{P.co.} \xi_o \Longleftrightarrow \forall \varepsilon > 0, \ \sum_{n\geq 1} P\left[d(\xi_n, \xi_o) \geq \varepsilon\right] < +\infty.$

$\xi_n \rightarrow_{N_P} \xi_o \Longleftrightarrow E\left[d(\xi_n, \xi_o)^P\right] \rightarrow 0$

Si $H = \mathbb{R}$, la convergence en moyenne d'ordre 2 s'appelle "convergence en moyenne quadratique" et se note $\rightarrow_{m.q.}$. si $H = L^2_\lambda$, elle s'appelle "convergence en moyenne quadratique intégrée".

Un estimateur (T_n) de g, paramètre à valeurs dans H, sera dit convergent suivant le mode stochatique $M\ (T_n \rightarrow_M g)$ si $(\forall \theta \in \Theta)$, $T_n \rightarrow_M g\ (\theta)$ pour $(\Omega, \ \mathcal{A}, \ \mathcal{P}) \ = \ (\mathbb{E}_\infty, \ \mathcal{A}_\infty, \ P_{\theta,\infty})$; on aura par exemple

$T_n \rightarrow_P g \Longleftrightarrow \ (\forall \theta \in \Omega), \ \ (\forall \varepsilon > 0), \ P_{\theta,\infty}\left(d\left(T_n, \ g\ (\theta)\ \right) \ \geq \varepsilon\right) \ \rightarrow_{n\rightarrow\infty} 0.$

Tableau des convergences:

$\boxed{p.co. \Rightarrow p.s. \Rightarrow p \Rightarrow L}$

$\boxed{N_p \Rightarrow p}$

1.3.2 Estimateurs convergents

Existence d'un estimateur convergent.

Etant donné le modèle $(\mathbb{E}^{\mathbb{N}}; \mathcal{A}^{\mathbb{N}}; P^{\mathbb{N}},\ p \in \mathcal{P}_o)$ et le paramètre $g:\ \mathcal{P}_o \to H$. On va enoncer une condition nécessaire et suffisante pour qu'il existe un estimateur de g convergent presque sûrement. On munit l'ensemble $\mathcal{P}\ (\mathbb{E},\ \mathcal{A})$ des probabilités sur $(\mathbb{E},\ \mathcal{A})$ de la structure uniforme U définie par la famille d'écarts:

$$e_{n,f}(P_o,\ Q_o) = \left| \int f\ dP_o^n - \int g dQ_o^n \right|;\ n \in \mathbb{N}^*$$

$f \in \mathcal{M}_b\ (\mathbb{E}^n,\ \mathbb{R})$ -Ensemble des applications mesurables et bornées de $\mathbb{E}^n$dans $\mathbb{R}$-où $P_o,\ Q_o \in \mathcal{P}\ (\mathbb{E},\ \mathcal{A})$. D'autre part, on suppose que $(H,\ d)$ est un espace métrique séparable: il est donc homéomorphe à un sous-espace cube $[0,1]^{\mathbb{N}}$ muni de la topologie associée à sa distance usuelle définie par:

$$\delta(x,\ y) = \sum_{i=0}^{+\infty} 2^{-i} |x_i - y_i|\,; x = (x_i,\ i \in \mathbb{N}),\ y = (y_i,\ i \in \mathbb{N});\ x,\ y \in [0,1]^{\mathbb{N}}.$$

Soit une hypothèse plus forte en supposant que d a été choisie de telle sorte que H soit isométrique à un tel sous-espace: Cela permettra d'identifier H à une partie du cube.

Condition nécessaire et suffisante d'existence d'un estimateur convergent

Théorème:

Les conditions suivantes sont équivalentes:

1) Il existe un estimateur de g qui converge en probabilité.

2) II existe un estimateur de g qui converge presque sûrement.

3) Il existe une suite (g_k) de paramètre à valeurs dans $[0,1]^{\mathbb{N}}$ telle que:

i) pour chaque k, g_k est $U - \delta$ uniformément continue.

ii) (g_k) tend vers g.

Corollaire: . Si H est un espace métrique séparable et g limite d'une suite:

$$(g_k,\ k \geq 1)$$

de paramètres, vérifiant l'une des conditions suivantes:

a) $\forall k \geq 1,\ g_k$ est $U{-}\delta$ uniformément continu

b) $\forall k \geq 1$, g_k admet un estimateur convergent en probabilité, uniformément sur $\mathcal{P}_o$

Alors:

g admet un estimateur qui converge presque sûrement

.

Application aux paramètres limites d'une suite de paramètres intégraux.

Si: $\mathbb{E} = [0,1]$, $H = C[0,1]$ muni de sa distance usuelle et

$$g_k\ (P)\ \ (\cdot)\ = \int_{[0,1]} \varphi_k\ (x, \cdot\ )\ dP\ (x); P \in \mathcal{P}_\circ; k \geq 1$$

où $\varphi_k \in C\ ([0,1] \times [0,1])$, alors:

$$g_k \text{ est } U - \delta \text{ uniformément continu}$$

Si g_k tend vers g dans $C[0,1]$ le corollaire permet d'en déduire l'existence d'un estimateur (T_n) qui converge presque sûrement vers g, autrement dit

$$(\forall P \in \mathcal{P}_o),\ \sup_{t\in[0,1]} |T_n(t) - g(P)(t)| \rightarrow o\ P^{\infty} p.\ s$$

Etude de la vitesse de convergence

On se contente d'énoncer un théorème de C. J. Stone (1980) relatif aux deux paramètres fonctionnels les plus importants: La densité et l'espérance conditionnelle. . On considère le modèle asymptotique:

$$(\mathbb{R}^{s\mathbb{N}},\ \mathcal{B}^{\mathbb{N}}, (P_\theta^{\mathbb{N}})\ _{\theta\in\Theta})$$

$\mathcal{B}$ tribu borélienne de $\mathbb{R}^s$ et Θ désigne une famille de fonctions réelles définies sur un voisinage ouvert U de l'origine dans $\mathbb{R}^{s_1}$. on se propose d'estimer le paramètre $g\ (\theta)\ = \theta\ (0)$, $\theta \in \Theta$ (à l'origine).

Définition:

a) Un réel positif r sera appelé borne supérieure de la vitesse de convergence vers $g\ (\theta)$) si pour tout estimateur (T_n) de $g\ (\theta)$ on a:

$$(\forall c > 0), \lim_{n\to\infty} \sup_{\theta\in\Theta} P_\theta^{\mathbb{N}}\, |T_n - g(\theta)| > cn^{-r}) > 0$$

et

$$\lim_{c\to 0} \frac{\lim}{n \to \infty} \sup_{\theta\in\Theta} P_\theta^{\mathbb{N}}(|T_n - g(\theta)| > cn^{-r}) = 1.$$

b) on dit que la vitesse de convergence r est atteinte s'il existe un estimateur-une suite d'estimateur (T_n) tel que:

$$\lim_{c\to\infty}\lim_{n\to\infty}\sup_{\theta\in\Theta} P_\theta^{\mathbb{N}}(|T_n - g.\theta)| > c(n^{-r}) = 0$$

c) on dit que r **est la vitesse de convergence optimale s'il vérifie a) et b).**

Remarque: On va maintenant préciser deux modèles pour lesquels la vitesse de convergence est bien déterminée, on introduit d'abord quelques notations:

Notations:

Soit $h \in C_k\ (U)$ et soit $h_k$le polynôme de Maclaurin de degré k associé à h. on désignera par $G_k\ (U)$ l'ensemble des $h \in C_k\ (U)$ tels que:

$$\|h(x) - h_k(x)\| \leq m\,\|x\|^{k+1}$$

ou m est une constante fixée

Exemple1:

Régression: On considère le modèle décrit dans l'exemple 5 du1. 1. 2, le paramètre θ est la fonction de régression .

$$x \longmapsto E(X_i^{(2)}/X_i^{(1)} = x) \text{ et } \Theta = \{\theta_o + h, h \in G_k(U)\}$$

ou $\theta_o$est fixée dans C_o.

On suppose que la variance conditionnelle vérifie:

$$(\forall x \in U),\ 0 < \alpha \leq V\ (X_i^{(2)}/X_i^{(1)} = x) \leq \beta < \infty$$

où α et β sont des réels ne dépendant que de θ, et que la loi de $X_i^{(2)}$ sachant que $X_i^{(1)} = x$ est de la forme:

$$\int f\ (y/x,\ \theta\ (x))d\lambda\ (y)$$

où $(x,\ t,\ y,\ ) \longmapsto f\ (y/x,\ t)$ est strictement positive, mesurable et telle que $\int yf\ (y/x,\ t)\ d\lambda\ (y) = 1$, pour t variant dans un ouvart $U_1 \supset \theta^{-1}\ (U)$ et x variant dans U. On suppose de plus que: $(x,\ t,\ y) \longmapsto f\ (y/x,\ t)$ est dans $C_2\ (U)$ et que l'égalité $\int f\ (y/x,\ t)\ d\lambda\ (y) = 1$ est deux fois différentiable sous le signe somme par rapport à t. enfin, on suppose qu'il existe ε_o et C positifs et $(x,\ t,\ y) \longmapsto M\ (y/x,\ t)$ mesurable tels que:

$$\frac{\partial^2 \log\ f(y/x, t+\varepsilon)}{\partial t^2} \leq M\ (y/x,\ t),\ \varepsilon < \varepsilon_o \text{ et } \int M\ (y/x,\ t) f\ (y/x,\ t) d\lambda\ (y) \leq C.$$

Exemple2: Densité. On considère le modèle décrit dans l'exemple 4 de 1. 1. 2, θ est la densité et $\Theta = \left\{\theta_o(1+h), h \in G_k,\ \ |h| \leq 1,\ \ \int \theta_o\ h\ d\lambda = 0\right\}$ ou $\theta_o \in C_o\ (U)$ vérifie $\theta_o\ (0)\ > 0$. ici $s_1 = s$

Théorème C. J. Stone (1980). Vitesse de convergence optimale en probabilité .

Dans les modèles de densité et de régression la vitesse de convergence optimale est:

$$\mathbf{k+1/\ (2\ (k+1)\ +s_1}$$

On retrouve le fameux: $r = \frac{2}{5}\ (k = 1, s_1 = 1)$

Preuve: Voir: C. J. Stone (1980).

1.4 Construction d'estimateurs Fonctionnelles

1.4.1 Etude de la convergence de la mesure empirique vers la mesure théorique

Soit un échantillon $X_{1,\ldots,}X_n$ de la loi μ on peut lui attribuer la mesure aléatoire:

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$$

appelée mesure empirique. C'est l'estimateur de μ le plus naturel puisque il s'agit de la distribution uniforme sur l'ensemble des observations.

Remarque: les statistiques utilisée en estimation non parametrique s'en déduisent par régularisation. Avant d'étudier les principales proprietés asymptotiques de μ_n. On présente une variante de l'inégalité de Bernstein

Inégalité de Bernstein-Frechet. Pour les variables bornées. Soit la version suivante:

Théorème: Soient $\xi_1,\ .\ .\ ,\ \ \xi_n$ des v. a. r indépendantes définies sur $(\Omega,\ A,\ P)$ un espace probabilisé.

(i) si $\alpha_i \leq \xi_i \leq \beta_i$, $i = 1,\ .\ .\ .\ .\ ,\ \ n$ où les α_i, β_i sont des constantes reélles, on a:

$$P[|\sum_{i=1}^{n}(\xi_i - E\xi_i)\ |\geq t] \leq 2\exp\left[-\frac{2t^2}{\sum\limits_{i=1}^{n}(\beta_i - \alpha_i)}\right] \quad (1)$$

(ii) S'il existe une constante $c > 0$ telle que:
$E \mid \xi_i - E\xi i \mid^P \leq c^{p-2}p!Var\xi_i.\ \ i = 1, \ . \ . \ . \ ,\ \ n; p = 3,\ 4.\ .\ .$ on a:

$$P[\mid \sum_{i=1}^{n}(\xi_i - E\xi_i) \mid \geq t] \leq 2\exp\left[-\frac{t^2}{4\sum\limits_{i=1}^{n} Var\xi_i + 2ct}\right] \quad (2)$$

Corollaire: Soit C une population de N nombres $c_1,\ c_2,\ .\ .\ .\ ,\ \ c_n$ et $\xi_1',\ .\ .\ .\ .\ ,\ \ \xi_n'$ un échantillon **sans** remise issu de C. Alors, l'inégalité de Bernstein-Fréchet s'applique aux variables bornées ξ_i'.

Théorème: Approximation uniforme de la mesure théorique par la mesure empirique.

$\forall\ \varepsilon > 0,$ on a:

$$P\left[\sup_{B\in C}|\mu_n(B) - \mu(B)| > \varepsilon\right] \leq 4e^{(4\varepsilon+4\varepsilon^2)}M(C,n^2)e^{-2n\varepsilon^2} \quad (3)$$

où $M(C,\ n)$ désigne la compexité d'une classe C, il est égale à:

$$M(C,n) = \max_{(x_1,\ldots,x_n)\in\mathbb{R}^{sn}} N_C(x_1,...,x_n)$$

où $N_C(x_1,\ \ldots,\ x_n)$ est le nombre d'ensembles distincts dans$\{(x_1,...,x_n)\cap B, B\in C\}$

Exemple: 1) Lorsque C est réduit à un singleton: $C = \{B\}$: $M(C,\ n) =$ 1

2) Lorsque $C = \left\{ {}_{j=1}^{s}\,]-\infty, a_j]\, ; a, ..., a \in \bar{\mathbb{R}} \right\}$: $M(C,\ n) =$ $(1+n)^s$

Corollaire: Convergence uniforme presque compléte de μ_n vers μ

Si $(\forall\varepsilon > 0)$, $\sum_{n\geq 1} M\ (C,\ n^2)\ e^{-2n\varepsilon^2} < +\infty$, alors:

$$\sup_{B\in C}|\mu_n(B) - \mu(B)| \rightarrow_{p.co.} 0 \quad (4)$$

Application: Estimation de la fonction de répartition

Pour le borélien B de $\mathbb{R}^s$,

$$B =]-\infty, x] = \prod_{j=1}^{s}\,]-\infty, x_j]$$

on retrouve la fonction de répartition Fde μ noté F_μ, L'estimateur le plus naturel de F_μ est la fonction de répartition empirique:

$$F_n\ (x) = \mu_n\ (]-\infty, x]) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,x]}\ (X_i); x\in\mathbb{R}^s$$

où $x = \ (x_1, \ . \ . \ . \ . \ x_s)$ et $]-\infty, x] = \prod_{j=1}^{s}]-\infty, x_j]$.

Le théorème suivant résume les propriétés d'optimalité et de convergence de F_n. On suppose que μ appartient à P_o. , sous-ensemble donnée de $P\ (\mathbb{R}^s)$.

Estimation de la $f.\ d.\ r.$ **par la** $f.\ d.\ r.$ **empirique.**

a) Si (S. O) la statistique d'ordre est complète alors, pour chaque x de $\mathbb{R}^s$, $F_n\ (x)$ est un estimateur sans biais optimal de $F_\mu\ (x)$. de plus, si m désigne une mesure bornée sur $(\mathbb{R}^s,\ \mathcal{B})$, F_n est un estimateur sans biais optimal de F_μ dans l'espace de Hilbert $L^2\ (\mathbb{R}^s,\ \mathcal{B},\ m)$.

b) On a l'inégalité suivante:

$$\forall \varepsilon > 0, \sup_{\mu \in \mathcal{P}_0} P\left[\sup_{x \in \mathbb{R}^s} |F_n(x) - F_\mu(x)| > \varepsilon\right] \leq 4e^{(4\varepsilon + 4\varepsilon^2)}(1 + n^2)^s e^{-2n\varepsilon^2} \quad (5)$$

et par conséquent:

Théorème de Glivenko-Cantelli:

$$(\forall \mu \in \mathcal{P}_o), \sup_{x \in \mathbb{R}^s} |F_n(x) - F_\mu(x)| \rightarrow_{p.co.} 0 \quad (6)$$

De plus

$$(\forall \mu \in \mathcal{P}_o), \sup_{x \in \mathbb{R}^s} |F_n(x) - F_\mu(x)| = 0(\frac{\log n}{n})^{\frac{1}{2}} p.s \quad (7)$$

En utilisant des méthodes directes, l'intégalité (5)devient pour $s = 1$:

$$P\left[\sup_{x \in \mathbb{R}} |F_n(x) - F_\mu(x)| > \varepsilon\right] \leq Ce^{-2n\varepsilon^2} \quad (8)$$

où C est une constante universelle, comme C n'est pas connue explicitement, cela diminue la portée pratique de (8). Quand $s \geq 1$, on obtient l'inégalité explicite: $\forall n \geq s^2 \varepsilon^{-2}$

$$P\left[\sup_{x \in \mathbb{R}^s} |F_n(x) - F_\mu(x)| > \varepsilon\right] \leq 2e^2(2n)^s e^{-2n\varepsilon^2} \quad (9)$$

Remarque:

La fonction de répartition empirique a des défauts habituels des estimateurs sans biais qui ne sont souvent **ni stricts** (i. e. ne prennent pas leurs valeurs dans l'ensemble Θ des valeurs possibles du paramètre) ni admissibles (i. e il existe des estimateurs qui leurs sont strictement préférables). Pour un modèle dominé, les estimateurs stricts de la densité induisent par intégration des estimateurs stricts de la fonction de répartition. les estimateurs admissibles usuels sont **bayésiens**. En statistique non paramètrique leur construction est difficile car la loi a priori doit répondre à deux exigences contradictoires: Avoir un support assez vaste et cependant permettre un calcul simple des lois a posteriori. Le processus de Dirichlet est un compromis acceptable.

Lois limites: Le théorème central limite multidimensionnel permet d'établir directement le résultat suivant:

Théorème: Pour toute probabilité μ sur $\mathbb{R}^s$, pour $k \geq 1$ et $(t_1, \ldots, t_k) \in \mathbb{R}^{sk}$, on a:

$$\sqrt{n}\,[F_n(t_1) - F_\mu(t_1), ..., F_n(t_k) - F_\mu(t_k)] \rightarrow_L \mathcal{N}$$

pour n tendant vers l'infini où $\mathcal{N}$ suit une loi de Gauss centrée et de même matrice de convariance que

$$(1_{\{X_1 \leq t_1\}}, \ldots, 1_{\{X_k \leq t_k\}})$$

On peut préciser cette convergence en montrant que le processus empirique:

$$(\sqrt{n}\ (F_n\ (t)\ - F_\mu\ (t),\ t \in \mathbb{R}^s)$$

converge vers un processus gaussien réel. Lorsque μ est la loi uniforme sur $[0,1]^s$, ce processus est un pont brownien:

$$(B(t), t \in [0,1]^s)$$

C. à. d: un processus gaussien séparable tel que:

$$EB\ (t)\ = 0$$

et

$$cov\ (B\ (t),\ B\ (u)\ )\ = \prod_{J=1}^{s} \min\ (t_J,\ u_J)\ - \prod_{J=1}^{s} t_J u_J$$

où $t = (t_1, \ldots, t_s) \in [0,1]^s$; $u = (u_1, \ldots, u_s) \in [0,1]^s$.

Remarque: Un pont brownien peut s'exprimer en fonction d'un processus de Wiener multidimensionnel $(W\ (t),\ t \in [0,1]^s)$ par la relation:

$$B(t) = W(t) - t_1....t_s W(\mathbf{1}); t \in [0,1]^s; \mathbf{1} = (1, ..., 1) \in \mathbb{R}^s$$

avec:

$$EW\ (t)\ = 0$$

$$Cov\ (W\ (t),\ W\ (u)\ )\ = \prod_{J=1}^{s} (t_J,\ u_J)$$

Notons Δ_n le processus empirique associé à la loi uniforme sur $[0,1]^s$, pour $s = 1$ on a:

$$\Delta_n = \sqrt{n}\ (F_n\ (x)\ - x)$$

où $F_n\ (x)\ = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}})$. On peut énoncer:

Théorème: Approximation forte du processus empirique par un pont brownien

Il existe un espace probabilisé $(\Omega,\ A,\ P)$, des versions Δ_n^* de Δ_n et B_n^* de B telles que:

$$\sup_{t\in[0,1]^s} |\Delta_n^*(t) - B_n^*(t)| = 0(\frac{(\log n)^s}{\sqrt{n}})\ p.s.$$

Pour $s = 1$, ce résultat a été obtenu par (K. M. T) abréviation de: Komlos-Major et Tusnady (1975) et par Borisov (1980) pour $s > 1$

1.5 Remarques et Conclusion.

♠ Les méthodes naturelles d'estimation sont basées sur la mesure empirique: Elles consistent à estimer le paramètre inconnu $\mathbf{g}\ (\mu)$ par le paramètre empirique correspondant $\mathbf{g}(\mu_n)$.

Les estimateurs obtenus ont alors des propriétés intéresantes. on a vu celles de F_n. Cependant dans bien des cas le paramètre empirique n'est pas défini (densité, régression, densité spectrale par exemple).

On est alors amenés à construire des estimateurs fonctionnels en régularisant la mesure empirique ou en modifiant le paramètre à estimer.

Si $X_1,\ .\ .\ ,\ \ X_n$ est un échantillon d'une loi à densité définie sur $\mathbb{R}^s$ on peut construire la fonction de répartition empirique régularisée F_n au moyen d'une fonction spline qui interpole F_n et admettra une densité.

Une autre méthode consiste à faire le produit de convolution

$$K_n * \mu_n$$

où K_n est une densité donnée; on obtient ainsi une densité empirique qui dépend de K_n où:

$$K_n\ (\cdot)\ = \frac{1}{h(n)} K\ (\frac{\cdot}{h(n)})$$

et:

$$\lim_{n\to\infty} h\ (n) = 0$$

avec$\frac{1}{n}$ et K un noyau. C'est cette méthode qu'on a appliqué pour l'estimation des paramètres étudiés, si le noyau K est égale à:

$$K\ (x)\ = 1_{\left[-\frac{1}{2},\frac{1}{2}\right]^s}\ (x)$$

on retrouve la méthode de la fenêtre mobile.

♠ La modification du paramètre à estimer s'effectue souvent par **projection** dans l'espace du parmètre.

Plus généralement, si g est un paramètre fonctionnel limite d'une suite (g_k) de paramètre intégraux, on peut remplacer $g\ (\mu)$ par $g_{k_n}(\mu)$ où la suite k_n sera choisie de façon à permettre de définir le paramètre empirique $g_{k_n}\ (\mu)$. Dans le cas d'un paramètre possédant le propriétés de localisation, il peut être intéressant de construire un estimateur spécifique pour en tenir compte.

Ainsi pour un paramètre localisé par un point, on peut construire un estimateur basé sur les observations situées au voisinage de ce point. On peut également se donner un entier k inférieur à n et n'utiliser que les k observations les plus proches de ce point (Méthode qui fait intervenir la statistique d'ordre). L'histogramme des fréquences et le régressogramme font également partie de ces méthodes locales.

♠ Les méthodes d'estimation par projection sont trés intéressantes lorsque les observations sont dans un éspace de dimension infinie.

♠ Ce qui précède peut faire penser qu'il existe une classification rigoureuse des méthodes d'estimation fonctionnelle en réalité: les familles d'estimateurs évoquées ci-dessus se recouvrent en partie. .

Chapter 2

Estimation Non-Paramétrique de l'espérance conditionnelle

2.1 Etude Bibliographique et Résultats

Le terme régression trouve son origine dans les travaux de Pearson et Lee (1903) qui introduisent la notion de régression à propos d'observations du couple $(X,\ Y)\ =$ (taille du père, taille du fils) afin d'étudier la corrélation entre la taille du pére et la taille du fils. La droite de régression est déterminée à l'aide d'une présentation graphique (Pearson et Lee1903) associée aux observations $(X_i,\ Y_i)\ \ i = 1,\ n$, et qui n'est autre chose qu'un régressogramme. Ce travail a été fondamental dans l'évolution ultérieure de la statistique non paramétrique. Le regressogramme est introduit pour simplifier le calcul (peu automatisé à cette époque) des coefficients de la droite de régression, mais l'étude mathématique du régressogramme trés tardive par rapport à Pearson et Lee (1903), repose sur le développement de la statistique non paramètrique, développement qui est lié à celui des moyens automatiques de calcul.

Les données:

Soient X un vecteur aléatoire de $\mathbb{R}^s (s \geq 1),\ Y$ une v. a. r et $(X_i,\ Y_i)_{i=1,\dots,n}$ un n échantillon de même loi que le couple $(X,\ Y)$, on désigne par r l'esperance conditionnelle de Y en X

$$\mathbf{r}(\mathbf{x}) = \mathbf{E}(\mathbf{Y}/\mathbf{X} = \mathbf{x})$$

et on considère l'estimation de cette fonction, lorsqu'on ne fait à priori sur la loi inconnue de $(X,\ Y)$aucune hypothèse autre que celle de l'existence de r, c. à. d: $E(|Y|) \quad < \infty$.

2.1.1 Présentation des méthodes

Les divers estimateurs de "régression" qu'on considére sont notés r_n.

Concernant leurs construction, la plupart des méthodes ont en commun l'approche suivant:

Pour tout $x \in \mathbb{R}^s$, $r_n(x)$ est la moyenne des Y_i**,** $i = 1,\ ...,\ n$ tel que X_i **soit proche de** x.

Ainsi les divers méthodes se différencient par le fait que la notion de **proximité** utilise ou non la notion de la **statistique d'ordre(SO)** .

Parmi les méthodes n'utilisant pas la SO, il s'agit p. ex de la méthode de la fenêtre mobile(cas particulier de la méthode du noyau)qui consiste à prendre comme voisinage de x une boule de centre x et de rayon h_n ainsi: $r_n(x)$ est la moyenne des $(Y_i)\ _{i=1,\dots,n}$ tel que X_i appartienne à la boule de centre x et de rayon h_n.

Parmi les méthodes utilisant la SO, à titre d'exemple la méthode des k points les plus proches(k-nearst neighbour notée $k - NN$) . La conception de cet estimateur est comme suit: $r_n(x)$ est la moyenne des$(Y_i)\ _{i=1,\dots,n}$ tel que X_i appartienne à la plus petite boule de centre x contenant au mons k_n points $X_j,\ j = 1,\ ...,\ n$.

Les résultats qu'on cherche sont de type **asymptotique** liant la convergence de r_n vers r, selon divers modes aux propriétés asymptotiques de suites réelles intervenant dans la définition de r_n, il s'agit de h_n (la fenêtre) ou de k_n

$$k_n = \frac{1}{h_n^s}$$

Ces résultats font également intervenir des hypothèses sur la loi du couple$(X,\ Y)$, et la loi marginale de X

Hypothèses générales

Les méthodes considérées ici sont qualifiées de non paramètriques car elles ne se ramènent pas à l'estimation d'un **nombre fini de paramètres** réels associés à la loi $P_{X,Y}$ du couple $(X,\ Y)$. Cette loi est cependant assujettie à des hypothèses trés générales, qui se réduisent à la seule condition:

$$E\,|Y|^q < \infty,\ q \geq 1$$

lorsque r_n vérifie la propriété de convergence universelle (Stone 1977) :

$$\forall q \geq 1, E\,|Y|^q < \infty \Longrightarrow E\,|r_n(X) - r(X)|^q \to 0$$

X indépendant de $\{(X_i, Y_i), i = 1, n\}$. Lorsqu'on s'intéresse à la convergence de $r_n(x)$ vers $r(x)$, x fixé

(resp; la convergence uniforme de $r_n$vers r, sur une partie non vide G de $\mathbb{R}^s$),

les hyphotèses sont toujours du type suivant :

i) r est continue en x (resp. sur un voisinage de G) .

ii) La loi marginale P_X admet par rapport à la mesure de Lebesgue sur $\mathbb{R}^s$ une densité f minorée par un nombre positif sur un voisinage de x. (resp. G)

iii) $EY^2 < \infty$ et la variance conditionnelle υ définie par:

$$\upsilon(x) = E((Y - r(x)\ )^2 / X = x)$$

$\forall x \in \mathbb{R}^s$est bornée sur un voisinage de x (resp. G)

Remarque : Ces hypothèses sont toujours de simples hypothèses de régularité, analytique sur r et probabiliste sur la loi marginale P_X et la loi conditionnelle $P_Y^{X=x}$.

D'autres approches font intervenir des hypothèses plus restrictives (par exemple Y bornée) ou, au contraire rendent moins restrictives ces même hypothèses (par exemple Geffroy (1975) pour l'existence de f, Collomb (1979) pour l'existence de EY)

Remarque: On donne dès maintenant un résultat: Conséquence directe (Bosq 1969) d'un théorème dû à Bickel et Lehmann (1969). Sous les hypothèses (i), (ii), (iii)

Il n'existe pas d'estimateurs sans biais de r, en ce sens que si $\mathcal{D}$ désigne l'ensemble des lois P vérifiant (i) (ii) et (iii), il n'existe pas d'estimateur r_n vérifiant:

$$\forall P \in \mathcal{D}, \forall x \in \mathbb{R}^s,\ Er_n(x) = r(x)$$

Cependant on note que les estimateurs r_n qu'on considére sont **asymptotiquement sans biais** .

2.1.2 Méthodes d'estimations: déscription et résultats

A) **Parmi les méthode n'utilisant pas la statistique d'ordre:**

1) **Méthode du régressogramme(à pas fixe) :**

Cet estimateur proposé par Tukey (1961) a une origine intuitive immédiate, (X à valeurs dans $[0, 1[$) on divise l'intervalle $[0, 1[$ en k_n intervalles égaux et disjoints et on estime la régression par la fonction en escalier qui sur chaque intervalle est constante et égale à la moyenne des Y_i, $i = 1,\ n$, tels que X_i

appartienne à cet intervalle et nulle si aucun X_i n'appartient à l'intervalle. Cette définition se généralise au cas où X prend ses valeurs dans $\mathbb{R}^s$, $s > 1$ en posant:

$$r_n(x) = \sum_{i=1}^n Y_i K_n(x, X_i) / \sum_{i=1}^n K_n(x, X_i) \text{ si } \sum_{i=1}^n K_n(x,\ X_i) \neq 0$$

$$= 0 \qquad \text{sinon}$$

où $K_n(x, z) = 1_{A_x}(z)$avec :

$$A_x = \left\{ \prod_{i=1}^s \left[h_n \left[\frac{x^{(l)}}{h_n} \right], h_n(\left[\frac{x^{(l)}}{h_n} \right] + 1) \right] \right\},\ h_n^s = \frac{1}{k_n}$$

(Notation: 1_A est la fonction indicatrice sur A, $\left[\frac{x^{(l)}}{h_n}\right]$ est la partie entière de$\frac{x^{(l)}}{h_n}$, $x^{(l)}$ la lièmme coordonnée de x),

Remarque: L'analogie avec l'histogramme (estimateur de la densité) est évidente .

L'étude des propriétés de convergence du régressogramme: A d'abord eu pour objet :

$$M_n = \sup_x |r_n(x) - r(x)|$$

Bosq (1969) donne pour r_n une condition suffisante de convergence presque sûr de M_n vers 0. Ce résultat a été amélioré par Geffroy (1975) qui montre que pour X à valeurs dans $[0, 1]^s$

$$k_n/n \to_{n\to\infty} 0 \text{ et } k_n/\log n \to_{n\to\infty} \infty \text{ où } k_n = \frac{1}{h_n^s}$$

est une condition nécessaire et suffisante de convergence uniforme presque sûr. Sabry (1978) obtient une condition suffisante de convergence presque sûre uniforme sur unintervalle (avec $s = 1$) dont la longueur dépend asymptotiquement de n . Collomb (1976) (1979) étudie les propriétés de convergence ponctuelle de r_n, en donnant pour tout x fixé une évaluation asymptotique du biais $b_n(x)$ et de la variance $V_n(x)$ de $r_n(x)$, en fonction de $r'(x)$, $f(x)$, $v(x)$ (variance conditionnelle) et h_n lorsque $h_n \to_{n\to\infty} 0$, $nh_n^s \to_{n\to\infty} \infty$.

2) **Méthode du noyau:**

Soient $(h_n)_n$ une suite réelle positive vérifiant :

$$\lim_{n\to\infty} h_n = 0$$

K un noyau de $\mathbb{R}^s$: Une fonction de densité sur $\mathbb{R}^s$bornée et vérifiant

$$\|x\|^s K(x) \quad \rightarrow_{|x|\rightarrow\infty} 0$$

(condition de Parzen), on suppose de plus qu'il est symétrique : $K(x) = K(-x)$, $\forall x \in \mathbb{R}^s$. Watson (1964) et Nadaraya (1964) on proposé, indépendamment et simultanément l'estimateur:

$$r_n(x) = \frac{\sum_{i=1}^{n} Y_i K((x - X_i)/h_n)}{\sum_{i=1}^{n} K((x - X_i)/h_n)} \text{si} \sum_{i=1}^{n} K((x - X_i)/h_n) \neq 0$$

$r_n(x) \quad = 0$ sinon. Le premier en utilisant les échantillons empiriques de loi connue, le second donnant quelques résultats de convergence pour le cas $s = 1$(les démonstrations sont analogues à celles de Parzen (1962) sur l'estimateur de densité) :

$$f_n(x) \quad = \frac{1}{nh_n^s} \sum_{i=1}^{n} K((x - X_i) \quad /h_n)$$

déja étudié par Rosenblatt (1956) .

Propriétés de convergence ponctuelle : On s'intéresse aux résultats concernant les propriétés de r_n en point x fixé, ou éventuellement, en un nombre fini de points . Rosenblatt (1969) et Konakov (1973) étudient $r_n(x)$ dans le cas $s = 1$, en l'écrivant sous la forme:

$$r_n(x) = \int y f_n(x, y) dy / \int f_n(x, y) dy$$

où

$$f_n(x, y) \quad = \frac{1}{nh_n^2} \sum_{i=1}^{n} K((x - X_i)/h_n, \ (y - Y_i)/h_n)$$

K noyau de $\mathbb{R}^2$tel que: $\int K(z, \ y) \quad dy = \mathbf{K}(z)$ est un noyau sur $\mathbb{R}$. Et en utilisant les propriétés de f_n en tant qu'estimateur de la densité f de la loi du couple $(X, \ Y)$, densité dont on suppose alors l'existance et la continuité . Ils obtiennent ainsi divers résultats de convergence ponctuelle, en imposant à $(h_n) \quad _{\mathbb{N}}$ la condition:

$$nh_n^{s+1} \rightarrow_{n\rightarrow\infty} \infty$$

La première étude directe de $r_n(x)$, x fixé, est celle de Schuster (1972) qui obtient la propriété de convergence asymptotiquement normale, en corrigeant le résultat du même type donné par Nadaraya (1964) . Il montre également que $r_n(x_1)$ et $r_n(x_2)$, $x_1 \neq x_2$, sont asymptotiquement indépendantes. Noda (1976) étudie les propriétés de convergence ponctuelle et donne une majoration asymptotique de l'erreur quadratique

$$E(r_n(x) \quad - r(x)\ )^2$$

Collomb (1976 ou 1979) donne une évaluation asymptotique du biais et de la variance de l'estimateur $r_n(x)$, x fixé, en fonction de h_n, $r'(x)$, $r''(x)$, $f(x)$, $f'(x)$ et $v(x)$ (variance conditionnelle)

$$b_n(x) = Er_n(x) - r(x) \simeq_{n\to\infty} \frac{1}{2} h_n^2 \sum_{l=1}^{s} \sum_{m=1}^{s} b_{l,m}(x) \int z^{(l)} z^{(m)} \mathbf{K}(z) dz$$

avec $b_{l,m}(m) \quad = \frac{\partial^2 r(x)}{\partial x_l \partial x_m} + \frac{\partial r(x)}{\partial x_l} \frac{\partial \log f(x)}{\partial x_m} + \frac{\partial r(x)}{\partial x_m} \frac{\partial \log f(x)}{\partial x_l}$ et

$$V_n(x) = E(r_n(x) - Er_n(x))^2{}_{n\to\infty} \simeq \frac{1}{nh_n^s} \frac{v(x)}{f(x)} \int \mathbf{K}^2(z) dz.$$

Ces formules permettent notamment de montrer que:

$$\min_{h_n \in \mathbb{R}^+} E(r_n(x) - r(x))^2_{n\to\infty} \simeq cn^{-4/(s+4)}$$

où c est une constante dépendant uniquement du noyau $\mathbf{K}$ et de la loi du couple $(X,\ Y)$.

Collomb (1979) généralise le résultat de Schuster (1972) et obtient en particulier:

$$P((T_n(x))^{-1/2}(r_n(x) - r(x)) \leq t) \to_{n\to\infty} (2\pi)^{-1/2} \int_{-\infty}^{t} e^{-u^2/2} du$$

où

$$T_n(x) \quad = \sum_{i=1}^{n} (Y_i - r_n(x)\ )^2 \mathbf{K}((x - X_i) \quad /h_n) \quad /(\sum_{i=1}^{n} \mathbf{K}((x - X_i) \quad /h_n)\ )^2$$

et

$$\lim_{n\to\infty} nh_n^{s+4} = 0.$$

(r resp. f dérivable à l'ordre 2 resp1). Ce résultat permet de définir pour $r(x)$ un intervalle de confiance de sécurité asymptotique donné .

Remarques:

1) Tous les résultats asymptotiques qu'on vient d'enoncer ne permettent pas de répondre à l'importante **question** que posent les praticiens de la statistique: Pour n fixé, **comment choisir la fenêtre** h_n?

Réponse:

Duhamel (1978) propose de remplacer dans la définition h_n par une *v. a. r.* minimisant une vraisemblance empirique ;Collomb (1979) propose une méthode du type 'cross-validation'consistant à remplacer h_n par une v. a. r minimisant une erreur quadratique empirique.

2) On note que tous les résultats sur la méthode du noyau(pour l'estimation de la densité ou de la regression) reposent sur un résultat (lemme de Bochner (1955)), relatif à la limite de l'intégrale:

$$\lim_{h\to 0}(g * K_h)(x) = \lim_{h\to 0}\int g(t)(1/h_n^s)K((x-t)/h_n)dt$$

où g est continue en x et

$$K_h(\cdot) \quad = \frac{1}{h_n^s}K(\frac{\cdot}{h_n})$$

Importance des méthodes n'utilisant pas la Statistiques d'ordre

L'importante place qu'occupent les méthodes n'utilisant pas des statistiques d'ordre a une double origine:

1) L'existence de résultats plus précis que de simples convergences

(évaluation du biais et de la variance de $r_n(x)$, la loi limite de *v. a. r.* associées à r_n) .

2) L'importance même du nombre des travaux et la diversification des méthodes: Importance et diversification qui sont liées au considérable développement, depuis Parzen (1962), des travaux sur l'estimation non paramètrique de la densité: Bosq(1973), Deheuvels(1977) ou Roussas (1987)) .

B) **Parmi les méthodes utilisant la statistique d'ordre:**

1) **Régressogramme à pas aléatoire:**

Le plus ancien travail relatif à un estimateur de régression défini à l'aide de statistiques d'ordre est dû à Bhattacharya et Parthasarathy (1961) qui étudient pour X à valeur sans $[0,1]$ et $n = k_n l_n$, la méthode consiste à diviser l'intervalle $[0,1]$ en l_n intervalles disjoints contenant chacun exactement k_n points X_i, $i = 1,\ n,$ et estimer la régression par la fonction en escalier qui, sur chaque intervalle de cette partition est constante et égale à la moyenne des Y_i tels que X_i appartienne à cet intervalle . Cet estimateur est nommé 'régressogramme à pas aléatoire ' en raison de l'analogie de sa définition avec la méthode' régressogramme à pas fixe'

2) **Méthode des k points les plus proches:**

Stone (1976) étudie cette méthode à l'aide d'échantillons empiriques de loi connue et montre que cet estimateur possède la propriété de convergence universelle si et seulement si:

$$k_n/n \to_{n\to\infty} 0,\ k_n\ \to_{n\to\infty} \infty$$

Devroye(1978a) obtient le résultat de convergence uniforme presque sûre:

$$k_n/n \to_{n\to\infty} 0, k_n/\log n \to_{n\to\infty} \infty \Longrightarrow \sup_x |r_n(x) - r(x)| \to_{n\to\infty} 0.$$

lorsque X prend ses valeurs dans un compact de $\mathbb{R}^s$.

2.1.3 Application à la prédiction:

Soit $X_n,\ n = 1,\ 2,\ ...$ un processus réel strictement stationnaire, l'objectif est prédire X_{n+k} à partir des observations $X_1,\ .....,\ X_n,\ n \geq k \geq 1,\ k < \infty$. Pour fixer les idées, on choisit l'horizon $k = 1$ et on suppose que le processus est markovien . Si ce processus est **connu**, le meilleur prédicteur est la régression

$$r(X_n) \quad = E(X_{n+1}/X_n)$$

Lorsque le processus est **inconnu** et non assujetti à des hypothèses permettant d'estimer r à l'aide de méthodes paramètriques, il est naturel d'utiliser comme prédicteur un estimateur non paramètrique de la régression r, défini par exemple (méthode du noyau) comme suit:

$$\hat{r}_n(X_n) = \sum_{i=1}^{n-1} X_{i+1} K((X_n - X_i)/h_n) / \sum_{i=1}^{n-1} K((X_n - Xi)/h_n)$$

Ce prédicteur est introduit par Watson (1964) à propos d'un problème concret relevant de la météréorologie (étude d'une suite de relevés de températures journalières) .

Exemple1: Si X_j désigne la pression moyenne de l'air le jour $j$où $j = 1,\ ...,\ n+1$. Y_j la quantité de pluie tombée le jour $j+1$, alors :

$\hat{r}_n(X_{n+1})$ est un prédicteur(ou une prévision) de Y_{n+1}; i. e un prédicteur de la quantité de la pluie qui tombera le jour $n+2$, lorsqu'on a observé $X_1,\ ...,\ X_{n+1}$ et $Y_1,\ ...,\ Y_n$;

Exemple 2: $X_1,\ ...,\ X_{n+1}$ sont les taux de cholestérol observés dans le sang de $n+1$ malades. $Y_1,\ ...,\ Y_n$ sont les taux de calcium observés dans le sang des n premiers malades: , alors: $\hat{r}_n(X_{n+1})$est un prédicteur(ou une prévision) du taux de calcium chez le $(n+1)^e$malade.

2.2 Estimation de la densité conditionnelle

2.2.1 Introduction et Hypothèses

L'intéret de l'estimation de l'espérance conditionnelle("régression") :

En statistique, on rencontre souvent le probléme de l'approximation, au sens des moindres carrés, de Y par une fonction affine des composantes de X : C'est le probléme de **la régression multiple**. Il peut être géneralisé en cherchant une application mesurable quelconque r, telle que $r(x)$ soit la plus proche de Y au sens des moindres carré. Si Y est de carré intégrable L'espérance de Y conditionnelle à X

$$r(x) \quad = E(Y/X = x)$$

est la solution du probléme. C'est une des raison pour laquelle l'estimation de la" regression" r est un problème fondamentale de la statistique non paramétrique. Sans oublier son importance dans la prévisio

Objectif: Dans ce paragraphe on montre qu'à partir de toute suite d'estimateurs de la densité possédant certaines propriétés asymptotiques, on peut construire une suite d'estimateurs de la "régression" qui converge uniformément presque sûrement . Pour cela, on passe par l'intermédiaire d'un estimateur de **la densité conditionnelle** Bhattacharya et Parthasarathy (1961) . Cette approche **généralise** la méthode du régressogramme et la méthode du noyau.

Les données:

On se place sur $(\mathbb{R}^2,\ \mathcal{B}_{\mathbb{R}^2})$, on désigne par F la fonction de répartition du couple de variables aléatoires $(X,\ Y)$, par f sa densité . On suppose que : F est absolument continue, f bornée et qu'elle est définie partout par la relation:

$$\frac{\partial^2 F}{\partial x \partial y}(x,\ y) \quad = f(x,\ y)$$

$g(x)$ désigne la densité marginale de X :

$$g(x) \quad = \int_{\mathbb{R}} f(x,\ y) \ \ dy$$

$h_{x,a}(y)$ est la densité de Y conditionnée par

$$X \in I_{x,a} = [x - a, x + a[$$

$x \in \mathbb{R},\ a > 0$. $h_x(y)$ est la densité de Y conditionnée par: $X = x,\ x \in \mathbb{R}$où

$$B(x,\ a) \quad = I_{x,a} \times \mathbb{R} = \left\{(x', y') \in \mathbb{R}^2 : x - a \leq x' < x + a, y' \in \mathbb{R}\right\}$$

$$p(x,\ a)\ \ = P\left[(X,Y) \in B(x,a)\right]$$

$(X_i,\ Y_i)\ _{1\leq i\leq n}$ est un échantillon de taille n tiré suivant F (de même loi que $(X,\ Y)$). $l_n(x,\ a)$ est le nombre de points de l'échantillon appartenant à $B(x,\ a)$; $l_n(x,\ a)$ est la réalisation d'une variable aléatoire $L_n(x,\ a)$. Enfin $y_{i_l},\ .\ .\ .\ .\ .\ ,\ y_{i_{l_n}}$ sont les ordonnées des points de l'échantillon appartenant à $B(x,\ a)$. On suppose que a est une suite positive de $n(a = a(n)\ > 0,\ \forall n \in \mathbb{N})$ et telle que:

$$\lim_{n\infty} a(n) = 0 \qquad (1)$$

$a(n)$ joue le role de la fenêtre.

2.2.2 Construction de l'estimateur et Etude de la convergence

Définition d'une classe d'estimateurs.

Pour définir un estimateur de $h_x(y)$, il semble naturel de partir d'un estimateur de la densite univariée . Comme on aura besoin dans la suite, que de certaines propriétés d'un estimateur de la densité, il est inutile de préciser la méthode utilisée pour construire cet estimateur . Cette construction est la suivante: On désigne donc par $\hat{f} = (\hat{f}_n)\ _{n=0,1,2,......}$ une suite d'estimateurs de la densité univariée, pour des raisons qui apparaitront ci-dessous, on pose arbitrairement :

$$\hat{f}_o(y) = 0, \forall y \in \mathbb{R} \quad (2)$$

et on suppose que $\hat{f}_n(y,\ .\ .\ .\ ,\ y; y)$ est une application mesurable de $\mathbb{R}^{n+1}$ dans $\mathbb{R}$ $\forall n \in \mathbb{N}$. Soit alors $\mathcal{F}(\hat{f})$ l'ensemble des densités univariées telles que :

1) $f_n^g(y)\ \ = E_g\left[(\hat{f}_n(y))\right] = \int_{\mathbb{R}^n} \hat{f}_n(y_1,\ .\ .\ .\ .\ .\ ,\ y_n; y)\ g(y_1)\ldots g(y_n)\ dy_1 \ldots\ldots dy_n.$ existe $\forall y \in \mathbb{R},\ \forall n \in \mathbb{N},\ \forall g \in \mathcal{F}(\hat{f})$.

2) $d(f_n^g,\ g)\ \ = \sup_{y\in\mathbb{R}} |f_n^g(y) - g(y)| \to 0$ avec $\frac{1}{n},\ \forall g \in \mathcal{F}(\hat{f})$.

3) Il existe deux constantes positives C_1 et C_2 et une fonction de n, α_n, non décroissante tendant vers $+\infty$ avec n et vérifiant la condition:

$$\lim \frac{n}{\alpha_n^2} = +\infty$$

telles que:

$$P\left[\sup_{y\in\mathbb{R}} \left|\hat{f}_n(y) - f_n^g(y)\right| > \varepsilon\right] \leq C_1 \exp\left[-C_2 \frac{n}{\alpha_n^2} \varepsilon^2\right]$$

$\forall \varepsilon > 0,\ \forall n \in \mathbb{N},\ \forall g \in \mathcal{F}(\hat{f})$.

Remarque: Les conditions précédents impliquent que

$$\sup_{y \in \mathbb{R}} \left| \hat{f}_n(y) - g(y) \right| \to_p 0$$

quand $n \to +\infty,\ \forall g \in \mathcal{F}(\hat{f})$ où $\hat{f}$ étant donnée et x étant fixé, on prend comme estimateur de h_x la fonction aléatoire définie par:

$$\hat{h}_{n,x}(y) = \hat{f}_{l_n}(y_{1_l}, ..., y_{i_{l_n}}; y)\ (\forall y \in \mathbb{R}, \forall n \in \mathbb{N}.\ (3)$$

où: $y_{1_l}, \ldots\ldots, y_{i_{l_n}}$ sont les ordonnées des points de l'echantillon $(X_i,\ Y_i)\ _{i=1,\ldots,n}$ appartenant à $B(x,\ a)\ = I_{x,a} \times \mathbb{R}$.

Etude de la Convergence de l'estimateur

On présente des lemmes préliminaires qui vont nous permettre d'aboutir au théorème: Résultat de convergence de l'estimateur. Etude ponctuelle: x est fixé une fois pour toutes .

Lemme 1: Si $g(x)\ > 0$ et s'il existe $\delta > 0$ tel que, pour $a \leq \delta$, f est lipchitzienne d'ordre γ_1 dans $B(x,\ a)$ et g est lipchitzienne d'ordre γ_2 dans

$$\bar{I}_{x,a} = [x - a, x + a]$$

alors il existe $\delta_1 \in]0, \delta]$ tel que $a < \delta_1$ entraine:

$$\sup_{y \in \mathbb{R}} |h_{x,a}(y) - h_x(y)| \leq A(x) a^{\gamma}. \qquad (4)$$

où $A(x)$est indépendante de y et de a et où l'on a posé:

$$y = \max(\gamma_1,\ \gamma_2)$$

Démonstration: Si $g(x)\ > 0$, on a:

$$F(x + a,\ +\infty)\ - F(x - a,\ +\infty)\ > 0$$

$\forall a > 0$ et l'on peut écrire :

$$h_{x,a}(y) = \frac{\frac{\partial F}{\partial Y}(x + a, y) - \frac{\partial F}{\partial Y}(x - a, y)}{F(x + a, +\infty) - F(x - a, +\infty)}$$

et pour $a \leq \delta$, f est lipchitzienne donc continue et, a fortiori, continue par rapport à x; il en est de même pour g et l'on peut écrire:

$$h_{x,a}(y) = \frac{f(x - a + 2\theta a, y)}{g(x - a + 2\theta' a)}, 0 < \theta < 1, 0 < \theta' < 1$$

d'autre part

$$h_x(y) \quad = \frac{f(x,y)}{g(x)}$$

et comme, pour a assez petit $(a < \delta_1)$ on a:

$$g(x - a + 2\theta' a) \quad > \frac{g(x)}{2}$$

il vient

$$|h_{x,a}(y) - h_x(y)| \leq \frac{2}{g(x)} \, |f(x - a + 2\theta a, y) - f(x,y)| + \frac{2f(x,y)}{g(x)^2} \left| g(x) - g(x - a + 2\theta' a) \right|$$

or, il existe deux constantes k_1 et k_2 telles que: $\forall a \leq \delta$ on a:

$$|f(x - a + 2\theta a, y) - f(x,y)| \leq k_1 a^{\gamma_1} \, |1 - 2\theta|^{\gamma_1} \leq k_1 a^{\gamma_1}$$

et

$$\left| g(x) - g(x - a + 2\theta' a) \right| \leq k_2 a^{\gamma_2} \leq \left| 1 - 2\theta' \right|^{\gamma_2} \leq k_2 a^{\gamma_2}$$

d'où

$$|h_{x,a}(y) - h_x(y)| \leq \frac{2}{g(x)} k_1 a^{\gamma_1} + \frac{2f(x,y)}{g^2(x)} k_2 a^{\gamma_2}$$

or, on sait que:

$$M = \sup_{(x,y) \in \mathbb{R}^2} f(x,y) < +\infty$$

on a donc: immédiatement l'inégalité (4) avec:

$$A(x) = \frac{2}{g(x)} k_1 + \frac{2M}{g(x)^2} k_2 \qquad (5)$$

On a bien $A(x)$ est indépendante de y et de a (la fenêtre) . ■

Pour simplifier l'énoncé du lemme suivant, on pose:

$$h_{n,x}(y) = f^{h_{x,a}}_{L_n(x,a)}(y), \forall n \in \mathbb{R}, \forall y \in \mathbb{R} \qquad (6)$$

autrement dit, $\forall y \in \mathbb{R}$et $\forall n \in \mathbb{N}$, $h_{n,x}(y)$ est une variable aléatoire liée à $L_n(x, \ a)$par la relation fonctionnelle suivante :

$$L_n(x,a) = l \Longrightarrow h_{n,x}(y) = E_{h_{x,a}} \left[\hat{f}_l(y) \right] \qquad (7)$$

$\forall l = 0, \ 1, \ . \ . \ . \ . \ , \ n$. On a alors le lemme suivant:

Lemme 2:

Sous les hypothèses du lemme 1 et en supposant, de plus que:

$$h_{x,a} \in \mathcal{F}(\hat{f})$$

au moins pour a assez petit et que:

$$\lim_{n\to\infty} \frac{na_n}{\alpha_n^2} = +\infty$$

on a:

$$\sup_{y\in\mathbb{R}} \left|\hat{h}_{n,x}(y) - h_{n,x}(x)\right| \to_p 0 \text{ quand } n \to \infty \qquad (8)$$

Démonstration:

En appliquant la formule des probabilités totales, on obtient :

$$P\left[d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon\right] = \sum_{l=0}^{n} C_n^l p(x,a)^l \left[1 - p(x,a)\right]^{n-l} P\left[d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon / L_n = l\right]$$

or d'après (7), on a:

$$P\left[d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon / L_n = l\right] = P\left[d(f_l; f_l^{h_{x,a}}) > \varepsilon\right]$$

Cette derniére probabilité étant calculée par rapport à la loi conditionnelle de densité $h_{x,a}$. Comme on a supposé que pour a assez petit

$$h_{x,a} \in \mathcal{F}(\hat{f})$$

on peut pour n assez grand appliquer 3 et écrire :

$$P\left[d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon\right] \leq C_l \sum_{l=0}^{l} C_n^l p(x,a)^l \left[1 - p(x,a)\right]^{n-l} \exp\left[-C_2 \frac{1}{\alpha_l^2} \varepsilon^2\right]$$

et comme α_l a été supposée non décroissante :

$$P\left[d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon\right] \leq C_l \sum_{l=0}^{n} C_n^l p(x,\ a)\ ^l \left[1 - p(x,a)\right]^{n-l} \exp\left[-C_2 \frac{1}{\alpha_n^2} \varepsilon^2\right]$$

$$\leq C_1 \left[1 - p(x,a)\left[1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2}\right]\right]^n$$

or pour a assez petit on a:

$$p(x,a) = F(x+a,+\infty) - F(x-a,+\infty) = 2ag(\zeta) > 2a.\frac{g(x)}{2} = a.g(x).$$

où $\zeta \in]x-a, x+a[$. Il vient alors :

$$P\left[d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon\right] \leq C_1 \left[1 - a_n g(x)\left[1 - \exp(C_2 \frac{\varepsilon^2}{\alpha_n^2})\right]\right]^n \tag{9}$$

Le deuxième membre de cette inégalité tend vers 0 si et seulement si :

$$n \log\left[1 - a_n g(x)\left[1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2}\right]\right] \to -\infty$$

donc, compte tenu du fait que:

$$\lim_{n\to\infty} a_n = 0$$

Si :

$$n a_n \left[1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2})\right] \to +\infty$$

or $\alpha_n \to +\infty$ donc $1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2}) \;\sim \frac{C_2 \varepsilon^2}{\alpha_n^2}$ et la condition s'écrit finalement. ■

$$\lim_{n\to\infty} \frac{n a_n}{\alpha_n^2} = +\infty \tag{10}$$

Lemme 3:
Sous les hypothèses du lemme 2 et l'hypothèse supplémentaire suivante:

$$\sum_{n=0}^{+\infty} \exp\left[-\lambda \frac{n a_n}{\alpha_n^2}\right] < +\infty, \ \forall \lambda > 0 \tag{11}$$

On a:

$$d(\hat{h}_{n,x}; h_{n,x}) \to_{p.s.} 0 \text{ avec} \frac{1}{n}$$

.

Démonstration : On sait puisque:

$$\log(1-u) \leq -u, \forall u < 1$$

et que

$$n \log\left[1 - a_n g(x)\right]\left[1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2})\right] \leq -n a_n g(x)\left[1 - \exp(-C_2 \frac{\varepsilon^2}{\alpha_n^2})\right] \tag{12}$$

pour n assez grand . D'autre part, on a pour n assez grand :

$$1-\exp(-\frac{C_2\varepsilon^2}{\alpha_n^2}) < 2\frac{C_2\varepsilon^2}{\alpha_n^2} \tag{13}$$

Les deux inégalités précedentes(12) et (13) entraînent alors :

$$n\log\left[1-a_n g(x)\left[1-\exp(-\frac{C_2\varepsilon^2}{\alpha_n^2}\right]\right] \leq -2C_2\varepsilon^2 g(x).\frac{na_n}{\alpha_n^2}$$

d'où, d'après (9) et toujours pour n assez grand :

$$P\left[d(\hat{h}_{n,x};h_{n,x}) > \varepsilon\right] \leq C_1\exp\left[-2C_2\varepsilon^2 g(x)\frac{na_n}{\alpha_n^2}\right] \tag{14}$$

La condition (11) permet alors de conclure via le fameux théorème de **Borel Cantelli:** *p. co* $\Rightarrow$ *p. s.* ■ .

Remarque: Pour le lemme suivant, on aura besoin d'une nouvelle propriété de la classe d'estimateurs utilisée

Définition: On dit que $\hat{f}$ est uniformément asymptotiquement sans biais sur $\mathcal{F}'(\hat{f}) \subset \mathcal{F}(\hat{f})$ si la condition 2) est valable uniformément pour $g \in \mathcal{F}'(\hat{f})$, autrement dit si :

$$\lim_{n\to\infty}\sup_{g\in\mathcal{F}'(\hat{f})} d(f_n^g;g) = 0 \tag{15}$$

Par abréviation $\hat{f}$ est *u. a. b.* sur $\mathcal{F}'(\hat{f})$.

Lemme 4:

Si les conditions suivantes sont vérifiées:

1) $g(x) > 0$

2) g est lipchitzienne dans un voisinage de x

3) $h_{x,a} \in \mathcal{F}(\hat{f}),\ \forall n \in \mathbb{N}$

4) $\hat{f}$ est *u. a. b.* sur l'ensemble$\{h_{x,a_n}\}_n \in \mathbb{N}$.

et $na_n \to +\infty$, alors :

$$d(h_{n,x};h_{x,a}) \to_p 0.$$

Démonstration : D'après (6), on peut écrire :

$$Pd(h_{n,x};h_{x,a}) > \varepsilon) = \sum_{l=0}^{n} C_n^l p(x,a)^l\left[1-p(x,a)\right]^{n-l} Pd(f_l^{h_{x,a}};h_{x,a}) > \varepsilon/L_n = l) \tag{16}$$

Les probabilités qui figurent au second membre de (16) valant 0 ou 1 puisque $d(f_l^{h_{x,a}}; h_{x,a})$ est une variable certaine quand $L_n = l$. Comme $\hat{f}$ est *u. a. b.* sur l'ensemble$\{h_{x,a_n}\}$ $n \in \mathbb{N}$, il existe un entier l_o indépendant de a_n(donc de n) tel que $l > l_o$ entraîne:

$$d(f_l^{h_{x,a}}; h_{x,a}) \quad < \varepsilon$$

$\forall a_n$ donc $\forall n \in \mathbb{N}$. Dans ces conditions, pour $n > l_o$, on peut écrire :

$$P\left[d(h_{n,x}; h_{x,a}) > \varepsilon\right] \leq \sum_{l=0}^{l_o} C_n^l p(x,a)^l \left[1 - p(x,a)\right]^{n-l} \tag{17}$$

Or pour n assez grand, on aura:

$$l_o < np(x,a) - [1 - p(x,a)] \tag{18}$$

En effet : On a vu dans la démonstration du lemme 2 qu'à partir d'un certain rang, on a l'inégalité:

$$p(x,\ a) \quad > ag(x)$$

et comme, par hypothèse, $na_n \to +\infty$ on aura toujours

$$l_o < na_n g(x)$$

à partir d'un certain rang, ce qui entraine (18). Donc, pour $n \geq \nu$,

$$C_n^l p(x,\ a)\ ^l \left[1 - p(x,a)\right]^{n-l}$$

sera fonction non décroissante de l pour $l \leq l_o$ (propriété connue de la loi binominale) .

Dans ces conditions, (17) implique :

$$P\left[d(h_{n,x}; h_{x,a}) > \varepsilon\right] \leq l_o C_n^{l_o} p(x,a)^{l_0} \left[1 - p(x,a)\right]^{n-l_o} \tag{19}$$

et en choisissant n assez grand pour que

$$ag(x) < p(x,a) < 3ag(x) \tag{19bis}$$

on obtient facilement :

$$P\left[d(h_{n,x}; h_{x,a}) > \varepsilon\right] \leq l_o \frac{n^{l_o}}{l_o!} \left[3ag(x)\right]^{l_o} \exp\left[-(n - l_o)ag(x)\right] \tag{20}$$

a partir de (20), on obtient donc une inégalité de la forme:

$$P\left[d(h_{n,x};h_{x,a})>\varepsilon\right]\leq A(na_n)^{l_o}e^{-B' na_n} \tag{21}$$

et à partir d'un certain rang

$$P\left[d(h_{n,x};h_{x,a})>\varepsilon\right]\leq Ae^{-Bna_n} \tag{22}$$

où A et B sont indépendants de n. l'hypothèse $na_n \to +\infty$ permet de conclure ■.

Lemme 5) :

Dans les mêmes conditions du lemme précedent et si de plus

$$\sum_{n=0}^{+\infty}\exp(-\gamma na_n)<+\infty, \forall\gamma>0 \tag{23}$$

Alors :

$$d(h_{n,x};h_{x,a})_{n\to\infty}\to_{p.s}0$$

.

Démonstration : C'est une conséquence immédiate du lemme4, en effet de l'inégalité (22) on permet de conclure■

Convergence presque sûr de l'estimateur.

Theorème 1: Sous les conditions suivantes :

1) pour a assez petit, f est lipchitzienne dans $B(x,\ a)$et g est lipchitzienne dans $\bar{I}(x,\ a)\ \ = [x-a,x+a]$(conditions du lemme1) .

2) $h_{x,a_n}\in\mathcal{F}(\hat{f}),\ \forall n\in\mathbb{N}$ et $\hat{f}$ est $u.\ a.\ b$ sur $\{h_{x,a_n}\}\ n\in\mathbb{N}$(conditions du lemme 4) .

3) $\sum_{n=0}^{+\infty}\exp\left[-\lambda\frac{na_n}{\alpha_n^2}\right]<+\infty,\ \forall\lambda>0$ (conditions du lemme5) On a:

$$d(\hat{h}_{n,x};h_x)_{n\to\infty}\to_{p.s}0.$$

Demonstration: Cette décomposition :

$$d(\hat{h}_{n,x};h_x)\leq d(\hat{h}_{n,x};h_{n,x})+d(h_{n,x};h_{x,a})+d(h_{x,a};h_x) \tag{24}$$

et l'application des lemmes préléminaires précedents permettent de conclure. ■

Etude de la vitesse de convergence

les inégalités (4), (14) et (22) montrent que, pour n assez grand, on a l'inégalité :

$$P\left[d(\hat{h}_{n,x};h_x) > \varepsilon\right] \leq K_1 \exp\left[-\frac{na_n}{\alpha_n^2}K_2\varepsilon^2\right] \tag{25}$$

En examinant les démonstrations des lemmes précédents, on constate que (25) reste valable, uniformément sur un intervalle compact, si les hypothèses faites sont uniformes sur cet intervalle . Les hypothèses du lemme17, la formule (5), les inégalités (19bis), la formule (15), l'inégalité (14) et les valeurs de A et de B dans l'inégalité (22) On peut donc énoncer le résultat suivant:

Théorème 2: Soit $[u,v]$ un intervalle compact tel que :

1) Il existe un voisinage V de $[u,v]$ dans $\mathbb{R}$ tel que f soit lipchitzienne sur $Vx\mathbb{R}$ et g lipchitzienne sur V.

2) $g(x) \;\; > 0, \; \forall x \in [u,v]$.

3) $h_{x,a_n} \in \mathcal{F}(\hat{f}), \; \forall x \in [u,v], \; \forall n \in \mathbb{N}$.

4) $\hat{f}$ est *u. a. b.* sur $\{h_{x,a_n}\}_{\;u\leq x\leq v}$; $n \geq \nu$ où $\nu \in \mathbb{N}$. Alors si

$$\frac{na_n}{\alpha_n^2} \to \infty$$

et si $\varepsilon > 0$ est donnée, il existe deux nombres positifs $K_1(\varepsilon)$ et $K_2(\varepsilon)$ et un entier $N(\varepsilon)$ tels que:

$$\sup_{u\leq x\leq v} P\left[d(\hat{h}_{n,x};h_x) > \varepsilon\right] \leq K_1(\varepsilon)\exp\left[-K_2(\varepsilon)\frac{na_n}{\alpha_n^2}\right] \tag{26}$$

$\forall n \geq \mathcal{N}(\varepsilon)$. Pour préciser un peu plus la vitesse de convergence, supposons que l'on ait:

$$\beta_n = \sup_{u\leq x\leq v, k\geq \nu} d(f_n^{h_{x,a_k}};h_{x,a_k}) \leq C_l' n^{-b}, C_l' > 0, b > 0 \tag{27}$$

$$\alpha_n = C_2'.\; n^c,\; 0 < c < \tfrac{1}{2}$$
(28)
$$a_n = C_3'.\; n^{-d},\; 0 < d < 1-2c$$
(29)

On peut énoncer alors le corollaire suivant:

Corollaire 1: Si:

$$\gamma_1 = \gamma_2 = 1$$

et si

$$0 < \delta < \inf(d,\ b(1-d),\ \frac{1-d}{2} - c)$$

Alors: Sous les hypothèses du théorème 2, on a:

$$n^{\delta} d(\hat{h}_{n,x}; h_x)_{n\to\infty} \to_{p.s.} 0, \forall x \in [u,v] \,.$$

Démonstration : Tout d'abord, compte tenu de (4), (29) et du fait que $\delta < d$, on a $\forall x \in [u,v]$:

$$n^{\delta} d(h_{x,a}; h_x) \to 0 \tag{30}$$

D'autre part, si l'on reprend les démonstrations des lemmes 2 et 3, on arrive à l'inégalité :

$$P\left[d(\hat{h}_{n,x}; h_{n,x}) > \varepsilon . n^{-\delta}\right] \leq C_1 \exp\left[-2C_2\varepsilon^2 n^{-2\delta} g(x) \frac{na_n}{\alpha_n^2}\right] \quad \text{(14bis)}$$

Remarque: (14bis) est la même que (14) avec ε remplacé par $\varepsilon.\ n^{-\delta}$ et le deuxième membre de l'inégalité (14bis) est le terme général d'une série convergente puisque

$$n^{-2\delta} . \frac{na_n}{\alpha_n^2} = \frac{C_3'}{C_2'^2} . n^{1-d-2\delta-2c}$$

et que

$$\delta < \frac{1-d}{2} - c$$

on a donc, $\forall x \in [u,v]$:

$$n^{\delta} . d(\hat{h}_{n,x}; h_{n,x}) \to_{p.s.} 0 \tag{31}$$

Enfin, si l'on reprend le démonstration du lemme 4, on constate que l'inégalité:

$$d(f_l^{h_{x,a}}; h_{x,a}) \quad < \varepsilon . n^{-\delta}$$

est vérifiée pour $l > l_o$ avec l_o tel que:

$$C_l' . l_o^{-b} < \varepsilon . n^{-\delta} \tag{32}$$

En prenant:

$$l_o = l_o(n,\ \varepsilon) \quad = K n^{\frac{\delta}{b}}$$

avec

$$K > (\frac{C'_l}{\varepsilon})^{\frac{l}{b}}$$

on a (32) et comme $\delta < b$ on a aussi $l_o < n$, au moins pour n assez grand : les inégalités (18), (19) et (20) restent donc valables si l'on remplace ε par $\varepsilon n^{-\delta}$. Il en est de même pour (21) puisque :

$$\exp\left[a_n \; l_o g(x)\right] \to l$$

quand $n \to +\infty$. d'autre part, comme

$$\log(na_n)^{\;\; l_o} = Kn^{\frac{\delta}{b}}(1-d) \quad \log {C'_3}^{\frac{l}{l-d}}.n$$

on a

$$\forall u > 0 : (na_n)^{\;\; l_o} < e^{n^{\frac{\delta}{b}+u}}$$

au moins pour n assez grand . en choisissant

$$u < 1 - d\frac{\delta}{b}$$

il vient :

$$P\left[d(h_{n,x}; h_{x,a}) > \varepsilon n^{-\delta}\right] < A' e^{-B' n a_n}$$

où A' et B' sont positifs et indépendants de n. donc:

$$n^{\delta} d(h_{n,x}; h_{x,a}) \to_{p.s.} 0 \tag{33}$$

ce qui avec (30) et (31) permet de conclure . ■

2.2.3 Application a la méthode du noyau

Examinons comment on peut appliquer les résultats obtenus aux deux classes usuelles d'estimateurs de la densité puis de la densité conditionnelle.

Méthode du noyau

L'estimateur utilisé est de la forme :

$$\hat{f}_n(y) \quad = \frac{\alpha_n}{n}\sum_{i=1}^{n} K\left[\alpha_n(y - y_i)\right] \tag{34}$$

où les y_i sont les points de l'échantillon et où K est une densité à variation bornée (un noyau), alors d'après un théorème de Nadaraya on sait que si g

est une densité uniformément continue, $g \in \mathcal{F}(\hat{f})$ de plus $\forall g \in \mathcal{F}(\hat{f})$ on a l'inégalité (Le fameux lemme de Bochner)

$$d(E_g(\hat{f}_n),\ g) \quad \leq \sup_{y\in\mathbb{R}} . \sup_{|z|\leq\delta} |g(y-z) - g(y)| + 2 \sup_{y\in\mathbb{R}} g(y). \int_{|u\geq\delta\alpha_n|} K(u) \ \ du$$

$\forall\delta > 0,\ \forall n \in \mathbb{N}$. On en déduit immédiatement que si $(g_i)\ _{i\in I}$ est une famille de densités uniformément continues, uniformément bornées et uniformément équicontinues, alots $\hat{f}$ est $u.\ a.\ b.$ sur $(g_i)\ _{i\in I}$.

Conséquence : Pour $\hat{f}_n$, si les hypothèses 1) et 2) du théorème 2 sont vérifiées, les hypothèses 3) et 4) le sont aussi. car la relation:

$$h_{x,a}(y) = \frac{f(x-a+2\theta a, y)}{g(x-a+2\theta' a)} \tag{35}$$

où $0 < \theta < 1,\ \ 0 < \theta' < 1$, montre que la famille $\{h_{x,a_n}\}\, n \in \mathbb{N}$, vérifie les conditions indiquées ci-dessus Enfin, d'après les résultats de Schuster(1969) on voit que, sous les hypothèses 1) et 2) du théorème 2, et en supposant, de plus .

$$\sup_{n\in\mathbb{N}} \left| h'_{x,a_n} \right| = D < +\infty$$

si l'on choisit $\delta \in \left]0, \frac{1}{4}\right[$ on aura:

$$n^{\delta}.d(\hat{h}_{n,x}; h_x)_{n\to\infty} \to_{p.s.} 0, \forall x \in [u,v]\,.$$

2.3 Estimation des fonctions de régression

2.3.1 Cas : Y est bornée

Soit $[u,v]$ un intevalle compact pour lequel les hypothèses du théorème 2 sont vérifiées . On suppose que

$$|Y| \leq M$$

où M est un nombre positif connu . On choisit a_n tel que

$$k_n = \frac{v-u}{2a_n}$$

soit entier, $\forall n \in \mathbb{N}$ et posons : $u_1 = u + a_n,\ u_2 = u + 3a_n,\ \ldots\ ,\ u_{k_n} = u + (2k_n - 1)\ \ a_n$. Soit alors:

$$r_\alpha(x) = \int_{-M}^{+M} y^\alpha h_x(y) dy,\ \alpha > 0$$

une fonction de régression de Y en X . On définit un estimateur $r_\alpha(x)$ en posant :

$$\hat{r}_{n,\alpha}(x) = \int_{-M}^{+M} y^\alpha \hat{h}_{n,u_i}(y)dy$$

pour $x \in [u + (2i-2)a_n, u + 2ia_n[, \forall i = 1, \ldots\ldots, k_n - 1$ et:

$$\hat{r}_{n,\alpha}(x) = \int_{-M}^{+M} y^\alpha \hat{h}_{n,u_{k_n}}(y)dy$$

pour: $x \in [u + (2k_n - 2)a_n, u + 2k_n a_n]$.

Théorème 3:

Sous les hypothèses du théorème 2 et l'hypothèse supplémentaire:

$$\sum_{n=0}^{\infty} a_n^{-1} \exp\left[-\lambda \frac{na_n}{\alpha_n^2}\right] < +\infty, \forall \lambda > 0 \ (36)$$

On a:

$$\lim_{n\to\infty} p.s. \sup_{x\in[u,v]} |\hat{r}_{n,\alpha}(x) - r_\alpha(x)| = 0, \forall \alpha > 0.$$

Démonstration : D'aprés les hypothèses 1) et 2) du théorème 2, on voit immédiatement que $h_x(y)$ est lipchitzienne sur $[u,v] \times \mathbb{R}$ d'où l'on déduit que $r_\alpha(x)$ l'est aussi sur $[u,v]$, $\forall \alpha > 0$: Il existe donc deux constantes positives δ et γ telles que

$$|r_\alpha(x) - r_\alpha(u_i)| \leq \delta\, |x - u_i|^\gamma$$

$\forall x \in [u,v]$, $\forall i \in \{1, ..., k_n\}$ d'où pour $x \in [u_i - a_n, u_i + a_n]$ on a:

$$|r_\alpha(x) - r_\alpha(u_i)| \leq \delta . a_n^\gamma$$

$\forall i \in \{1, ., k_n\}$. Compte tenu du fait que :

$$\lim_{n\to\infty} a_n = 0$$

et de l'inégalité:

$$|\hat{r}_{n,\alpha}(x) - r_\alpha(x)| \leq |\hat{r}_{n,\alpha}(x) - r_\alpha(u_i)| + |r_\alpha(u_i) - r_\alpha(x)|$$

$\forall x \in [u_i - a_n, u_i + a_n]$, $\forall i \in \{1, ..., k_n\}$, il vient pour n assez grand :

$$P\left[\sup_{x\in[u,v]} |\hat{r}_{n,\alpha}(x) - r_\alpha(x)| > \varepsilon\right] \leq P\left[\sup_{l\leq i\leq k_n} |\hat{r}_{n,\alpha}(u_i) - r_\alpha(u_i)| > \frac{\varepsilon}{2}\right] \quad (37)$$

où ε est un nombre positif donné. Or, $\forall i \in \{1, ..., k_n\}$, on a l'inégalitè

$$|\hat{r}_{n,\alpha}(u_i) - r_\alpha(u_i)| \leq \int_{-M}^{+M} |y|^\alpha \, dy. \sup_{|y|\leq M} \left|\hat{h}_{n,u_i}(y) - h_{u_i}(y)\right|$$

d'où

$$P\left[\sup_{x\in[u,v]} |\hat{r}_{n,\alpha}(x) - r_\alpha(x)| > \varepsilon\right] \leq \sum_{i=1}^{k_n} P\left[\sup_{|y|\leq M} \left|\hat{h}_{n,u_i}(y) - h_{u_i}(y)\right| > \frac{\alpha+1}{4}.\frac{\varepsilon}{M^{\alpha+1}}\right]$$

en appliquant (26) on obtient, en posant

$$\frac{\alpha+1}{4}.\frac{\varepsilon}{M^{\alpha+1}} = \eta$$

:

$$P\left[\sup_{x\in[u,v]} |\hat{r}_{n,\alpha}(x) - r_\alpha(x)| > \varepsilon\right] \leq \frac{v-u}{2a_n} K_l(\eta) \quad \exp\left[-K_2(\eta)\frac{na_n}{\alpha_n^2}\right]$$

$\forall n > N(\eta)$ et l'hypothèse(36) permet de conclure . ■

2.3.2 Cas : Y n'est pas nécessairement borné

Moyennant de légères modifications, la méthode précédente s'applique au cas où Y n'est pas nécessairement borné : on suppose que les hypothèses du théorème 2 et du corollaire sont vérifiées et que :

$$r_\alpha(x) \quad = \int_{\mathbb{R}} y^\alpha h_x(y) \quad dy$$

$\alpha > 0$ existe, $\forall x \in [u, v]$. On définit alors un estimateur de $r_\alpha(x)$ en posant :

$$r^*_{n,\alpha}(x) = \int_{-T_n}^{+T_n} y^\alpha \hat{h}_{n,u_i}(y)dy$$

pour $x \in [u + (2i-2)a_n, \ u + 2ia_n]$ $i = 1, \ . \ . \ . \ . \ . \ , \ k_n - 1$ et

$$r^*_{n,\alpha}(x) \int_{-T_n}^{+T_n} y^\alpha \hat{h}_{n,u_{k_n}}(y)dy$$

pour $x \in [u + (2k_n - 2)a_n, v]$, on aura alors:

Théorème 4 Si:

1) Les hypothèses du théorème 2 et de son corollaire sont vérifiées .
2) $\lim_{T+\infty}$. $\sup_{x\in[u,v]}\int_{|y\geq T|}|y|^{\alpha}\, h_x(y)\quad dy=0.$
3) $T_n=n^{\mu}$ avec $0<\mu<\frac{1}{\alpha+1}$. $\inf(d,\ b(1-d),\ \frac{1-d}{2}-c)$. Alors :

$$\lim_{n\to\infty} p.s. \sup_{x\in[u,v]}\left|r^*_{n,\alpha}(x)-r_{\alpha}(x)\right|=0.$$

Démonstration : Posons

$$r_{n,\alpha}(x)\quad=\int_{-T_n}^{+T_n} y^{\alpha}h_x(y)\quad dy$$

d'après le 2), il suffit de montrer que

$$\lim_{n\to\infty}\sup_{x\in[u,v]}\left|r^*_{n,\alpha}(x)-r_{n,\alpha}(x)\right|=0.\ p.s$$

la démonstration est alors analoque à celle du théorème 3 : $r_{n,\alpha}(x)$ est lipchitzienne d'ordre 1 et il existe une constate δ telle que :

$$|r_{n,\alpha}(x)-r_{n,\alpha}(u_i)|\leq\frac{2}{\alpha+1}.\delta.T_n^{\alpha+1}.a_n$$

$\forall x\in[u_i-a_n,u_i+a_n]\,,\ \forall i\in\{1,....,k_n\}$. comme on a

$$\mu<\frac{d}{\alpha+1}$$

il vient

$$\lim_{n\to\infty}T_n^{\alpha+1}.a_n=0$$

on a donc, pour n assez grand et ε positif donné, une inégalité analogue à (37), c'est -à-dire que :

$$P\left[\sup_{x\in[u,v]}\left|r^*_{n,\alpha}(x)-r_{n,\alpha}(x)\right|>\varepsilon\right]\leq P\left[\sup_{l\leq i\leq k_n}\left|r^*_{n,\alpha}(u_i)-r_{n,\alpha}(u_i)\right|>\frac{\varepsilon}{2}\right]$$

or, $\forall i\in\{1,....,k_n\}$ on a l'inégalité:

$$\left|r^*_{n,\alpha}(u_i)-r_{n,\alpha}(u_i)\right|\leq\int_{+T_n}^{+T_n}|y|^{\alpha}\,dy.\sup_{|y|\leq T_n}\left|\hat{h}_{n,u_i}(y)-h_{u_i}(y)\right|$$

d'où: $P\left[\sup_{x\in[u,v]}\left|r^*_{n,\alpha}(x)-r_{n,\alpha}(x)\right|>\varepsilon\right]\leq\sum_{i=l}^{k_n}P\left[\sup_{y\in\mathbb{R}}\left|\hat{h}_{n,u_i}(y)-h_{u_i}(y)\right|>\frac{\varepsilon(\alpha+1)}{4.T_n\alpha+1}\right]$
(38)

or dans la démonstration du corollaire du théorème 2 on a: Pour

$$0 < \delta < \inf(d,\ b(1-d),\ \frac{1-d}{2} - c)$$

une inégalité de la forme:

$$\sup_{x\in[u,v]} P\left[n^{\delta}.d(\hat{h}_{n,x}; h_x) > \varepsilon\right] \leq C_1 \exp\left[-C_2 n^{\beta}\right] \tag{39}$$

où $C_1,\ C_2$ et β sont positifs. Alors (38) et (39) entraînent :

$$P\left[\sup_{x\in[u,v]} \left|r^*_{n,\alpha}(x) - r_{n,\alpha}(x)\right| \varepsilon\right] \leq \frac{v-u}{2a_n} C_1 \exp\left[-C_2 n^{\beta}\ \right] \tag{40}$$

car

$$0 < \mu(\alpha+1) \quad < (d,\ b(1-d),\ \frac{1-d}{2} - c)$$

or on a:

$$a_n = C'_3.n^{-d}$$

$(d > 0)$, donc le deuxième membre de (40) est le terme général d'une série convergente et la démonstration est ainsi achevée. ■

Cas: particulier important(Le régressogramme)

Tukey(1961) a proposé d'estimer la courbe de régression sur un intervalle compact $[u, v]$ à l'aide du régressogramme . Cet estimateur est défini de la manière suivante : On fait une partition de $[u, v]$ en intervalles de même longueur et on estime la fonction de régression par la fonction en escalier qui, dans chaque intervalle de la partition, est constante et égale à la moyenne des ordonnées des points de l'échantillon dont l'abscisse appartient à l'intervalle considéré, et 0 si aucun point de l'échantillon n'est dans ce cas. Tukey pose alors le problème du choix de la partition : le théorème 3 permet de répondre partiellement à cette question ;en effet, le régressogramme appartient à la famille d'estimateurs de la régression qu'on vient d'étudier . Pour le voir, il suffit d'estimer $h_x(y)$ à l'aide d'un noyau K possédant les propriétés suivantes : Il existe un nombre positif A vérifiant $K(t) \quad = 0$ pour $|t| \geq A$ et tel que:

$$\int_{-A}^{+A} K(t)dt = 1 \tag{41}$$

$$\int_{-A}^{+A} t.K(t)dt = 0 \tag{42}$$

L'estimateur de $r_1(x) \quad (\alpha = 1)$ sera alors défini par:

$$r^*_{n,1}(x) = \frac{\alpha_{l_n}}{l_n} \sum_{k=1}^{l_n} \int_{-T_n}^{+T_n} yK\left[\alpha_{l_n}(y - y_{i_k})\right] dy, \text{si } l_n \geq l.$$

et $r^*_{n,l}(x) \quad = 0$, si $l_n = 0$ d'où en faisant les changements de variables $t_{i_k} = \alpha_{l_n}(y - y_{i_k}),\ \ k = 1,\ \ldots\ldots,\ l_n$. $r^*_{n,1}(x) \quad = 0,$ si $l_n = 0$,

$$r^*_{n,1}(x) = \frac{1}{l_n} \sum_{k=1}^{l_n} y_{i_k} \int_{\alpha_{l_n}(-T_n - y_{i_k})}^{\alpha_{l_n}(T_n - y_{i_k})} K(t_{i_k})\ dt_{i_k} - \frac{1}{l_n \alpha_{l_n}} \sum_{k=1}^{l_n} \int_{\alpha_{l_n}(-T_n - y_{i_k})}^{\alpha_{l_n}(T_n - y_{i_k})} t_{i_k} K(t_{i_k})\ dt_{i_k}$$

si $l_n \geq l$. Comme $\lim_{n\to\infty} \alpha_n = +\infty$ on peut toujours choisir $\alpha_n \geq 1,\ \forall n \in \mathbb{N}$;alors, si l'on suppose

$$|y| \leq M$$

et si l'on tient compte du fait que:

$$\lim_{n\to\infty} T_n = +\infty$$

il vient pour n assez grand :

$$\alpha_{l_n}(T_n - y_{i_k}) \quad \geq T_n - M > A$$

et

$$\alpha_{l_n}(-T_n - y_{i_k}) \quad \leq -T_n + M < A$$

d'où, en tenant compte de (41) et (42) (spécifités du noyau K) :

$$r^*_{n,1}(x) = \frac{1}{l_n} \sum_{k=1}^{l_n} y_{i_k} \text{ si } l_n \geq l$$

$r^*_{n,1}(x) \quad = 0$ si $l_n = 0$ et on obtient ainsi: Le régressogramme associé à la partition de $[u, v]$ en intervalles de longueur $2a_n$. Remarquons que l'estimateur obtenu ne dépend pas de la suite α_n. on aura donc le corollaire suivant:

Corollaire: Sous les hypothèses du théorème 4, le régressogramme associé aux intervalles de longueur

$$0(n^{-d})$$

$0 < d < 1$, converge presque sûrement vers $r_1(x)$, uniformément pour $x \in [u, v]$.

2.3.3 Synthèse et Conclusions:

L'approche qu'on vient de présenter est la synthèse et l'adaptation de plusieurs articles dont l'article principale est celui de Bosq(1969) qui a repris les travaux de Nadaraya(1965) et de Tukey(1961).

C'est une méthode analytique, parmi ses avantages: Elle permet de retrouver d'autres méthodes comme celle du régressogramme(la plus ancienne), outre elle utilise des rudiments simples dans le raisonnement (propriétes simples d'analyse et de probabilité).

PartieII
Introduction Générale

Dans cette partie on présente une étude détaillée du paramètre de position **(médiane)**et de régression (**médiane conditionnelle**), on remarque une grande similitude avec **l'espérance conditionnelle** paramètre plus connu.

On constate que la médiane conditionnelle n'est autre que la médiane par rapport à la loi conditionnelle i. e: Le quantile conditionnel d'ordre $\alpha = \frac{1}{2}$, qu'on note m.

Ainsi dans tout ce qui suit $m(x)$désigne la médiane conditionnelle de la variable aléatoire (v, a)Y sachant la v, a $X = x$ avec $(X,\ Y) \in (\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ i. e: $m(x)$ est caractérisé par l'équation :

$$F(m(x)/x) = \frac{1}{2}$$

où $F(Y/x)$désigne la fonction de répartion conditionnelle de Y sachant $X = x$, Le cas: $d_2 = 1$(cas unidimensionnel) est traité dans le chapitre 4, On a énoncé: Les différentes définitions et les propriétés les plus importantes (Inégalité de Jensen, de Chebychef et la relation avec l'espérance conditionnelle ainsi les propriétes essentielles d'éxistence et la non unicite du paramètre). Outre, on a remarqué pour certaines classes de lois dont les lois conditionnelles sont symétriques comme les lois uniformes et les lois normales, les deux paramètres (espérance conditionnelle et médiane conditionnelles) coincident.

En voila un exemple simple pour ilustrer cette egalité.

Exemple: Soit la loi uniforme sur un pavé de $\mathbb{R}^2$ du couple$(X,\ Y)$ éxprimée par sa densité:

$$f(x, y) = 2I_T$$

où I est la fonction indicatrice sur le triangle T de sommets: $(0,\ 0)$, $(0,\ 1)$, $(1,\ 0)$;un calcul simple donne la densité conditionnelle de Y sachant $X = x$,

$$f(Y/x) = \frac{1}{1-x} \text{ pour } 0 \leq x \leq 1 \text{ et } 0 \leq y \leq 1-x$$

donc l'éspérancce conditionnellede Y sachant $X = x$ noté:

$$r(x) = E(Y/X = x) = \frac{1}{2}(1-x) \text{ pour } 0 \leq x \leq 1$$

de même la fonction de répartition conditionnelle de Y sachant $X = x$:

$$F(y/x) = \frac{y}{1-x} \text{ pour } 0 \leq x \leq 1.$$

Ainsi la médiane conditionnelle de Y sachant $X = x; m(x)$qui est solution de l'équation:

$$m(x) = F^{-1}(\frac{1}{2}/x)$$

est égale à:

$$r(x) = \frac{1}{2}(1-x)$$

Ensuite, on présente l'estimation (lorsque la loi $P_{X,Y}$ du couple $(X,\ Y)$ est inconnue) par la méthode du médianogramme-une méthode similaire à celle du régressogramme utilisée pour l'estimation de l'espérance conditionnelle. On remarque que le paramètre n'ést pas sans biais- il est biaisé-c'est une conséqence du théorème de Bickel Leymann, donc l'étude se fait asymptotiquement: Elle consiste à faire intervenir une suite positive $h(n)$avec

$$\lim_{n\to\infty} h(n) = 0$$

$h(n)$ est appelée fenêtre, et l'objectif de toute l'étude: est de trouver des conditions -conditions de régularités mathématiques -sur la loi du couple, la loi conditionnelle et sur cette suite au voisinage de l'infini pour réaliser la convergence de l'éstimateur:

$$\hat{m}_{h(n)} = \hat{F}^{-1}_{h(n)}(\frac{1}{2}/x)$$

vers $m_{h(n)}(x)$ médiane conditionnelle locale(X au voisinage de$x$) puis vers $m(x)$ (globale)en prenant, en premier lieu :

Le couple $(X,\ Y)$ issu des observations $(X_{i,}Y_i)$ non corrélées à travers d'un n **échantillon** $(X_i,\ Y_i)_{i=1,n}$ et en second cas le couple $(X,\ Y)$de même loi que les observations$(X_i,\ Y_i)_{i=1,n}$corrélées à partir de l'étude des **processus** $U_i = (X_i,\ Y_i)_{i=1,n}$($\alpha-$ mélangeants) avec le cas géométriquement mélangeant comme cas particulier:

$$\alpha(k) = a\rho^k, a > 0 \text{ et } 0 < \rho < 1)$$

Les mêmes étapes ont été reprises avec la méthode du noyau de convolution identique à celle utilisée pour l'estimation de l'espérance conditionnelle.

Le chapitre 5 est réservée à la généralisation au cas multidimensionnel $(d_2 > 1)$.

L'étude des différentes définitions, les propriétés ainsi l'estimation de la médiane conditionnelle multivariée avec les deux méthodes: L 'estimateur empirique(médianogramme) et la méthode du noyau ont été abordées.

Quant au dernier chapitre est consacrée à l'application à la prévision dans le cas d'un processus strictement stationnaire, Markovien d'ordre k. La comparaison avec le modèle paramétrique, à travers la déscription de la démarche de **Box et Jenkis en quatre étapes** pour le modèle

$$ARIMA(p, q, d)$$

clos ce chapitre.

Avant de commencer l'etude de la médiane conditionnelle, on présente dans le chapitre 3 des résultats importants pour la suite de notre étude concernant la médiane (cas univarié et cas multivariés), les différentes caractérisations et les lois limites, ainsi que l'estimation dans le cas d'un échantillon et dans le cas d'un processus .

Chapter 3

Estimation Non Paramètrique de la Médiane

3.1 Cas univarié

Définitions et Propriétés

Fonction quantile: Soit X une variable aléatoire réelle (v. a. r) de loi P_X de fonction de répartition F et $\alpha \in]0,1[$ .La fonctions quantile d'ordre α de X (ou de la loi P_X) noté $q_X(\alpha)$ou tout simplement $q_\alpha$est : l'inverse de sa fonction de répartition. Quand cette derniére est strictement croissante, son inverse est définie sans aucune ambiguité. Mais une fonction de répartition reste constante sur tout intervalle dans lequel la variable ne peut pas prendre des valeurs . Pour cela, on énonce la définition suivante:

Définition 1:

$$q_X(\alpha) = \inf \{x : F_X(x) \geq \alpha\}$$

par convention: $q_X(0)$ est la plus petite des valeurs possibles pour X
et $q_X(1)$ est la plus grande, elles sont éventuellement infinies.

Remarque 1: La fonction quantile est un moyen de décrire la dispersion d'une loi.

Une valeure importante est la médiane: $m = q_X(\frac{1}{2})$

Remarque 2: Les valeurs de la fonction quantiles sont souvent plus utilisées en statistique que les valeurs de la fonction de répartition, on note à titre d'exemple l'utilisation fréquente des intervalles de dispersion

$$I_\alpha = [q(\beta), q(1 - \alpha + \beta)]$$

avec $0 \leq \beta \leq \alpha$, $I_\alpha$est appelé intervalle de dispersion de niveau $1 - \alpha$

3.1.1 Définitions et Propriétés

Définitions de la Médiane d'une $v.\ a$ **univariée.**

Soit X une $v.\ a.\ r.$ Notons m_X la médiane de la variable aléatoire X de loi P_X, de fonction de répartition F, de densité f.

Définition 1: m_X est caractérisé par :

$$P(X \geq m_X) \geq \frac{1}{2} \leq P(X \leq m_X)$$

Définition 2: : m_X est caractérisé par :

$$F(m_X) = \frac{1}{2}$$

Remarque 1: cette équation n'a de solution unique que dans des cas rares.

Proposition 1: Lorsque F est continue et strictement croissante ($F' = f > 0$) Alors:

$$\exists ! m_X : F(m_X) = \frac{1}{2}$$

Neamoins on peut donner une définition plus génerale:

Définition 3: On pose : $\mathcal{N} = \left\{x : F(x) \geq \frac{1}{2}\right\}$, $\mathcal{N}' = \left\{x : F(x) \leq \frac{1}{2}\right\}$

$$m_X = \inf \mathcal{N}$$

et

$$\acute{m}_X = \sup \mathcal{N}'$$

sont des médianes de X et toute valeur intermediaire est considérée comme médiane:

$$(m_X + m'_X)/2$$

Exemples de médiane de lois connues: On présente des quantiles q_αd'ordre α $(0 < \alpha < 1)$ de quelques lois connues et on remplace α par $\frac{1}{2}$ pour obtenir la médiane.

1-lois uniforme sur $[0, 1]$: $q_\alpha = \alpha$ et ainsi la médiane $m_X = \frac{1}{2}$

2-Lois Normale $N(0,\ 1)$: $q_\alpha = \phi^{-1}(\alpha)$ et la médiane est $m_X = 0$

3-Lognormale : $q_\alpha = Exp(\phi^{-1}(\alpha))$.

4-Weibull : $q_\alpha = \frac{1}{\beta}\left\{\log(1-\alpha)^{-1}\right\}^{\beta}$.

5-Logistique: $q_\alpha = \log(\alpha/1-\alpha)$.

6-Cauchy: $q_\alpha = \log \pi(\alpha - 1/2)$.

7- Paréto: $q_\alpha = \frac{1}{\beta}(1-\alpha)^{-\beta}$.

8-Exponentielle$\mathcal{E}(\lambda)$: $q_\alpha = -\frac{1}{\lambda}\log(1-\alpha)$

où $(\phi(x) = \frac{1}{2\pi}\int_{-\infty}^{x}\exp(-\frac{t^2}{2})dt$: fonction de répartition de la loi $N(0,\ 1)$

Remarque: pour **les lois discrètes,** la fonction quantile est **une fonction en escalier**, comme la fonction de répartition, si X prend les valeurs $x_k,\ k = 1,\ 2.\ .\ .\ .\ .$, rangées par ordre croissant, la fonction de répartition (cumulative)est égale à :

$F_k = \sum_{j=1}^{k} P(X = x_j)$sur l'intervalle $[x_k, x_{k+1}[$. la fonction quantile vaut :

$Q_X(u) = x_1 \quad$ pour $\quad u \in]0, F_1]$

$\qquad\quad = x_k \quad$ pour $\quad u \in]F_k, F_{k+1}]$

par exemple, pour la loi géométrique $G(p)$ la fonction quantile est la fonction qui, pour tout $k = 1,\ 2,\ .\ .\ .\ .\ .$, vaut k sur l'intervalle $\left]1-(1-p)^{k-1}, 1-(1-p)^k\right]$,

Propriétés de la médiane d'une $v.\ a.$ **univariée** :

Propriété 1: **Non unicité de la médiane**

Provient de la définition 2 que la médiane d'une $v.\ a.\ r.\ X$

est unique que dans des cas rares, mais il existe toujours une **"plus petite** " et une "**plus grande** " médiane.

Toute valeur intermédiaire est considérée comme médiane de X.

Propriété2: relation entre m_X **et** $E(X)$ **(La moyenne de** X**)**

Proposition 2:

$$\forall u > 0, |m_X - E(X)| \leq \frac{1}{u}\log 2\mathcal{E}_X(u)$$

où: $\mathcal{E}_X(u) = E(\exp u\,|X - E(X)|)$

Démonstration : D'après l'inégalité de Chernov (shorack, Welner(1986)) on a:

$$\forall u > 0, P(|X - E(X)| \geq \alpha) \leq \frac{E(\exp u\,|X - E(X)|)}{\exp \alpha u} = 1/2$$

et:

$$\frac{\mathcal{E}_X(u)}{\exp \alpha u} = 1/2 \Longrightarrow \alpha = \frac{1}{u}\log(2\mathcal{E}_X(u))$$

donc:

$$E(X) - \frac{1}{u}\log(2\mathcal{E}_X(u)) \leq m_X \leq E(X) + \frac{1}{u}\log(2\mathcal{E}_X(u)) \ \blacksquare$$

Caractérisation: Réalisation de l'inf **au sens de** L^1, m_X vérifie la propriété suivante:

$$\|X - m_X\|_{L^1} \leq \|X - c\|_{L^1}\,\forall c \in \mathbb{R}$$

Poposition 3: On a:

$$E(|X - m_X|) = \inf \{E(|X - a|, a \in \mathbb{R}\}$$

Pour toute médiane de X.

Remarque sur l'Estimation de la médiane: Non existence d'un estimateur sans biais:

Conséquence du Théorème de Bickel et Lehmann.

On présente **un contre exemple**. Soient: $\mathcal{D}$ Ensemble des mesures de probabilité absolument continue,

$\Omega = \Theta = \mathbb{R}$, on a $\mathcal{D}$ convexe. Soit P_X la loi uniforme de X sur $[-1, 0]$ sa densité est:

$$f(x) = 1_{[-1,0]}(x)$$

La fonction de répartition:

$$F(x) = 0 \quad \text{si: } x < 1,\ x + 1 \quad \text{si: } x \in]-1, 0] \quad \text{et } 1 \quad \text{si : } x > 0.$$

Soit: Q_X la loi exponentielle sur $\mathbb{R}^+$ sa densité est:

$$g(x) = e^{-x} 1_{\mathbb{R}+}$$

sa fonction de répartition $G(x)$ est alors définie par:

$$G(x) = 1 - \exp -x, x \geq 0.$$

Soit: $\alpha \in]0, 1[$, la loi $\alpha P_X + (1 - \alpha) Q_X \in \mathcal{D}$ est de densité: $\alpha f(x) + (1 - \alpha) g(x)$. Déterminer la médiane de $\alpha P_X + (1 - \alpha) Q\ _X$ revient à résoudre l'équation :

$$\alpha F(x) + (1 - \alpha) G(x) = 1/2,\ (1)$$

Pour cela, on distingue deux cas :

(i). $x < 0$, (1)$\iff \alpha(1 + x) = 1/2$.

(ii). $x \geq 0$, (1)$\iff \alpha + (1 - \alpha)(1 - \exp -x) = 1/2$

En (i). (1) n'a de solution m que si $\alpha < 1/2$ et

$$m = \tfrac{1}{2\alpha} - 1.$$

En (ii). (1) n'a de solution m que si $\alpha < 1/2$ et

$$m = \log 2(1 - \alpha).$$

Aucune de **ces solutions n'est polynôme en** α. Ce qui démontrela propriété suivante:

Propriété 3: La non existence d'estimateur sans biais de la médiane.

3.1.2 Estimation de la médiane et lois limites:

On étudie deux cas, variables corrélées et non corrélées.

♦ **I. Variables non corrélées: Cas d'un échantillon:**

$(X_1, \ . \ . \ . \ , \ X_n)$désigne un échantillon issu de X associé au modèle $(\mathbb{R}^n, \ \mathcal{B}_{\mathbb{R}^n})$

Nous allons présenter et étudier plusieurs estimateurs (quatre) non paramétrique de la médiane de la $v. \ a. \ r. \ X$ à partir des observations $(X_i \ )_{i=1,n}$.

Des résultats sur différents types de convergences seront aussi exposés .

1-Estimateur empirique : (Statistique d'ordre)

a) Définition de la statistique d'ordre : On appelle statistique d'ordre l'application mesurable T :

$(\mathbb{R}^n, \ \mathcal{B}_{\mathbb{R}^n}) \to (\mathbb{R}^n, \ \mathcal{B}_{\mathbb{R}^n})$ classant les $(X_i)_{i=1,n}$ par ordre croissant .

$(X_1, \ . \ . \ . \ . \ , \ X_n) \longmapsto (X_{(1)}, \ . \ . \ . \ . \ . \ . \ , \ X_{(n)})$, avec $X_{(1)} \leq X_{(2)} \leq \ldots\ldots \leq X_{(n)}$

$X_{(k)}$ n'a de sens que si n est fixé . sinon $X_{(k)}$ serait noté $X_{(k:n)}$.

b) Définition de l'estimateur empirique : Soit X une $v. \ a. \ r.$, $\{X_i\}_{i=1,n}$ une suite d'observations i. i. d. de même loi que X. Un estimateur de la médiane de X noté $\hat{m}_n$ est donné par:

$$\hat{m}_n = X_{([n/2]+1)}$$

où [] désigne la partie entière.

L'estimateur empirique du α-quantile est donné par :

$$\hat{q}_n(\alpha) = X_{([\alpha n]+1)}$$

Plusieurs travaux ont été consacrés à la recherche de ces lois. On énonce le résultat suivant et la généralisation aux $\alpha-$quantiles avec des exemples.

Théorème 1 :

Si F est continue et strictement croissante et $F'(m_X) = f(m_X) > 0$

Alors :

$$n^{1/2}(\hat{m}_n - m_X)_{n\to\infty} \to^L N(0, \frac{1}{4f^2(m_X)})$$

Géneralisation de cette loi aux $\alpha-$quantiles

$$n^{1/2}(X_{[n\alpha],n} - q(\alpha))_{n\to\infty} \to^L N(0, \frac{\alpha(1-\alpha)}{f^2(q(\alpha))})$$

où $X_{[n\alpha],n}$désigne l'estimateur empirique de $q(\alpha) = F^{inv}(\alpha)$.

En effet: pour $\alpha = \frac{1}{2}$ on a $q(\alpha) = m_X$

Corollaire1:

1)**Loi uniforme**

On a:

$$n^{1/2}(U_{[n/2],n} - \frac{a+b}{2})_{n\to\infty} \to^{L} N(0, \frac{(b-a)^2}{4})$$

En effet si $X \sim U\,[a,b]$ où $(a<b)$ on a:

$$q(\frac{1}{2}) = (a+b)/2$$

médiane théorique, et $f(q(\frac{1}{2}) = \frac{1}{b-a}$

2). **loi normal**

On a:

$$n^{1/2}(X_{[n\alpha],n} - m)_{n\to\infty} \to^{L} N(0, \frac{\sigma^2\pi}{2})$$

En effet si $X \sim N(m,\ \sigma^2)$ où $(m \in \mathbb{R},\ \sigma \in \mathbb{R}^+)$, on a:

$$q(\frac{1}{2}) = m$$

et $f(m) = \frac{1}{\sigma\sqrt{2\pi}}$

Erreur quadratique moyenne et convergence en moyenne quadratique (m. q.):

Notons $r_n = [n/2] + 1$.

Falk (1983) a démontré le résultat suivant :

Théorème 2 :

On suppose que : $\forall\delta > 0,\ \lim_{t\to\infty} t^{\delta}(1 - F(t) + F(-t)) = 0$, que F^{-1} existe, est deux fois différentiables au voisinage de 1/2 et que la dérivé seconde est bornée, alors : La moyenne et la variance de $\hat{m}_n$ existent et on a :

$$E\left[(\hat{m}_n - m_X)^2\right] = (n+2)^{-1}\frac{r_n}{n+1}(1-\frac{r_n}{n+1})(F^{-1})'^2(\frac{r_n}{n+1})+0(\frac{r_n}{n}-1/2)+0(n^{-3/2})$$

d'où le :

Corollaire 2 : Sous les conditions du théorème 2, on a:

$$\hat{m}_n \longrightarrow^{m.q.} m_X$$

Médiane non unique :

On suppose qu'il existe a et b tels que $a < b$ et, $\quad a = \inf\{x/F(x) \geq 1/2\}$
$b = \sup\{x/F(x) \leq 1/2\}$

Une présentation de la loi du logarithme itérée :

Notons: $\sigma_n^2 = \sigma^2(X_n) = E(X_n^2)$

$S_n = \sum_{k=1}^n X_k,\ \ E(S_n) = 0$

$s_n^2 = \sigma^2(S_n) = \sum_{k=1}^n \sigma_k^2$

un énoncé de la loi du logarithme itérée est le suivant :

si $s_n^2 \longrightarrow \infty,\ n \to \infty,\ \frac{|X_n|}{S_n} = 0(\log\log s_n^2)^{-\frac{1}{2}}$ et $t_n = (2\log\log s_n^2)^{\frac{1}{2}}$

alors: $P(\lim_{n\to\infty}\sup \frac{S_n}{s_n t_n} = 1) = 1$ et si on remplace X par $-X$ on a :

$$P(\lim_{n\to\infty}\inf \frac{S_n}{s_n t_n} = -1) = 1$$

ou plus simplement: Sous les hypothèses:

Z_i v. a. r, i. i. d avec $E(Z_i) = \mu$

et $Var(Z_i) = \sigma^2 < \infty,\ S_n = \sum_{i=1}^n Z_i$

On a:

$$\lim\sup_{n\to\infty} \pm\frac{\{S_n - n\mu\}}{\sigma\sqrt{2n\log\log n}} = 1\ \ p.s$$

qui est équivalent à:

$$\begin{cases} P(\lim\sup_{n\to\infty} \frac{\{S_n-n\mu\}}{\sigma\sqrt{2n\log\log n}} = 1) = 1 & (1) \\ P(\lim\inf_{n\to\infty} \frac{\{S_n-n\mu\}}{\sigma\sqrt{2n\log\log n}} = -1) = 1 & (2) \end{cases},$$ et en remplaçant dans (1) Xpar $-X$, on obtient (2)

1: Exemple classique: Application de la loi du logarithme itéré au processus empirique uniforme $\alpha_n(x)$

But : Déduction du comportement symptotique:

Soit X_iv. a. r i. i. d de loi F

$$\alpha_n(x) = \sqrt{n}\,\{F_n(x) - F(x)\}$$

$Z_i = 1_{\{X_i \leq x\}}; S_n = \sum_{i=1}^n Z_i;\ F(x) = P(X_i \leq x)$

$F_n(x) = \frac{1}{n}\sum_{i=1}^n Z_i = \frac{1}{n}S_n$

$Z_i = z$	0	1
$P(Z_i = z)$	$1 - F(x)$	$F(x)$

on a: $\mu = E(Z_i) = F(x)$, $\sigma^2 = VarZ_i = F(x)(1 - F(x))$ et donc:

$$\lim\sup_{n\to\infty} \pm\frac{\{S_n-n\mu\}}{\sigma\sqrt{2n\log\log n}} = \lim\sup_{n\to\infty} \pm\frac{\{nF_n(x)-nF(x)\}}{\sqrt{F(x)(1-F(x))2n\log\log n}}$$

$$= \lim\sup \pm\frac{\sqrt{n}(F_n(x)-F(x))}{\sqrt{2F(x)(1-F(x))\log\log n}}$$

en posant $\alpha_n(x) = \sqrt{n}\,\{F_n(x) - F(x)\}$processus empirique de$F(x)$, la loi du logarithme itéré devient:

$\lim\sup \pm\frac{\alpha_n(x)}{\sqrt{2F(x)(1-F(x))\log\log n}} = 1$, pour un échantillon de loi uniforme $F(x) = x$ on obtient:

$$\limsup \pm \frac{\alpha_n(x)}{\sqrt{2\log\log n}} = \sqrt{x(1-x)}$$

Exemple: Application du logarithme itéré à l'estimateur empirique de la médane $\hat{m}_n$

But: Etude de l'effet de l'oscillation de $\hat{m}_n$sur $[a,b]$

où $a = \inf\{x : F(x) \geq 1/2\}$et $b = \sup\{x : F(x) \leq 1/2\}$

Effet d'oscillation de la médiane :

Proposition 1:

la suite d'estimateurs $\{\hat{m}_n\}$ obéit a un effet d'oscillation par rapport à l'intervalle $[a,b]$, c-a-d:

$$P(\hat{m}_n \leq a \quad i.s.) = P(\hat{m}_n \geq b \quad i.s.) = 1 \qquad\qquad \forall F(f.r.)$$

($i.\ s.$ = infiniment souvent, noté $i.\ o$).

Démonstration: On va montrer que: $P(\hat{m}_n \geq b,\ i.\ s.\ ) = 1$ indépendament de Fen appliquant la loi du log itéré avec: $Zi = 1_{\{X_i \geq b\}} - 1_{\{X_i \leq a\}}$

les v. a. r Z_i sont indépendantes de loi :

$Z_i = z$	-1	$+1$
$P(Z_i = z)$	$\frac{1}{2}$	$\frac{1}{2}$

Montrons que: $\{S_n = \sum_{i=1}^{n} Z_i \geq 0\} \subset \{\hat{m}_n \geq b\}$

Rappelons que $\hat{m}_n = X_{\left[\frac{n}{2}\right]+1}$(la kième statistique d'ordre avec $k = \left[\frac{n}{2}\right] + 1$)

$$\{\hat{m}_n \geq b\} = \left\{\text{au moins } n - \left[\tfrac{n}{2}\right] \text{ des } X_i, i = 1, ..., n, \text{ sont } \geq b\right\}$$

c-à-d: $\{\hat{m}_n \geq b\} = \left\{\sum_{i=1}^{n} Z_i \geq n - 2\left[\frac{n}{2}\right] + 1\right\} \supset \{\sum_{i=1}^{n} Z_i \geq 0\}$

il es claire que $\mu = E(Z_i) = 0$, et $\sigma^2 = Var Z_i = 1$

En utilisant la loi du log itéré, on obtient: $P(\limsup_{n\to\infty} \frac{\sum_{i=1}^{n} Z_i}{\sqrt{2n\log\log n}} = 1) = 1$ (1)

Ainsi $\forall\varepsilon > 0$, $P(\left\{\frac{\sum_{i=1}^{n} Z_i}{\sqrt{2n\log\log n}} > (1-\varepsilon) = i.s\right\}) = 1$

ou $P(\sum_{i=1}^{n} Z_i \geq 0\ p.\ s) = 1$, ceci prouve que:

$$P(\hat{m}_n \geq b \quad i.s) = 1$$

De la mêm façon, en changeant Z_i par $-Z_i$ on prouve que:

$\{\sum_{i=1}^{n} Z_i \leq -1\} \subset \{\hat{m}_n \leq a\}$

en utilisant la loi du log itéré on obtient : $P(\liminf_{n\to\infty} \frac{\sum_{i=1}^{n} Z_i}{\sqrt{2n\log\log n}} = -1) = 1$ (2)

ce qui prouve que :

$$P(\hat{m}_n \leq a \quad i.s.) = 1$$

ce qui acheve la démonstration■

Remarque : L'effet d'oscillation de l'estimateur de la médiane ne permet pas sa convergence vers une constante,

l'existence d'estimateurs convergents de a et b et un problème auquel nous répondrons dans les paragraphes suivants .

Existence d'estimateurs de a et étude de ses modes de convergence :

On remarque que $\hat{m}_n = X_{([n/2]+1)}$ n'est pas un estimateur consistant de a . Cependant on va montrer que sous certaines hypothèses, il existe $\hat{m}_{L(n)} = X_{(L(n))}$, $L(n) < [n/2] + 1$, un estimateur qui converge vers a.

les résultas suivants sont des cas particuliers des théorèmes de Feldman et Tuker (1966) .

Proposition 2 :

Si pour $K > 0$, $\delta > 0$, la suite $\{L(n)\}$ satisfait:

(i) $0 < [n/2] + 1 - L(n) \leq Kn^{\frac{1}{2}+\varepsilon}$ avec $0 < \varepsilon < 1/2$

(ii) $[n/2] + 1 - L(n) \geq (1 + \delta)(2n \log \log n/2)^{\frac{1}{2}}$

Alors:

$$\hat{m}_{L(n)} \longrightarrow a \quad p.\ s.$$

Proposition 3:

Une condition nécessaire et suffisante pour que $\hat{m}_{L(n)}$ converge en probabilité vers a est la suivante:

$$C: \begin{cases} \text{(i)} & \lim_{n\to\infty} L(n)/n = 1/2 \\ \text{(ii)} & \lim_{n\to\infty} \sqrt{n}(1/2 - L(n)/n) = +\infty \end{cases}$$

Estimation de la médiane d'une loi symétrique :

Soit X une $v.\ a.\ r.$symétrique (La loi de X est identique à la loi de $-X$), F sa $f.\ r$, on appelle médiane centrale de X la médiane m_X^s vérifiant :

$$m_X^s = \frac{a+b}{2}$$

Dans ce cas $F^*F(2x)$ admet une médiane unique, comme on va le voir dans le résultat suivant :

Proposition 4 :

si F est symétrique alors $F^*F(2x)$ a une médiane unique: $m = m_X^s$.

Proposition 5 :

sous les hypothèses de la proposition 4, on a : $\hat{m}_n^s \longrightarrow m_X^s,\ p.\ s.$

Remarque :

Ce résultat n'est valable que si F est symétrique pour deux raisons :

(1) si F n'est pas symétrique, $F * F(2x)$ peut avoir une médiane en dehors de l'intervalle des médianes de F.

(2) au contraire du cas symétrique, $F * F(2x)$ peut avoir un intervalle non dégénéré de médianes .

Ceci est confirmé par le contre exemple suivant dû à Feldman et Tukey (1966) :

On considère $0 < a < b$, soit $\varepsilon > 0$ tel que $0 < \varepsilon < a$, soit $c > 2b + \varepsilon$ et $0 < \delta < \frac{1}{8}$.

soit f la densité de F définie par:

$$f(x) = \begin{cases} \frac{1}{2\varepsilon} & \text{si} \quad x \in (a - \varepsilon,\ a) \\ \frac{\delta}{(c-b)} & \text{si} \quad x \in (b,\ c) \\ (\frac{1}{2} - \delta)/\varepsilon & \text{si} \quad x \in (c,\ c + \varepsilon) \\ 0 & \text{ailleurs} \end{cases} .$$

Soient X et Y deux $v.\ a.\ \ i.\ i.\ d.$ de densité f (ci dessus), les médianes de X sont tous les points de l'intervalle $[a, b]$ et, on montrera que la plus petite médiane de $(X + Y)/2$ est plus grande que b en prouvant que :

$P\left[(X + Y)/2 > b\right] > 1/2$.

Comme $c > 2b$, on a juste à montrer que: $P(X + Y > c) \geq 1/2$ ou $P(X + Y \leq c) \leq 1/2$

$P(X + Y \leq c) < \frac{1}{4} + 2(\frac{\delta}{2}) + \delta^2 = (\frac{1}{2} + \delta)^2$.

et comme $\delta < 1/2$, on a : $P(X + Y \leq c) < (\frac{\delta}{8})^2 < 1/2$.■

2- Estimateur alternatif: noté $(\hat{m}_{k+1/2,n})$

Nous présentons dans ce qui suit cet estimateur sans détailler ses propriétés .

Cet estimateur a été explicité dans l' article de Kaigh (1983) .

Définition :

Soit $(X_1,\ \ldots\ ,\ X_n)$ un n-échantillon de $v.\ a.\ r.\ \ i.\ i.\ d.$ et k un entier fixé .

soit $(Y_{(1:k)},\ \ldots\ ,\ Y_{(k:k)})$ une sous statistique d'ordre de k $v.\ a.$ prise sans remise parmi les n $v.\ a.\ r.\ X_i$.

Un estimateur de la médiane des Y_i est déterminé par $Y_{[k/2]:k}$.

dans le n-échantillon défini précédemment il y a C_n^k k-sous échantillons, et l'estimateur $\hat{m}_{(k+1)/2,n}$ est défini comme suit:

$$\hat{m}(_{k+1)/2,n} = \sum_{j=1}^{r+n-k} P(Y_{(r:k)} = x_{(j:n)}) X_{(j:n)}$$

$r = [(k + 1)/2]$, et

$$P(Y_{(r:k)} = x_{(j:n)}) = \tbinom{j-1}{r-1}\tbinom{n-j}{k-r}/\tbinom{n}{k}$$

$r \leq j \leq r+n-k$. où $\binom{n}{k} = C_n^k$

Estimation de la médiane par la méthode du noyau et la méthode des L-statistique :

3- Méthode du noyau:

a. Définition d'un noyau de probabilité :

Un noyau K est une application de $\mathbb{R}$ vers $\mathbb{R}$ bornée, intégrable par rapport à la mesure de Lebesgue et d'intégrale 1.On dit qu'un noyau est de Parzen -Rosemblatt si: $\lim_{|X|\to\infty} |X|^s K(x) = 0$

b.**Estimation de la médiane :**

Dans tout ce paragraphe on suppose que la médiane de X, m_X est unique .

Un estimateur de la fonction de répartition F par la méthode du noyau K a la forme suivante :

$$F_n^K(x) = \int_{-\infty}^{x} \frac{1}{nh(n)} K(\frac{Xi-t}{h(n)})dt, \text{ où } h(n) \to 0 \text{ avec } \frac{1}{n}$$

où $(X_i)_{\leq 1 i \leq n}$est un $n-$échantillon issu de X.

Il en résulte l'estimateur suivant de la médiane :

$$m_n^K = \inf \left\{ x : F_n^K(x) \geq 1/2 \right\}$$

Reiss (1980) a défini un estimateur basé sur la "moyenne " des termes de la statistique d'ordre $(X_1, ..., X_n)$.

Cet estimateur est repris par Falk (1984 (a)) et est présenté sous la forme suivante :

$$m_n^K = \int_0^1 F_n^{-1}(x) \frac{1}{h(n)} K(\frac{1/2-x}{h(n)})dx \quad (*)$$

m_n^K est un estimateur de la médiane par la méthode du noyau .

Etude de la convergence de m_n^K:

Soient les conditions suivantes :

(1) X a une densité $f(x)$ continue et positive au voisinage de m_X.

(2) $f'(x)$ existe et est continue au voisinage de m_X.

(3) Le noyau $K(x)$ est de support fini .

(4) $K(x)$ est bornée .

(5) $\lim_{x\to\infty} x^a [F(-x) + 1 - F(x)] = 0$, pour $a > 0$.

(6) on suppose que F^{-1}est $(m+1)$ différentiable au voisinage de 1/2 , avec une dérivée seconde continue si $m = 1$, et de dérivée $(m+1)$ème bornée si $m \geq 2$.

(7) $\int K(x)dx = 1$, $\int xK(x)dx = 0$.

(8) $\lim_{n\to\infty} n^{1/4}h(n) = 0$

$$m_n^K = (*) = \sum_{i=1}^{n} X_{(i:n)} \frac{1}{h(n)} \int_{(i-1)/n}^{i/n} K(\frac{t-1/2}{h(n)})dt$$

Evaluation de l'erreur quadratique,

et étude de la convergence en moyenne quadratique (m.q) de m_n^K

Il s'agit d'étudier : $E(\left[m_n^K - m_X\right]^2) = M.S.E.(m_n^K)$, et de conclure pour n tendant vers l'infini :

Théorème 3:

On suppose que les conditions (3) , (5) , (6) , et (7) sont vérifiées , alors :

pour n assez grand la $M.S.E.(m_n^K)$ existe et :

$$n.M.S.E.(m_n^K) = (F^{-1})'^2(1/2)\left\{\frac{1}{4} - 2h(n)\int xK(x)k(x)dx\right\} + o(nh(n)^{2m+2}) + o(n^{-1/4}) + o(h(n))$$

où $k(x) = \int_{-a}^{x} K(y)dy$.

Corollaire 3 :

Sous les conditions du théorème 3 , on a

$$m_n^K \to^{m.q} m_X$$

Pour n tendant vers l'infini, la $M.S.E.(m_n^K)$ tend vers 0, ce qui prouve le corrollaire3.

Etude de la loi limite :

Yang (1985) a démontré le théorème suivant :

Théorème 4 :

On suppose que les conditions (1) , (2) , (3) , (7) , et (8) sont vérifiées , alors :

$$n^{1/2}\left[m_n^K - m_X\right] = -n^{1/2}\left[F_n(m_X) - 1/2\right]/f(m_X) + o_p(1)$$

où $o_p(1)$ converge vers 0 en probabilité quand n tend vers ∞.

ce théorème permet d'avoir le résultat suivant :

Corollaire 4 :

Sous les conditions du théorème 4, on a :

$$n^{1/2}(m_n^K - m_X) \to^{L.} N(0, \frac{1}{4}f^2(m_X))$$

Remarque :

On obtient la même loi limite que l'estimateur empirique $\hat{m}_n$(voir **1 -Théorème1**)

Etude de la vitesse de convergence de la variance et du biais de m_n^K :

Ce résultat est du à Yang (1985) .

Théorème 5 :

On suppose que les conditions (1) , (2) , (3) ,(4) , (7) et (8) sont vérifiées , alors :

$$\begin{aligned} Var(m_n^K) &= o(1/nh(n)^2) \qquad \text{et} \\ \left|E(m_n^K) - m_X\right| &= o(n^{-1/2}) + 0(h(n)^2) \end{aligned}$$

4- Méthode des L-statistiques :

Définitions des L-estimateurs :

Un L-estimateur est une fonction de la statistique d'ordre, de la forme:

$$T_n = \sum_{i=1}^{n} a_{n_i} H(X_{(i)})$$

H étant une application quelconque de $\mathbb{R}$ dans $\mathbb{R}$, les a_n sont des constantes .

Estimation de la médiane par la méthode des L-statistiques:

En s'inspirant de la méthode du noyau , on définit cet estimateur, qu'on note $\hat{m}_n^L$, de la façon suivante :

$$\hat{m}_n^L = \frac{1}{n} \sum_{i=1}^{n} X_{(i:n)} \frac{1}{h(n)} K \left[\frac{(i/n) - (1/2)}{h(n)} \right]$$

K est un noyau et $h(n)$ une suite tendant vers 0.

Remarque:

L'estimateur m_n^K classé dans la méthode du noyau est aussi un L-estimateur .

du fait de la relation étroite entre m_n^K et $\hat{m}_n^L$, on a des résultats semblables concernant les convergences :

Loi limite , variance et biais :

Corollaire 5 :

En supposant que les conditions (1) , (2) , (3) et (7) ainsi que la condition:

$$(A) : \lim_{n\to\infty} 1/(n^{5/8}h(n)) = 0$$

sont vérifiées ,alors les résultats des théorèmes 4 et 5 s'appliquent à l'estimateur $\hat{m}_n^L$.

Equivalence asymptotique entre m_n^K et $\hat{m}_n^L$:

Du fait de leur définitions ,ces estimateurs sont assez liés .

Théorème 6:

On suppose vérifer les conditions (3) et (5) et que le noyau K est Lipchitzien alors :

$$E((\hat{m}_n^L - m_n^K)^2) = o(1/(nh(n)^2 a_n^2))$$

avec $a_n \to \infty$ et $a_n/(nh(n)) \to 0, n \to \infty$.

(On dira que les estimateurs $\hat{m}_n^L$ et m_n^K sont asymptotiquement équivalents en moyenne quadratique).

Remarque:

1- La démonstration du théorème 6 est dû à Yang (1985).

2- Quand $h(n)$ est grand , la variance de m_n^K est petite alors que le biais est grand , si $h(n)$ est petite on a la phénomène inverse .

3- En raison de l'équivanlence asymptotique de m_n^K et $\hat{m}_n^L$, on peut conclure aux mêmes phénomènes pour $\hat{m}_n^L$.

4- On a défini précédemment la médiane d'une $v.a.X$ comme étant une constante m telle que: $|X - m|_{L_1} = \inf |X - c|_{L_1}, c \in \mathbb{R}$.

Soit $(X_1, ..., X_n)$ un n-échantillon de $v.a.r$ $i.i.d.$ de même loi que X , un estimateur de la médiane de X est la valeur θ qui minimise la somme suivante : $\sum_{i=1}^{n} |X_i - \theta|$.

si: n est impair $n = 2r + 1$, cet estimateur est donné par $X_{(r+1:n)}$.

si: n est pair $n = 2r$, on peut prendre pour θ toute valeur de l'intervalle $\left[X_{(r:n)}, X_{(r+1:n)}\right]$.

3.1.3 Comparaison des différents estimateurs de la médiane :

Nous avons choisi comme critère de comparaison des estimateurs$T_n$et T_n' de la médiane m_X le rapport de leurs erreurs quadratiques, noté (RE.Q) et défini ainsi:

$$R.E.Q. = \frac{E(T_n - m_X)^2}{E(T_n' - m_X)^2}$$

a .Comparaison des estimateurs empirique $\hat{m}_n$ et alternatif $\hat{m}_{(k+1)/2,n}$

Les données du tableau suivant proviennent de Kaigh (1983) .

$$R.E.Q. = \frac{E(\hat{m}_n - m_X)^2}{E(\hat{m}_{k+1/2,n} - m_X)}$$

Tableau 1:		¨R.E.Q		
Lois	uniforme	normale	Double exponentielle	Cauchy
$n = 99, k = 39$	$1,16$	$1,12$	$1,05$	$1,08$
$n = 99, k = 9$	$1,48$	$1,31$	$0,96$	$0,94$
$n = 99, k = 19$	$1,28$	$1,22$	$1,04$	$0,96$

Constatation :

Pour ces lois connues , les simulations montrent que dans 2/3 des cas , l'estimateur alternatif de la médiane est plus performant que l'estimateur empirique .

cette performance varie suivant k dont il serait intéressant de connaitre la valeur optimale, un début de solution est proposé par Lecoutre et Tassi (1987) .

b. Comparaison des estimateurs empirique et L-estimateur :

Le noyau K utilisé pour la définition de $\hat{m}_n^L$ est la suivant :

$$K(u) = (1 - |u|)I_{(|u| \leq 1)}$$

Ce même noyau est aussi utilisé pour la définition de m_n^Kles données figurants dans le tableau qui suit sont obtenues par simulation et sont dûs à Yang (1985) .

$$A = \frac{E(\hat{m}_n - m_X)^2}{E(\hat{m}_n^L - m_X)^2}, \qquad B = \frac{E(\hat{m}_n - m_X)^2}{E(m_n^K - m_X)^2}$$

Tableau 2:

$h(n)$	$0,06$	$0,08$	$0,10$	$0,12$	$0,18$	$0,20$	$0,22$	$0,24$	$0,26$

Normale

$A =$

$n = 30$	$0,94$	$0,94$	$1,06$	$1,05$	$1,11$	$1,14$	$1,14$	$1,17$	$1,18$
$n = 50$	$1,06$	$1,08$	$1,10$	$1,12$	$1,17$	$1,19$	$1,21$	1.22	$1,24$

Double exponentielle

$A =$

$n = 30$	$0,89$	$0,91$	$0,99$	$0,96$	$0,97$	$0,99$	$0,97$	$0,97$	$0,96$
$n = 50$	$1,00$	$1,00$	$1,01$	$1,01$	$1,00$	$0,99$	$0,98$	$0,97$	$0,96$

Exponentielle

$A =$

$n = 30$	$1,13$	$1,18$	$1,28$	$1,27$	$1,35$	$1,39$	$1,39$	$1,42$	$1,43$
$n = 50$	$1,19$	$1,21$	$1,23$	$1,25$	$1,30$	$1,32$	$1,33$	$1,34$	$1,35$

Constatation:

Pour les lois exponentielle et normale , l'estimateur $\hat{m}_n^L$ de la médiane est plus performant que l'estimateur $\hat{m}\ _n$.

cette performance augmente quand $h(n)$ croit .

pour la simulation double exponentielle , l'estimateur empirique est plus performant que $\hat{m}_n^L$.

d'une manière générale , m_n^L est plus performant que $\hat{m}_n$.

Tableau 3 : $n = 100$

$h(n)$	0,05	0,09	0,11	0,13	0,15	0,17	0,19	0,21	0,25

Normale

$A =$	1,03	1,07	1,08	1,10	1,12	1,14	1,15	1,17	1,20
$B =$	1,04	1,07	1,09	1,010	1,12	1,14	1,16	1,17	1,20

Double exponentielle

$A =$	1,03	1,04	1,04	1,04	1,04	1,03	1,03	1,02	1,00
$B =$	1,04	1,05	1,06	1,06	1,06	1,05	1,04	1,04	1,02

Exponentielle

$A =$	1,05	1,09	1,11	1,13	1,15	1,16	1,17	1,18	1,18
$B =$	1,03	1,07	1,08	1,10	1,11	1,11	1,12	1,12	1,10

Interprétation

Ce tableau confirme les constatations du tableau 1 concernant l'estimateur $\hat{m}_n^L$, il nous permet de tirer les mêmes conclusions pour l'estimateur m_n^K. Plus n est grand , plus la performance des estimateurs m_n^K et $\hat{m}_n^L$ est meilleure.

c. Comparaison des estimateurs m_n^K et $\hat{m}_n^L$:

$$C = \frac{E(m_n^K - m_X)^2}{E(\hat{m}_n^L - m_X)^2}$$

Tableau 4:

$h(n)$	0,05	0,09	0,11	0,13	0,15	0,17	0,19	0,21	0,25

Normale

$C =$	0,99	1,00	0,99	1,00	1,00	1,00	0,99	1,00	1,00

Double exponentielle

$C =$	0,99	0,99	0,98	0,98	0,98	0,98	0,98	0,98	0,98

Exponentielle

$C =$	1,01	1,01	1,02	1,02	1,03	1,04	1,04	1,05	1,07

¨Ce tableau illustre le théorème 6 concernant la relation entre les estimateurs m_n^K et $\hat{m}_n^L$

d .Comparaison des estimateurs alternatif $\hat{m}_{k+1/2,n}$ et $\hat{m}_n^L$

:

$$D = \frac{E(\hat{m}_{k+1/2,n} - m_X)^2}{E(m_n^L - m_X)^2}$$

Les données de ce tableau figurent dans l'article de Yang (1985)

Tableau 5:

$h(n)$	$0,05$	$0,09$	$0,11$	$0,13$	$0,15$	$0,17$	$0,21$	$0,23$	$0,25$

Normale

$k = 39$	$0,92$	$0,94$	$0,96$	$0,98$	$1,00$	$1,00$	$1,04$	$1,06$	$1,07$
$k = 79$	$0,99$	$1,01$	$1,03$	$1,05$	$1,07$	$1,08$	$1,12$	$1,14$	$1,15$

Double exponentielle

$k = 39$	$0,98$	$0,99$	$0,99$	$0,99$	$0,99$	$0,98$	$0,97$	$0,96$	$0,95$
$k = 79$	$0,98$	$0,99$	$0,99$	$0,99$	$0,99$	$0,98$	$0,97$	$0,96$	$0,95$

Exponentielle

$k = 39$	$0,95$	$0,99$	$1,00$	$1,01$	$1,03$	$1,04$	$1,06$	$1,06$	$1,06$
$k = 79$	$1,00$	$1,04$	$1,05$	$1,07$	$1,09$	$1,10$	$1,12$	$1,12$	$1,12$

Interprétation :

Suivant les distributions des résultats simulés , l'un ou l'autre des deux estimateurs m_n^K ou $\hat{m}_{k+1/2,n}$ est meilleur . cependant , il apparait que généralement m_n^K est plus performant . Cette performance varie avec $h(n)$.

Yang (1985) a proposé un choix optimale de $h(n)$ en minimisant l'erreur quadratique moyenne déterminée par la méthode des Bootstrap ou la méthode du Jacknife .

⧫ II. Variables corrélees : Cas d'un processus.

Soit $(X_i)_{i=1,n}$ une suite d'obsrevations corrélées de la $v.a.r.$ X, m_X désigne encore la médiane de X supposée unique et $\hat{m}_n$ représente son estimateur empirique défini comme dans le cas d'observations non corrélées .

plus précisément , nous supposons que les observations $(X_i)_{i=1,n}$ sont α-mélangentes , c'est à dire :

Définition de variables α-mélangeantes :

Une suite de $v.a.$ est dite α-mélangeante (ou fortement mélangeante) s'il existe une suite $\alpha(n)$ vérifiant :

$1 \geq \alpha(1) \geq \alpha(2) \geq \alpha(n)$ avec $\alpha(n) \to 0$ avec $\frac{1}{n}$

et

$$\sup_{k \geq 1} \sup_{B \in M_1^k} \sup_{A \in M_{k+n}^{\infty}} |P(A \cap B) - P(A)P(B)| \leq \alpha(n)$$

où: M_1^k représente la tribu engendrée par $(X_1, ..., X_k)$

M^{α}_{k+n} représente la tribu engendrée par $(X_{k+n}, X_{k+n+1}, ...)$

Carbon (1983) a établi le résultat suivant :

Lemme 1:

Soit $(X_i)_{i\in\mathbb{N}}$ une suite de *v.a.* α-mélangeantes définies sur (Ω, A, P) et vérifiant :

$\forall i \in \mathbb{N}, E(X_i) = 0, |X_1| \leq d$ et $E(X_i^2) \leq D$.

On pose :

$\forall k \in \mathbb{N}^*$, $\xi(k) = \sum_{i=1}^{k} \alpha(i)$ (La somme des coefficients de mélangence)

$\forall \varepsilon > 0$ et $\forall n \in \mathbb{N}^*$ tel que $n \geq 3$ on a :

$$P(\left|\sum_{i=1}^{n} X_i\right| > \varepsilon) \leq 2\exp\left[-\alpha\varepsilon + 4\alpha^2 e(D + 8d^2\xi(k))n + 2\sqrt{e}(\alpha(k))^{2k/3n}.\frac{n}{k}\right]$$

α vérifiant : $0 < \alpha \leq \frac{1}{4kde}$, et $k \in \mathbb{N}^*$ tel que $k < n/2$.

les nombres α et k ainsi que D et d peuvent dépendre de n.

Convergences de m_n vers m_X :

Soit $k(n)$ une suite entière coissante vérifiant avec α :

(H) $\lim_{n\to\infty} \alpha(k(n))^{\frac{2k(n)}{3n}} = 0$ et $\lim_{n\to\infty} k(n) = \infty$.

(H)' $k(n)/\log n \to \infty, n \to \infty$.

Lemme 2 :

Sous les conditions du lemme 1 et sous l'hypothèse (H) , on a pour n assez grand ,

$$\sup_{x\in\mathbb{R}} P(|F_n(x) - F(x)| > \varepsilon) \leq 2\exp(-\frac{n\varepsilon}{32ek(n)})$$

Démonstration :

Soit $x \in \mathbb{R}$, on pose $\Delta_i = 1_{\{X_i \leq x\}} - F(x)$,

$|\Delta_i| \leq 2 = d$ et $E(\Delta_i^2) \leq 2 = D$.

$F_n(x) - F(x) = \frac{1}{n}\sum_{i=1}^{n} \Delta_i$, donc :

$P(\frac{1}{n}|\sum_{i=1}^{n} \Delta_i| > \varepsilon) = P(|F_n(x) - F(x)| > \varepsilon) = (*)$

En appliquant le lemme 1 et en choisissant $\alpha = \frac{1}{4kde}$, on obtient pour tout $\varepsilon > 0$,

$(*) \leq 2\exp(-\frac{n}{k}\left[\frac{\varepsilon}{8e} - \frac{1}{16e}(\frac{2}{k} + 32\frac{\xi(k)}{k}) - 2\sqrt{e}\alpha(k)^{2k/3n}\right]$

d'aprés (H) , $\frac{2}{k}$, $\frac{\xi(k)}{k}$ et $\alpha(k)^{2k/3n} \to 0$, quand $n \to \infty$.

Pour n assez grand , on obtient :

$\frac{1}{16e}(\frac{2}{k} + 32\frac{\xi(k)}{k}) + 2\sqrt{e}\alpha(k)^{2k/3n}) \leq \frac{3\varepsilon}{32e}$, donc :

$P(|F_n(x) - F(x)| > \varepsilon) \leq 2\exp(-\frac{n\varepsilon}{32ke})$

Il suffit de remarquer que le majorant ne dépend pas de x pour finir la démonstration ■.

Ces deux lemmes permettent de démontrer le résultat suivant :

Théorème 7:

Sous les hypothèses (H) et (H)' , on a :

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to^{p.co.} 0$$

Démonstration :

Il suffit d'utiliser le lemme 2 et la démonstration classique du théorème de Glevenko-Cantelli.

Théorème 8:

Sous les hypothèses (H) et (H)' ,on a :

$$\hat{m}_n \to^{p.co.}_{n\to\infty} m_X$$

Démonstration :

D'aprés le théorème 1 , si $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to^{p.co.} 0$ et si la médiane est unique alors:

$$m_n \to^{p.co.} m_X$$

Il suffit d'appliquer le théorème 7 pour terminer la démonstration .

Application :

Corollaire:

Si $(X_i)_{i=1,n}$ est processus réel géométriquement mélangeant $(\alpha(k) = a\rho^k, a \geq 0, 0 \leq \rho < 1, k \in \mathbb{N}^*)$

Alors ,

$$\hat{m}_n \to^{p.co.} m_X$$

Démonstration :

Il suffit de vérifier les hypothèses (H) et (H)' et d'appliquer le théorème 8.

$$\alpha(k) = a \exp k \log \rho,$$

$$\alpha(k)^{2k/3n} = a^{2k/3n} \exp \frac{2k^2}{3n} \log \rho = (^*)$$

On pose: $$k(n) = \beta_n n^{(1+\beta)/2},\ \ 1 < \beta_n < 2,\ 0 < \beta < 2.$$

$$(^*) = \exp \frac{2}{3}\beta_n n^{(\beta-1)/2} \log a. \exp \log \rho \frac{2}{3}\beta_n^2 n^\beta \to 0, n \to \infty.$$

(H)' est immédiate.■

3.2 Cas multivarié:

3.2.1 Définitions et Propriété

Définitions de la Médiane d'un Vecteur Aléatoire:

Définition 1: Mood(1941)

Soit: $X = (X_1, \ . \ . \ . \ , \ X_s), \ s > 1$ un vecteur aléatoire, Mood(1941) a défini une médiane m_X par le vecteur:

$$(m_{X_1}...m_{Xs}), \text{où} \ \ m_{X_i}, i = 1, ..., s$$

est une médiane de la $v.\ a.\ r\ X_i$, Median(s): X_i

Remarque1: Cette définition de la médiane multivariée présente quelques faiblesses : Si on remplace $X_i, \ (i = 1, \ s)$, par une fonction monotone de X_i la médiane est invariante mais elle ne le sera plus si le repére subit une rotation. Ceci appelle des définitions plus opérationnelles.

Autres définitions :

Définition 2: Médiane de Haldane (1948) Soit X un vecteur aleatoire intégrable, soit Φ la fonction de $\mathbb{R}^s$ dans $\mathbb{R}^+$ définie par

$$\Phi(\alpha) = \int_{\mathbb{R}^s} \|x - \alpha\| \, dF_X(x)$$

Une médiane de X est le vecteur M_x qui minimise Φ. La médiane suivant Mood est appelée "**Médiane arithmétique**" et celle de Haldane est appelée "**Médiane géomètrique"** de manière générale Kemperman(1987) a défini la médiane d'une mesure finie sur un Banach :

Définition 3: Kemperman: Soit λ une mesure finie sur un espace de Banach ($E, \ \|\cdot\|$). La médiane de λ est tout point $y \in E$ minimisant f où:

$$f(y) = \int_E (\|x - y\| - \|x\|)\lambda(dx) \qquad (*)$$

Remarque 2:

Si $\int_E \|x\| \, \lambda(dx) < \infty$ on retrouve la définiton de Haldane. Isogai (1985) a repris la définition de Haldane pour la généraliser de la façon suivante :

Définition 4: Isogai. Soit Ψ une fonction strictement croissante défine sur $[0, +\infty[$, ayant une dérivée second continue sur $(0, \ +\infty)$ et vérifiant :

$\Psi(0) \geq 0$

$\Psi(u_1 + u_2) \leq \Psi(u_1) + \Psi(u_2)$ pour $u_1, \ u_2 \geq 0 \quad (1)$

(1) implique que $\Psi(u) \leq \Psi(1).\ (u + 1)$. On définie la distance entre X et θ par :

$$\Psi(X, \theta) = \Psi(\|X - \theta\|)$$

où $\|X-\theta\|$ représente la norme défine par:

$$\{(X-\theta)^t\beta(X-\theta)\}^{1/2}\text{avec } \ \beta=(\beta_{i,j}); i,j=1,s$$

est la matrice de convariance du vecteur aléatoire X^t : $(X_1, \ . \ . \ . \ , \ X_s)$ transposé du vecteur X de moyenne $M(M=(m_1, \ . \ . \ . \ , \ m_s))$. Posons: $\Delta(\theta)=E(\Psi(X, \ \theta))$ et notons θ_0 la valeur qui minimise $\Delta(\theta)$:

$\theta_0$est appelé médiane généralisée de X associée à Ψ

Médiane d'un echantillon:

Définition 5:

Soit un n-echantillon multivarié d'oservations X_i, une médiane de cet échantillon est une des valeurs m_n qui minimisent la somme suivante :

$$\sum_{i=1}^{n} d(\alpha, X_{i)}$$

d etant une distance de $\mathbb{R}^s$ où $\alpha \in \mathbb{R}^s$. Si d est la distance euclidienne, la médiane de l'échantillon appartient au plud petit polygone convexe C contenant la répartition. C'est aussi l'intersection des demi-plans contenant les X_i.

Existence et unicite de la médiane multidimensionnelle :

L'existence de la médiane d'un vec. a. X de fonction de répartition F est assuré par le fait que sa loi P_X est tendue- Notion de tightness(Billingsley1968, p9).

Valadier(1984) a prouvé le résultat suivant:

Théorème 1: Valadier(1984): Soit E un espace de Banach séparable et réflexif et P une mesure de probabilité sur E. On suppose que P est d'ordre 1. Alors:

La fonction ϕ : $E \to \mathbb{R}$ définie par: $\phi(a) = \int_E \|a-x\| P(dx)$, est convexe, lipchitzienne de rapport 1, on a:

$$\lim_{n\to\infty} \frac{\phi(a)}{\|a\|} = 1$$

et elle atteint sa borne inférieure sur un convexe fermé borné non vide M. Pour $a \in M$, on a:

$$\|a\| \leq 2\int \|x\| P(dx)$$

Kemperman (1987) a démonté que f définie par:

$$f(y) = \int (\|x-y\| - \|x\|)\lambda(dx) \quad (*)$$

est continue et lipchitzienne.

Défintion 6: Soit E un espace de Banach, E est dit strictement convexe si :

$$\forall (X,Y) \in E^2 \text{ et } (\|X\| = \|Y\| = 1, X \neq Y), \text{on a : } \|(X+Y)/2\| < 1.$$

Remarque : Tout espace de dimension finie muni de la norme euclidienne est strictement convexe.

Théorème 2: (Kemperman (1987)): La fonction $f : E \rightarrow \mathbb{R}$, est convexe et l'ensemble des médianes $M = M(P)$ est fermé et convexe. Si en plus E est strictement convexe et si le support de P n'est pas une droite. Alors: **f est strictement convexe et P possède au plus une médiane.**

Démonstration : C'est le théorème de Kemperman (1987).

Dans toute la suite, $E = \mathbb{R}^s$ et est muni de la norme euclidienne (sauf indication contraire).

Corollaire 1: Si le support de P n'est pas une droite, la médiane de P est unique.

Démonstration : On présente deux démonstrations pour ce résultat; la deuxiéme qui est directe est due àMilasevic -Ducharme (1987(b)).

1. E est strictement convexe d'aprés la remarque a, l'existence de la médiane est assurée par le théorème1 et on utilise le théorème 2 pour conclure à l'unicité de la médiane. ■

2. Supposons que $\alpha_1,\ \alpha_2$ sont deux médianes de P, $\alpha_1 \neq \alpha_2$, L est la droite qui passe par α_1 et α_2. Soit $0 < \lambda < 1$ et $x \in E/L$, on a:

$$\|x - \lambda\alpha_1 - (1-\lambda)\alpha_2\| - \|x\| < \lambda(\|x - \alpha_1\| - \|x\|) + (1-\lambda)(\|x - \alpha_2\| - \|x\|)$$

comme le support de P n'appartient pas à L, alors :

$$f(\lambda\alpha_1 + (1-\lambda)\alpha_2) < \lambda f(\alpha_1) + (1-\lambda) f(\alpha_2) = \min_{\alpha \in E} f(\alpha)$$

Ceci est en contradiction avec le fait que α_1 et α_2 sont deux médianes distinctes de P. ■

Relation entre médiane et distance:

Proposition1: La mediane dépend de la distance utilisée En effet: Soit l'exemple suivant: On munit $\mathbb{R}^2$ de la norme L_∞:

$$\|x\| = \max(\left|x^1\right|, \left|x^2\right|)$$

où $x = (\left|x^1\right|,\ \left|x^2\right|)$. Soient P la mesure de support les points $(x_i)_{i=1,4}$de coordonnées $(1,\ 0)$, $(-1,\ 0)$, $(0,\ 1)$ et $(0,\ -1)$affectés uniformément des poids$(\omega_i)_{i=1,4} = \frac{1}{4}$et g la fonction définie par:

$$g_1(y) = g_1(y^1, y^2) = \sum_{i=1}^{4} \max(\left|y^1 - x_i^1\right|, \left|y^2 - x_i^2\right|$$

La médiane de cette mesure est tout point y minimisant g. On remarque que g est symétrique et convexe donc elle est minimale à son origine et $g_1(0) = 1$. $g_1(x_i) = 1(\forall i = 1,\ 4)$, donc le carré Q, de sommets $x_i(i = 1,\ 4)$, est inclus dans $M(P)$car $M(P)$ est convexe d'après le théorème 1 et donc:

$$M(P) = Q.$$

Changeons la métrique, soit $\mathbb{R}^2$ maintenant muni de la norme euclidienne, **il est strictement convexe**. Soit g la fonction définie par :

$$g_2(y) = g_2(y^1, y^2) = \sum_{i=1}^{4} \sqrt{(y^1 - x_i^1)^2 + (y^2 - x_i^2)^2},$$

$g_2$est convexe et symétrique. Sa valeur minimale est obtenue en son origine, **donc $y(0,\ 0)$ est l'unique médiane de P. Donc la médiane dépend de la métrique utilisée.■**

Chapter 4

Mediane conditionnelle: Cas univarié.

4.1 Résultats Théoriques:

4.1.1 Définitions

Définitions de la mediane conditionnelle: Soient $(\Omega,\ \mathcal{F},\ P)$ un espace probabilisé, Y une v. a. réelle, $\mathcal{G}$ une σ-algébre telle que $\mathcal{G}\ \subseteq\ \mathcal{F}$.

Définition 1: La médiane conditionnelle de Y sachant la $\sigma-$algébre $\mathcal{G}$, qu'on note m, est une v. a. vérifiant propriétes suivantes:

i) m est $\mathcal{G}-$mesurable

ii) $P(Y \geq m\ /\mathcal{G}\ )\ \geq\ 1/2\ \leq\ P(Y\ \leq m/\mathcal{G})\ \ p.\ \ s.$ Cette définition est à rapprocher de la définition de la médiane non conditionnelle.

Soient les données suivantes:

X un vec. a. de $\mathbb{R}^d$, Y une v. a. r. , $\mathrm{P}_{X,Y}$ la probabilité du couple $(X,\ Y) \in (\mathbb{R}^{d+1},\ \beta(\mathbb{R}^{d+1})$ où: $\beta(\mathbb{R}^{d+1})$ est la tribu borélienne de $\mathbb{R}^{d+1}$;c. à. d: la tribu engendrée par les pavés de $\mathbb{R}^{d+1}$. Soit P_Y^X une version régulière de la probabilité conditionnelle de Y par rapport à X et $F(.\ /x)$ une version de la fonction de répartition conditionnelle de Y sachant $X = x$ et notons :

$\mathcal{N} =\ \ \{y\ :\ F(y/x)\ \geq 1/2\}$ (resp. $\mathcal{N}'\ =\ \{y\ :\ F(y/x)\ \leq\ 1/2\}$

Définition 2: Une médiane conditionnelle de$Y$ sachant $X = x.\ m(x)$, est la borne inférieure de $\mathcal{N}$: (resp. $m'(x)$ est la borne supérieure de $\mathcal{N}'$)

$$m(x) = \inf \mathcal{N} \quad \text{et} \quad m'(x) = \sup \mathcal{N}'$$

En cas d'unicité de la médiane conditionnelle on a :

$$m(x) = m'(x)$$

Analogie avec l'Esperance conditionnelle:

On suppose que: $E(|Y|) < \infty$. Par analogie à l'espérance conditionnelle (Meilleure appoximation de Y par un fonction de X au sens de L^2). la Médiane conditionnelle de Y sachant $X = x$ est la meilleure approximation de Y par une fonction de X au sens de la norme L^1 soit:

Définition 3: $m(x)$ est telle que:

$$E(|Y - m(x)| \ / \ X = x) \ = \ \inf_{a \in \mathbb{R}} E(|Y - a| \ / \ X = x)$$

4.1.2 Géneralisation *aux* $\alpha-$quatiles conditionnels

La généralisation de la définition (1) aux α-quantiles conditionnels se fait comme suit:

Soit $\alpha \in]0, 1[$, le $\alpha-$quantile conditionnel de Y sachant la σ-algèbre $\mathcal{G}$, noté q_α, est une v. a. vérifiant :

i) q_α est $\mathcal{G}$-mesurable

ii) $P(Y \ \geq q_\alpha \ /\mathcal{G}) \ \geq \ 1 - \ \alpha, \ \ P(Y \ \leq \ q_\alpha \ /\mathcal{G}) \ \geq \alpha.$

La généralisation de la définition 2 se fait de la façon suivante :

Notons: $\mathcal{N}_\alpha = \{ \ y \ : \ F(y/x) \ \geq \alpha\}$ (resp $\mathcal{N}'_\alpha = \{ \ y : \ F(y/x) \ \leq \alpha\}$),

On appelle α-quantile conditionnel de Y sachant $X = x$, noté $q_\alpha \ (x)$, la borne inférieure de $\mathcal{N}_\alpha$. (resp $q'_\alpha \ (x)$ la borne supérieure de $\mathcal{N}'_\alpha$):

$$q_\alpha(x) = \inf \mathcal{N}_a \text{ et } \ q'_\alpha(x) = \sup \mathcal{N}'_\alpha$$

Quant à la définition 3 sa généralisation est comme suit : **Stone(1977)**

Pour $E(|Y| \ < \ \infty$, posons $\mathcal{H}(y, \ a) = c(\alpha \ (y - a)^+ + (1 - \alpha)(y - a)^-)$, $c > 0$.

Le α-quantile conditionnel $q_\alpha(x)$ est tel que:

$$E(\mathcal{H}(Y, q_\alpha \ (x) \ ) = \ \inf_{a \in \mathbb{R}} \ E(\mathcal{H}(Y, a) \ / \ X = x)$$

où: $x^+ = \max(x, \ 0), \ \ x^- = -\min(x, \ 0)$

Lorsque la loi conditionnelle est symétrique, la médiane conditionnelle et l'espérance

conditionnelle coïncident ($p. \ s.$) lorsqu'elles existent.

C'est une des motivation de notre choix d'étudier ces deux paramètres simultanément, qui ont beaucoup de propriétés communes.

4.1.3 Propriètés de la médiane conditionnelle:

Propriéte 1: Existence de la médiane conditionnelle.

C'est le résultat du théorème suivant:

Théorème 1: Soit $\mathcal{G}$ une $\sigma-$algèbre.

Pour toute v. a. r. il existe une médiane conditionnelle m **sachant** G.

Preuve: Si Y est une v. a. r. , il existe une version régulière de la fonction de répartition conditionnelle $P(Y \leq \ y \ / \ \mathcal{G})$, on définit, pour chaque ω dans Ω,

$$m(\omega) = \inf \{ \ y \in \ \mathbb{R} : P(Y \leq y \ / \ \mathcal{G}) \ (\omega) \ > \ 1/2\}$$

m est une v. a. $\mathcal{G}$-mesurable. Soit n un entier positif fixé et k un entier quelconque.

Lorsque l'évènement

$$[\omega : k-1 \ < \ nm(\omega) \ \leq k]$$

est réalisé, on a :

$$\begin{array}{l} P(Y \ < \ m - n^{-1}/ \ \mathcal{G}) \ \leq \ P(Y \leq \ \frac{k-1}{n} \ / \ \mathcal{G}) \leq \ 1/2 \\ \leq P(Y \leq \ \frac{k}{n} \ /\mathcal{G}) \ \leq \ P(Y \ \leq \ m + n^{-1} \ / \ \mathcal{G}) \quad \ p. \ s. \end{array}$$

et, en faisant varier k dans $\mathbb{Z}$, on obtient les inégalités suivantes :

$$\forall \, n \ \geq 1 \ P(Y < m - n^{-1}/ \ \mathcal{G}) \ \leq \ 1/2 \ \leq \ P(Y \ \leq m + n^{-1}/ \ \mathcal{G}) \ p.s.$$

Il suffit de faire tendre n vers l'infinit, et remplacer la premiére inéquation

$$P(Y < m - n^{-1}/ \ \mathcal{G}) \ \leq \ 1/2$$

par:

$$P(Y \geq m - n^{-1}/ \ \mathcal{G}) \ \geq \ 1/2$$

(passage à l'evènement complémentaire)on retrouve la définition1:

$$\text{ii)} \ P(Y \geq m \ /\mathcal{G} \ ) \ \geq \ 1/2 \ \leq \ P(Y \leq m/\mathcal{G}) \ \ p.s,$$

et on a i) m est une v. a. $\mathcal{G}$-mesurable, pour conclure de l'existence de la médiane conditionnelle. ■

Théorème 2: (Valadier(1984)

Notons $\mathcal{M}(x)$ l'ensemble des médianes conditionnelles de Y sachant $X = x$, alors:

$$\sup\{\|a\| \ : \ a \ \in \mathcal{M}(x)\} \ \leq \ 2E\ (\|Y\| \ / \ X = x)\ p.p$$

Preuve: Voir Valadier (1984)

Propriété 2:

Non unicité de la médiane conditionnelle.

Tomkins (1976) a démontré le résultat suivant:

Théorème 3:

Il existe une médiane conditionnelle m^* et une médiane conditionnelle m^{**} telles que :

$$m^* \leq \ m \ \leq \ m^{**}\ p.s$$

pour toutes médianes conditionnelles m et Y sachant $\mathcal{G}$.

Propriété 3: Inégalité de JENSEN pour les médianes conditionnelles.

Théorème 4:

Soit ϕ une fonction réelle convexe définie sur un intervalle I ouvert de $\mathbb{R}$, alors :

pour toute σ-algèbre $\mathcal{G}$, pour toute v. a. r. Y telle que $P(Y \in \ I) = 1$ et pour toute v. a. r. $m(Y/\ \mathcal{G})$

Il existe une médiane conditionnelle de la v. a. $\phi(Y)$ qui satisfait :

$$m(\phi(Y)\ /\ \mathcal{G})\ \geq\ \phi(m(Y\ /\ \mathcal{G})\ p.s$$

En particulier : $m(|Y|\ /\mathcal{G})$ et $m(Y^2\ /\ \mathcal{G})$ existent et vérifient :

$$m(|Y|\ /\mathcal{G})\ \geq |m(Y/\mathcal{G})| \qquad \text{et} \qquad m(Y^2/\mathcal{G})\ \geq m^2(Y \mid \mathcal{G}) \qquad p.s$$

(les fonctions$x \mapsto |x|$ et $x \mapsto x^2$ sont des de fonctions convexes sur $\mathbb{R}$)

Propriété 4: Inégalité de Chebychef pour les médianes conditionnelles.

Proposition 1:

Soit ε une *v. a. r.* positive alors :

$$P(|Y| > \varepsilon\ /X = x) \leq \varepsilon^{-\alpha}\ E((|Y|^{\alpha}\ /X = x)\ p.s$$

pour $\alpha > 0$

Preuve : Posons $Z = \varepsilon^{-1}\,|Y|$, alors:

$$E(|Y|^{\alpha}\ /X = x)\varepsilon^{-\alpha} = E((\varepsilon^{-1}\,|Y|^{\alpha})/X = x) = E(Z^{\alpha}/X = x)$$

$$\geq\ E(Z^{\alpha}1_{\{Z>1\}}/X=x)\ \geq E(1_{\{Z>1\}}/X=x)$$
$$=P(Z>1\ /X=x)=P(|Y|>\varepsilon/X=x).$$

Proposition 2: Pour tout $\alpha > 0$

$$|m(x)| \leq (2E(|Y|^{\alpha}/X=x)^{1/\alpha} \quad p.s$$

Preuve: D'après la proposition 1, pour tout $\delta > 0$,
$P(|Y| > [(2+\delta)\ E(|Y|^{\alpha}/X=x)]^{1/\alpha}/X=x) \leq \frac{E(|Y|^{\alpha}/X=x)}{(1+\delta)\ E(|Y|^{\alpha}/X=x)} = \frac{1}{2+\delta}\ p.\ s.$
Le théorème 4 permet de finir la démonstration. ■

Propriété 5:

Relation entre Médiane conditionnelle et Espérance conditionnelle.

Notons $V(Y/X=x)$ la variance conditionnelle de Y sachant $X=x$.

Proposition 3: On a $|m(x) - E(Y/X=x)| \leq \sqrt{2V(Y/X=x)}$

Preuve: IL suffit de remplacer dans la proposition 1
α par 2 e ε par $\sqrt{2E(|Y-E(Y/\ X=x)|^2/X=x)}$. ■

Dans ce qui suit on propose une suite de résultats qui se démontrent assez facilement en utilisant les définitions de la médiane conditionnelle(Tomkins(1976))

Théorème 5:

Soit ϕ une fonction strictement monotone sur $\mathbb{R}$, alors:
(i) $\phi(m)$ est une médiane conditionnelle de$Y$ sachant $\mathcal{G}$
Soit T une v. a $\mathcal{G}$-mesurable, alors :
(ii) si $m(T/\mathcal{G})$ est une médiane conditionnelle de T sachant $\mathcal{G}$ alors: $m(T/\mathcal{G}) = T\ p.\ s.$
(iii) $m(Y/\mathcal{G})\ + T$ est une médiane conditionnelle de $Y+T$, i. e.
$m(Y+T/\mathcal{G}) = m(Y/\mathcal{G}) + T.$
(iv)$Tm(Y/\mathcal{G})$ est une médiane conditionnelle de TY, i. e. $m(YT/\mathcal{G}) = Tm(Y/\mathcal{G})$.
Soient a un réel et Z une v. a. , alors :
(v) a$m(Y/\mathcal{G})$ est une médiane conditionnelle de aY, i. e.
$m(aY/\mathcal{G}) = a\ m(Y/\mathcal{G})$.
(vi) si $Y \geq Z\ p.\ s.$, $(Y,\ Z$ deux $v.\ a.\ r.$), alors: Il existe une médiane conditionnelle $m(Z/\mathcal{G})$
telle que: $m(Y/\mathcal{G}) \geq m(Z/\mathcal{G})\ p.\ s$

Propriété 6:

Remarques sur la médiane conditionnelle d'une somme de v. a. , sur la médiane conditionnelle d'une

médiane conditionnelle et sur l'inégalité de Cauchy-Schartz pour les médianes conditionnelles.

Au contraire de l'Espérance conditionnelle,
la Médiane conditionnelle ne vérifie pas les égalités suivantes :

$$m(Y + Z/\mathcal{G}) = m(Y/\mathcal{G}) + m(Z/\mathcal{G})$$

où $Y,\quad Z$ deux $v.\ a.\ r.$ et:

$$m(M(Y/\mathcal{G}) \mid \mathcal{G}_1) = m(Y/\mathcal{G}_1)$$

où $\mathcal{G}_1$ étant une $\sigma-$algèbre et $\mathcal{G}_1 \subseteq \mathcal{G}$ et M une médiane conditionnelle.

L'inégalité de Cauchy-Schartz n'est pas toujours vérifiée pour les médianes conditionnelles

(i. e) L'inégalité:

$$m^2(|ZY|/\mathcal{G}) \leq m(Y^2/\mathcal{G})\ M(Z^2/\mathcal{G})$$

n'est pas vraie.

Estimation de la médiane conditionnelle cas univarié

Introduction: La loi $P_{X,Y}$ de $(X,\ Y)$ est inconnue et n'est assujettie qu'à des hypothèses simples de régularité mathématique. On va utiliser pour l'estimation de la médiane conditionnelle deux méthodes non paramétriques. La première est empirique appelé médianogramme, est à rapprocher du régressogramme étudié notamment par Bosq (1969) et Geffroy (1980). La deuxième est la méthode du noyau abondamment utilisée dans le cas de l 'espérance conditionnelle

Méthode du Médianogramme.

Methode du Médianogramme:

Estimation Empirique de la Médiane conditionnelle.

$(X_i,\ Y_i)_{i=1,n}$ désigne une suite d'observations du vec. a. i. i. d. $(X,\ Y)$ dont l'espace d'états est $(\mathbb{R}^{d+1},\ \mathcal{B}(\mathbb{R}^{d+1}))$.

4.1.4 Construction du Médianogramme.

Soit:

$$J_n = \left\{ I_{n,z}, z \in \mathbb{Z}^d \right\}$$

le pavage de $\mathbb{R}^d$ défini par :

$$z = (z_1, ..., z_d) \in \mathbb{Z}^d \ \ et \ \ I_{n,z} = \prod_{k=1}^{d} [z_k h(n), (z_k + 1)h(n)[$$

$h(n)$ étant une suite positive tendant vers 0 avec $\frac{1}{n}$. Pour tout $x \in \mathbb{R}^d$, on désigne par $I(x,\ h(n))$ le pavé de J_n contenant x, $F_{h(n)}(./x)$ représente la f. r. c de Y sachant que $X \in I(x,\ h(n))$, on suppose qu'elle est absolument continue et qu'elle a une densité notée $f_{h(n)}(.\ /x)$.

$$m_{h(n)}\ (x) = F^{-1}_{h(n)}(1/2/x)$$

est une médiane conditionnelle locale de Y sachant que $X \in I(x,\ h(n))$. On note: $q(x,\ h(n))$ le nombre de X_i appartenant à $I(x,\ h(n))$: C'est la réalisation d'une v. a. $Q(x,\ h(n))$

$$Y_{i_1}, Y_{i_2}, ., Y_{i_{q(x,h(n))}}$$

sont les Y_i telles que $X_i \in\ I(x,\ h(n))$.

On définit l'estimateur empirique de $F_{h(n)}(y/x)$ qu'on note $\hat{F}_{h(n)}(y/x)$ de la façon suivante:

$$\widehat{F}_{h(n)}(y/x) = F_{q(x,h(n))}(Y_{i_1}\ , \ .\ .\ .\ .\ ,\ Y_{i_{q(x,h(n))}}\ ,\ y)$$

$$= \begin{cases} \frac{1}{q(x,h(h(n)))} \sum_{j=1}^{q(x,h(n))} 1_{\left\{Y_{i_j}\ \leq y\right\}} & \text{si : } q(x,h(n)) \neq 0 \\ 0 & \text{si : } q(x,h(n)) \neq 0 \end{cases}$$

Ainsi l'estimateur empirique $\hat{m}_{h(n)}(x)$ de $m_{h(n)}(x)$ est défini de la façon suivante:

$\hat{m}_{h(n)}(x) = \hat{F}^{-1}_{h(n)}(1/2/x)$

$$= \begin{cases} Y_{([i_{q(x,h(n)1/2)}])} & \text{si } q(x,h(n)) \text{ est imp} \\ (Y_{([i_{q(x,h(n)1/2)}])}\ +\ Y_{([i_{q(x,h(n)1/2+1)}])}\ )/2\ , & \text{si } q(x,h(n)) \text{ est pair} \\ m_o & \text{si } q(x,h(n)) = 0 \end{cases}$$

avec m_o valeur arbitraire

4.1.5 Etude des modes de convergences

Etude des Différents Modes de convergences du Médianogramme.

les principales hypothèses H sont :

(H1) la f. r. $F(.\ ,\ .)$ du couple aléatoire $(X,\ Y)$est absolument continue et ayant une densité f bornée.

(H2) la f. r. c. $F(.\ /x)$est absolument continue et on note$f(.\ /x)$ sa densité.

(H3) la loi marginale de X a une densité g strictement minorée.

(H4) $f(.\ ,\ .\ )$ est lipschitzienne d'ordre 1 sur $\bar{I}(x,\ h(n)) \times \mathbb{R}$.

(H5) $g(.\ )$est lipschitzienne d'ordre 1 sur $\bar{I}(x,\ h(n))$.

(H6) la médiane conditionnelle $m(x)$ de Y sachant $X = x$ est unique.

(H7) la médiane conditionnelle locale $m_{h(n)}$ de Y sachant $X \in I(x,\ h(n))$ est unique.

(H8) $\sum\limits_{n=0}^{\infty} \exp(-\lambda n h^d(n)) < \ \forall \lambda > 0.$

♦ Cas d'un échantillon: Observations Non Corrélées

Dans ce qui suit $(X_i,\ Y_i)_{i=1,n}$désigne une suite d'observations indépendantes et de même lois que le couple$(X,\ Y)$

Convergence Ponctuelle de $m_{h(n)}$ vers $m(x)$

On commence par établir (ou rappeler) les lemmes qui seront utilisés pour les démonstrations des théorèmes qui vont suivre:

Lemme 1: Bosq(1969)**:**

Sous les hypothéses: (H1), (H2), (H3), (H4) et (H5), on a :

$$\sup_{y \in \mathbb{R}} \left| F_{h(n)}(y/x) - F(y/x) \right|_{n \to \infty} \longrightarrow 0\,,$$

Preuve: Bosq (1969) a démontré :

$$\sup_{y \in \mathbb{R}} \left| f_{h(n)}(y/x) - f(y/x) \right| \longrightarrow 0\,,\ n \longrightarrow \infty$$

la démonstration du lemme est une conséquence de ce résultat et d'un théorème bien connu de Scheffe Billingsley (1968)p. 224. Des indications complémentaires se trouvent dans l'article de Bosq (1973). ■

Le lemme suivant a été établi par Dvoretzky-Kieffer-Wolfowitz (1956)

Lemme 2:

Soit $F(.\ )$ la fonction de répartition d'une v. a. r. Y. $\hat{F}_n$désigne l'estimateur empirique de F défini par

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=n}^{n} 1_{\{Y_i \,\le\, y\}}$$

$(Y_i)_{i=1,n}$ étant des v. a. de même loi que Y. Alors, $\forall\ \varepsilon\ >\ 0$

$$P(\sup_y \mid \hat{F}_n(y) - F(y) \mid >\ \varepsilon\ )\ \le C\ \exp(-2n\varepsilon^2)$$

.

C étant une constante universelle positive

Preuve: Se reférer a l'article de Devroy (1982). $\blacksquare$.

Ceci servira pour la démonstration du lemme suivant:

Lemme 3: Sous l'hypothèse (H3), si:

$$\lim_{n\to\infty} nh^d(n) \;=\; \infty$$

alors:

$$\sup_y \left|\hat{F}_{h(n)}(y/x) - F_{h(n)}(y/x)\right| \;_{n\to\infty} \to^p 0$$

Preuve: Soit:

$$B(x,h(n)) \;=\; \{(x,y) \in I(x,h(n)) \times \; \mathbb{R}\}$$

On note

$$p(x,h(n)) = Prob((X,Y) \in B(x,h(n))) \; = F(x+h(n),+\infty) - F(x-h(n),+\infty)$$

et:

$$d(\hat{F}_{h(n)}(\cdot/x), F_{h(n)}(\cdot/x)) = d(\hat{F}_{h(n)}, F_{h(n)}) = \sup_y \left|\hat{F}_{h(n)}(y/x) - F_{h(n)}(y/x)\right|$$

En utilisant la formule de probabilité totale, on peut écrire

$P(d(\hat{F}_{h(n)},\; F_{h(n)}) \;>\; \varepsilon) = \sum_{q\,=\,0}^{n} C_h^q \, p\,(x,\; h(n))^q(1 - p(x,\; h(n))^{n-q} \times$

$P(d(\hat{F}_{h(n)},\; F_{h(n)}) > \varepsilon/q(x,\; h(n)) = q)$

D'aprés la définition de $\hat{F}_{h(n)}$, on peut écrire

$$P(d(\hat{F}_{h(n)}, F_{h(n)}) > \varepsilon/q(x, h(n) = q) = P(d(\hat{F}_q, F_{x,h(n)}) \; > \varepsilon)$$

$\hat{F}_q$désignant la f. r. empirique obtenue à partir des qY_i. Le lemme 2 permet d'avoir :

$$P(d(\hat{F}_{h(n)}, F_{h(n)}) \;>\; \varepsilon \; \le c \sum_{q=0}^{n} C_h^q \; p\;(x, h(n))^q(1 - p(x, h(n))^{n-q} \times \exp(-2q\varepsilon^2\;)$$

$$\le c \sum_{q\,=\,0}^{n} C_h^q \; p\;(x, h(n) \exp(-2\varepsilon^2))^q(1 - p(x, h(n))^{n-q} \le \; c(1 - (p(x, h(n))(1 - \exp(-2\varepsilon^2)))^n$$

or l'expression:

$$(1-(p(x,h(n))(1-\exp(-2\varepsilon^2)))^n = \exp(n\log(1-p(x,h(n))(1-\exp(-2\varepsilon^2)))$$

En utilisant le théorème de la moyenne, on peut affirmer qu'il existe: $\xi \in I(x,\ h(n))$ tel que:

$$p(x,h(n)) = h^d(n)g(\xi)$$

D'aprés l'hypothèse (H3), g est strictement minorée, alors il existe $b\ >\ 0$ tel que : $\forall\ x \in \mathbb{R}^d,\ g(x)\ > b$. On a:

$$\log(1-x) \leq -x$$

Ceci entraine:

$$n\log(1-h^d(n)b(1-\exp(-2\varepsilon^2))) \leq -nh^d(n)b(1-\exp(-2\varepsilon^2))$$

comme:

$$1-\exp(-2\varepsilon^2) \leq 2\varepsilon^2$$

on a :

$$n\log(1-h^d(n)b(1-\exp(-2\varepsilon^2))) \leq -nh^d(n)b(1-\exp(-2\varepsilon^2) \leq -nh^d(n)2\varepsilon^2 b$$

donc:

$$P(d(\hat{F}_{h(n)}, F_{h(n)})\ >\ \varepsilon) \leq c\exp(-2\varepsilon^2 bnh^d(n)) \quad (1)$$

$\lim_{n\to\infty} nh^d(n) = \infty$ d'aprés (H8) donc: la serie de terme général

$$\exp(-2\varepsilon^2 bnh^d(n))$$

est convergente et donc le majorant de

$$P(d(\hat{F}_{h(n)}, F_{h(n)})\ >\ \varepsilon)$$

tend vers 0. ■

Lemme 4: Sous les hypothèses: (H3) et (H8) on a :

$$d(\hat{F}_{h(n)}, F_{h(n)})_{n\to\infty} \longrightarrow^{p.co} 0$$

Preuve: En utilisant (1) dans la démonstration du lemme 3, on a

$$\sum_{n=0}^{\infty} P(d(\hat{F}_{h(n)}, F_{h(n)})\ >\ \varepsilon) \leq c\sum_{n=0}^{\infty}\exp(-2\varepsilon^2 bnh^d(n))$$

Le majorant est fini d'aprés (H8). ■

Lemme 5: Sous (H6) et (H7), si $\hat{F}_{h(n)}(.\ /x)$ converge uniformément vers $F_{h(n)}(.\ \mid x)$: (C)

Alors, pour tout x fixé :

$$\hat{m}_{h(n)}(x)_{n\to\infty} \ \rightarrow \ m_{h(n)}(x) \quad p.s$$

Preuve: $\left|F_{h(n)}\ (m_{h(n)}(x)/x) - F_{h(n)}\ (\dot{m}_{h(n)}(x)/x)\right| \leq$

$\left|F_{h(n)}\ (m_{h(n)}(x)/x) - (\hat{F}_{h(n)}\ (m_{h(n)}(x)/x)\right| + \left|\hat{F}_{h(n)}\ (m_{h(n)}(x)/x) - F_{h(n)}\ (m_{h(n)}(x)/x)\right|$

$= \qquad A \qquad + \qquad B$

$A \leq \sup_{x\in\mathbb{R}} \left|F_{h(n)}(y/x) - \hat{F}_{h(n)}(y/x)\right| \longrightarrow 0 \ \ p.\ s.$ en raison de (C).

$B \ = 0$ d'aprés la définition de $\hat{m}_{h(n)}(x)$ et $m_{h(n)}(x)$. L'unicité de la médiane conditionnelle et la continuité de $F_{h(n)}(.\ /x)$ permettent de finir la démonstration. ■

Théorème 1: Convergence ponctuelle p. s

Sous les hypothéses **H**: (H1)- (H8), $\forall x$ fixé on a:

$$\lim_{n \longrightarrow \infty} \hat{m}_{h(n)}(x) \ = m(x) \quad p.s.$$

Preuve:

$$\left|m_{h(n)}(x) - m(x)\right| \leq \left|m_{h(n)}(x) - m_{h(n)}(x)\right| + \left|m_{h(n)}(x) - m(x)\ \right| \ \leq \ A \ + \ B$$

A converge vers 0 ($p.\ s.$) d'aprés les lemmes 4 et 5. Pour terminer la démonstration du théorème, il faut prouver que B converge vers 0

$$\begin{aligned}\left|F(m_{h(n)}(x)/x) - F(m(x)/x)\right| \ \ \leq \ \ & \left|F(m_{h(n)}(x)/x) - (F_{h(n)}(m_{h(n)}(x)/x)\right| + \\ & \left|F_{h(n)}(m_{h(n)}(x)/x) - F(m(x)/x)\right|\end{aligned}$$

or

$$\left|F(m_{h(n)}(x)/x) - F_{h(n)}\ (m_{h(n)}(x)/x)\right| \leq \sup_{y} \left|F(y/x) - F_{h(n)}\ (y/x)\right|$$

d'aprés le lemme 1, le deuxième membre de cette inégalité tend vers 0 et à fortiori le premier.

$$\left|F_{h(n)} m_{h(n)}(x)/x) - F(m(x)/x)\right| = 0$$

en raison de la définition de $m_{h(n)}(x)$ et de $m(x)$. L'unicité de $m(x)$ et la continuité de $F(.\ /x)$ permettent de conclure que B converge vers 0, et ainsi d'achever la démonstration. ■

Convergence uniforme de $\hat{m}_{h(n)}$vers m

Soit l'hypothèse (H9): C un compact de $\mathbb{R}^d$, on peut lui construire un recouvrement par un nombre ℓn fini de cubes de $\mathbb{R}^d$d'arête $h(n)$, en ce sens qu'il existe $x_1,\ x_2,\ .\ .\ .\ ,\ x_{\ell n}$ dans C, tels que

$$C \subset \bigcup_{i=1}^{\ell n} I(x_i, h(n)$$

On choisit $h(n)$de sorte que:

$$\ell n \leq \frac{L}{h^d(n)}$$

L est une constante positive indépendante de n.

On supose (H10) que $F(.\ /x)$ est lispschitzienne de rapport $\delta\ > 0$ et de l'ordre 1 sur C uniformément par rapport à y, (l'avantage de ce choix est la simplification des calculs), alors:

$$|F(y/x) - F(y/x_i)| \leq \delta \|x - x_i\| \ \forall\, y\ \in \mathbb{R}$$

Soit (U): L'hypothèse supplémentaire suivante dite **d'unicité uniforme** :

$\forall\, \varepsilon\ > 0,\ \exists\, \alpha\ >\ 0\ ,\ \forall\, t:\ C \longrightarrow \mathbb{R}$

$$\sup_{x\in C} |m(x) - t(x)| \geq \varepsilon\ \Rightarrow \sup_{x\in C}\ |F(m(x)/x) - F(t(x)/x)| \geq \alpha$$

Lemme 6: Sous l'hypothèse: (U), et si on a :

$$\sup_{x\in C}\sup_{y\in\mathbb{R}} \left|\hat{F}_{h(n)}(y/x)) - F(y/x)\right|_{n\to\infty} \to^{p.co} 0$$

Alors:

$$\sup_{x\in C} \left|\hat{m}_{h(n)}(x) - m(x)\right|_{n\to\infty} \longrightarrow^{p.co} 0$$

Preuve: En remplaçant dans l'hypothèse (U) t par $\hat{m}_{h(n)}$, on peut écrire :

(a) $\forall\, \varepsilon\ >\ 0,\ \ \exists\, \alpha\ >\ 0$ tel que:

$$\sup_{x\in C} \left|\hat{m}_{h(n)}(x) - m(x)\right| \geq \varepsilon \Longrightarrow \sup_{x\in C} \left|F(\hat{m}_{h(n)}(x)/x) - F(m(x)/x)\right| \geq \alpha$$

or on a:

$$\left|F(\hat{m}_{h(n)}(x)/x) - F(m(x)/x)\right| \leq \left|F(\hat{m}_{h(n)}(x)/x) - \hat{F}_{h(n)}(\hat{m}_{h(n)}(x)/x)\right| + \left|\hat{F}_{h(n)}(\hat{m}_{h(n)}(x)/x) - F(m(x)/x)\right|$$

ceci donne:

$$\left|F(\hat{m}_{h(n)}(x)/x) - F(m(x)/x)\right| \leq 2\sup_{x\in C}\sup_{y\in\mathbb{R}}\left|\hat{F}_{h(n)}(y/x)) - F(y/x)\right|$$

En utilisant (a), on obtient : $\forall\, \varepsilon > 0$, $\exists\, \alpha > 0$ tel que:

$$\sup_{x\in C}\left|\hat{m}_{h(n)}(x) - m(x)\right| \geq \varepsilon \Longrightarrow \sup_{x\in C}\sup_{y\in\mathbb{R}}\left|\hat{F}_{h(n)}(y/x)) - F(y/x)\right| \geq \alpha/2$$

D'où : $\forall\, \varepsilon > 0$, $\exists\, \alpha > 0$ tel que:

$$P(\sup_{x\in C}\left|\hat{m}_{h(n)}(x) - m(x)\right| \geq \varepsilon) \leq P(\sup_{x\in C}\sup_{y\in\mathbb{R}}\left|\hat{F}_{h(n)}(y/x)) - F(y/x)\right| \geq \alpha/2)$$

Si la série dont le terme général est la partie droite de l'inéquation de dessus, est convergente, alors il sera de même pour la série dont le terme général est la partie gauche par comparaison de la nature des deux séries. ■

Théorème 2:

Sous les hypothèses **H**: (H1)- (H7), (H9), (H10), (U) et où (H8)est remplacée par

$$\lim_{n\to\infty} nh^d(n)/\log n = \infty \qquad \text{(H11)}$$

Alors:

$$\sup_{x\in C}(m_{h(n)}(x) - m(x)) \underset{n\to\infty}{\overset{p.co.}{\longrightarrow}} 0$$

Preuve: Il sufit de démontrer:

$$\sup_{x\in C}\sup_{x\in\mathbb{R}}\left|\hat{F}_{h(n)}(y/x)) - F(y/x)\right| \underset{n\to\infty}{\overset{p.co.}{\longrightarrow}} 0$$

et d'appliquer le lemme 6. D'aprés l'hypothèse (H10),

$$|F(y/x) - F(y/x_i)| \leq \delta\,\|x - x_i\|\,,\ \forall\, x \in C \text{ et } \forall\, i\ \in \{1, ..., \ell n\}\,.$$

Ainsi pour chaque $x \in I(x_i,\ h(n))$, on a:

$$|F(y/x) - F(y/x_i)| \leq \delta.h^d(n),\ \forall\, i \in 1, ..., \ell n$$

Compte tenu du fait que $\lim\limits_{n\longrightarrow\infty} h(n)=0$ et de l'inégalité:

$$\begin{aligned}\left|\hat{F}_{h(n)}(y/x)-F(y/x)\right| &\leq \left|\hat{F}_{h(n)}(y/x)-F_{h(n)}(y/x_i)\right| \\ &+\left|F_{h(n)}(y/x_i)-F(y/x_i)\right|+|F(y/x_i)-F(y/x)|\end{aligned}$$

$\forall\ x\in I(x_i,\ h(n)),\ \forall\ i\ \in 1,\ \ldots,\ \ell n$. En utilisant le lemme 1 et (1) de la démonstration du lemme 3, On obtient pour n assez grand :

$$P(\sup_{x\in C}\sup_{y\in\mathbb{R}}\left|\hat{F}_{h(n)}(y/x)-F(y/x)\right|>\ \varepsilon)\leq P(\sup_{1\leq i\leq \ell n}\sup_{y\in\mathbb{R}}\left|\hat{F}_{h(n)}(y/x_i))-F(y/x_i)\right|\geq\varepsilon\ /3)$$

$$\leq\sum_{i=1}^{\ell n}P(\sup_{y\in\mathbb{R}}\left|\hat{F}_{h(n)}(y/x_i))-F(y/x_i)\right|\geq\varepsilon\ /3)\leq c\sum_{i=1}^{\ell n}\exp-\frac{2\varepsilon^2}{9}bnh^d(n).$$

Notons

$$\mathbf{b(n)}=C\ell n\exp-\frac{2\varepsilon^2}{9}bnh^d(n)$$

$\lim_{n\to\infty}nh^d(n)=\infty$, alors : On peut trouver $\eta\ >\ 0$ telle que: $(nh^d(n))^{-1}<\eta$ et il s'en suit que :

$$\mathbf{b(n)}\leq CL\eta n^{1-\gamma(n)}$$

avec:

$$\gamma(n)=\frac{2\varepsilon^2}{9}\ bnh^d(n)\ /\ \log n\longrightarrow\infty$$

avec n. L'hypothèse (H11) permet de definir la démonstration. ■

Etude de laVitesse de convergence.

Soient les hypothèses supplémentaires suivantes :

a) Il existe une constante M_0 telle que :

$$\left|m(x)-m(x')\right|\leq M_0\left\|x-x'\right\|$$

$\forall\ x,\ x^{'}$ dans C

b) Il existe M_1vérifiiant :

$$M_1^{-1}\ <g(x)\ <\ M_1$$

$\forall\ x$ dans Coù g est la marginale de X

c) $f(y/x)$ est bornée sur un voisinage de la médiane de sorte qu'il existe une constante positive ε_0, telle que :

$$M_1^{-1}\leq f(y/x)\leq M_1$$

$\forall\ x \in C$ et $y \in (m(x) - \varepsilon_0\ ,\ m(x) + \varepsilon_0)$. On pose:

$$r = 1/(d+2)$$

Truong (1989) a établi et démontré le résultat suivant:

Théorème 3: Sous les condition : a), b), c), si:

$$h(n) \simeq (n^{-1} \log n)^r$$

il existe $\tau \rangle\ 0$ telque:

$$\lim_n P(\sup_{x \in C} |\hat{m}_n(x) - m(x)| \geq \tau(n^{-1} \log n)^r) = 0$$

Remarque: Ce théorème est la suite des travaux de Stone (1982) qui a démontré que :

$$(n^{-1} \log n)^r$$

est la vitesse convergence optimale de $\|r_n(x) - r(x)\|_\infty$où $r_n(x)$ étant un estimateur de $r(x) = E(Y/X = x)$.

4.2 La Méthode du Noyau

On rappelle qu'une fonction de $L_1(\mathbb{R}^d)$ est un noyau si $\sup_{x\in\mathbb{R}^d} |K(x)| <$ ∞ et $\lim \|x\|^d K(x) = 0$Desdétails et des exemples se trouvent dans l'article de Rosemblatt(1956). On supposera par la suite que K est un noyau strictement positif

4.2.1 Construction de l'estimateur.

Définition de l'estimateur m_n de la médiane conditionnelle par la méthode du noyau:

Notons:

$$K_{h(n)}(x) = \frac{1}{h^d(n)}\ K(\frac{x}{h(n)})\ \forall x \in \mathbb{R}^d$$

$h(n)$ étant une suite réelle strictement positive de limite nulle. Soit P_Y^X une version régulière de la probabilité conditionnelle de Y par rapport à X. $\forall x \in \mathbb{R}^d$, on considère l'estimateur $(P_Y^{X=x})_n$ de $(P_Y^{X=x})$ défini par :

$$(P_Y^{X=x})_n = \frac{\sum\limits_{i=1}^n \delta_{Y_i} K_{h(n)}(x - X_i)}{\sum\limits_{i=1}^n K_{h(n)}(x - X_i)}$$

d'où :

$$\hat{F}_n(y/x) = \int_{\{Y\ \leq\ y\}} (P_Y^{X=x})_n(dy)$$

est un estimateur de la fonction de répartition conditionnelle de Y sachant $X = x$ définie par:

$$\hat{F}_n(y/x) = \int_{\{Y\ \leq\ y\}} (P_Y^{X=x})_n(dy) = \frac{\sum_{i=1}^{n} 1_{\{Y_i \leq y\}} K_{h(n)}(x - X_i)}{\sum_{i=1}^{n} K_{h(n)}(x - X_i)}.$$

Soient:

$$\hat{m}_n^P(x) = \inf\left\{y\,:\,\hat{F}_n(y/x) \geq 1/2\right\}$$

$$\hat{m}_n^G(x) = \sup\left\{y\,:\,\hat{F}_n(y/x) \leq 1/2\right\}$$

$$\hat{m}_n(x) = \frac{\hat{m}_n^P(x) + \hat{m}_n^G(x)}{2}$$

Alors: $\hat{m}_n(x)$ est un estimateur par la méthode du noyau de $m(x)$.

4.2.2 Etude des Modes de convergences

Etude des Différents Modes de convergence de m_n.

Hypothèses généralesH' : Soient les hypothèses suivantes

Le noyau K est borné à support borné et $\displaystyle\int K(u)du = 1$ (H'1)

la suite $nh^d(n) \longrightarrow \infty,\ n \longrightarrow \infty$ (H'2)

La loi marginale de X a une densité f continue est positive au voisinage de x (H'3)

La fonction $F(y/.)$ est continue au voisinage de x (H'4)

Lafonction $F_Y(./x)$ est continue (H'5)

La médiane conditionnelle $m(x)$ est unique (H'6)

On suppose l'existance $(\forall y)$ d'une application définie sur $\mathbb{R}^d$ telle que :

$$E(I_{\{Y_i \leq y\}}/.) = F(y/.) \quad (1)$$

On désigne par C un compact de $\mathbb{R}^d$ et on suppose que la loi P_{X_i} de X_i est uniforment absolument continue par rapport à la mesure (λ) de lebesgue c. a. d. :

$$\left\{\begin{array}{l} \exists\Gamma \ < \ \infty : P(X_i \in B) \leq \Gamma\lambda(B), \forall i \in \mathbb{N}, \forall B \in \mathbb{B}_{\mathbb{R}}d \\ \exists\gamma \ > \ 0 : P(X_i \in B) \geq \gamma\lambda(B), \forall i \in \mathbb{N}, \forall B \in \mathbb{B}_{\hat{C}} \end{array}\right. \qquad (2)$$

$\widehat{C}$ étant un $\varepsilon-$ voisinage compact de C dans $\mathbb{R}^d$

Le noyau K de $\mathbb{R}^d$ vérifie :

$$\left\{\begin{array}{l} K(.) < K_1 < \infty, \int K(z)dz = 1 \\ \|z\|^d K(z) \longrightarrow 0, \|z\| \longrightarrow \infty \end{array}\right. \qquad (\ 3)$$

On se rétreint au cas d'un échantillon: Observations non corrélées

Convergence en probabilité et dans L^r de $m_n(.\ )$:

Lemme 1. Hardle et Tsybakov (1988):

Sous les hypothèses (H'1)-(H'6), on a :

$$\sup_{y\in\mathbb{R}} \left|\hat{F}_n(y/x) - F(y/x)\right| \underset{n\to\infty}{\overset{p.}{\longrightarrow}} 0$$

Preuve: voir Hardle et Tsybakov (1988).

Lemme 2:

Sous (H1)-(H6), on a : $\forall\ \varepsilon > 0$

$$\begin{array}{l} \lim_n P(m_n(x) \geq m(x) - \varepsilon) = 1 \\ \lim_n P(m_n(x) \leq m(x) + \varepsilon) = 1 \end{array}$$

Preuve: On pose

$$W_{ni} = \frac{K_{ni}(x - X_i)}{\sum\limits_i^n K_{ni}(x - X_i)}$$

et on montre seulement la première partie du lemme, la seconde est identique. Posons:

$$\varphi(x) = F(m(x) - \frac{\varepsilon}{2}/x)$$

D'aprés le lemme 6 et le théorème 1 de Stone(1977), on a:

$$\lim_{n\longrightarrow\infty} E\left|\sum_i W_{ni}(x) I_{\left\{Y_i \leq m(x_i) - \frac{\varepsilon}{2}\right\}} - \varphi(x)\right| = 0$$

ceci implique:

$$\lim_{n\longrightarrow\infty} P(\sum_i W_{ni}(x) I_{\{Y_i \leq m(x_i) - \varepsilon/2\}} \geq \frac{1/2 + \varphi(x)}{2}) = 0.(*)$$

la proposition 4 de Stone(1977) permet d'écrire:

$$\sum_i W_{ni}(x) I_{\{m(x_i) \,\leq\, m(x)-\varepsilon/2\}} \; {}_{n\to\infty} \overset{p.}{\longrightarrow} 0$$

d'où:

$$\lim_{n\longrightarrow\infty} P(\sum_i W_{ni}(x) I_{\{m(x_i) \,\leq\, m(x)-\varepsilon/2\}} \geq \frac{1/2-\varphi(x)}{2}) = 0 \quad (**)$$

(*) et (**) impliquent

$$\lim_{n\longrightarrow\infty} P(\sum_i W_{ni}(x) I_{\{Y_i \,\leq\, m(x)-\varepsilon\}} < 1/2) = 1$$

et donc

$$P(m_n(x) \geq m(x)-\varepsilon) = 1$$

On suppose maintenant que: $E(|Y|^r) < \infty,\;\; r \geq 1$, alors :

Théorème1: Sous les hypothèses (H2)-(H6), si le noyau K est à support compact et minoré sur ce compact par un nombre strictement positif, alors:

$$m_n(x)_{n\to\infty} \longrightarrow m(x)$$

en probabilité et dans L^r

Preuve: On a:

$$P(|m_n(x)-m(x)| \geq \varepsilon) = 1 - P(|m_n(x)-m(x)| < \varepsilon)$$

or:

$$\begin{aligned} P(|m_n(x)-m(x)| &< \varepsilon) = P(m(x)-\varepsilon < m_n(x) < \varepsilon + m(x)) = \\ P(m_n(x) &< \varepsilon + m(x)) - P(m_n(x) > m(x)-\varepsilon) \end{aligned}$$

$$= 1 - P(m_n(x) \geq \varepsilon + m(x)) - P(m_n(x) > m(x)-\varepsilon)$$

Il suffit d'appliquer le lemme 2 pour conclure:

$$m_n(x)_{n\to\infty} \overset{p.}{\longrightarrow} m(x)$$

D'aprés la proposition (2)

$$|m(x)|^r \leq (2E(|Y|^r /X = x)) \Longrightarrow E(|m(x)|^r) \leq 2E(|Y|^r) \quad (1)$$

propriété de l'espérance conditionnelle, pour $m(x) > 0$,

$$|m(x)|^r I_{\{m_n(x) \geq m(x)\}} \leq (2 \sum_i W_{ni}(x) |y|^r I_{\{|Y_i| \geq m(x)\}}),$$

en utilisant le lemme 2 et le théorème 1 de Stone(1977), on peut écrire (2)

$$E(|m_n(x)|^r I_{\{m_n(x) \geq m(x)\}} \leq 2E(|Y|^r I_{\{|Y| \geq m(x)\}}). \quad (2)$$

Comme:

$$E(|Y|^r) < \infty,$$

(1), (2) et la convergence en probabilité permettentd'avoir **la convergence dans** L^r de $m_n(x)$ vers $m(x)$. ■

Convergence simple(ponctuelle en x**) de** $m_n(x)$ **ver** $m(x)$

On remplace dans ce qui précède l'hypothèse (H2) par :

$$\sum_{n \geq 1} \exp(-\lambda n h^d(n)) < \infty \ , \forall \lambda > 0 \quad \text{(H2')}$$

.

Théorème2: Sous les hypothèses du théorème 1, on a : Pour presque tout x

$$m_n(x)_{n \to \infty} \longrightarrow m(x) p.s.$$

Preuve: Sous la condition (H2')on a: $\sup_{y \in \mathbb{R}} \left| \hat{F}_n(y/x) - F(y/x) \right|_{n \to \infty} \longrightarrow 0$

$$\begin{aligned} \left| \hat{F}_n(m_n(x)/x) - F(m(x)/x) \right| &\leq \left| \hat{F}_n(m_n(x)/x) - F(m_n(x)/x) \right| \\ &\quad + |F(m_n(x)/x) - F(m(x)/x)| \\ &= A + B \end{aligned}$$

$$A \leq \sup_{y \in \mathbb{R}} \left| \hat{F}_n(y/x) - F(y/x) \right|_{n \to \infty} \xrightarrow{p.s.} 0$$

on a: $B = 0$, d'aprés la définition de la médiane conditionnelle, ceci donne :

$$\left| \hat{F}_n(m_n(x)/x) - F(m(x)/x) \right| \leq \sup_{y \in \mathbb{R}} \left| \hat{F}_n(y/x) - F(y/x) \right| {}_{n \to \infty} \xrightarrow{p.s.} 0$$

L'unicité de la médiane conditionnelle et la continuité de $F(.\ /x)$ permettent de conclure. ■

Convergence uniforme de m_n **ver** m

Posons:

$$F^{1,n}(x,y) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Y_i \le y\}} K_{h(n)}(x - X_i)$$

avec:

$$K_{h(n)}(t) = \frac{1}{h^d(n)} K(\frac{t}{h(n)})$$

et:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h(n)}(x - X_i)$$

alors :

$$\hat{F}_n(y/x) = \frac{F^{1,n}(x,y)}{\hat{f}_n(x)}$$

Ceci donne :

$$\hat{F}_n(y/x) - F_y(x) = \frac{F^{1,n}(x,y)}{\hat{f}_n(x)} - F(y/x) = \frac{F^{1,n}(x,y) - \hat{f}_n(x)F(y/x)}{\hat{f}_n(x)}$$

,

$$\sup_{x\in C}\sup_{y\in\mathbb{R}} \left|\hat{F}_n(y/x) - F(y/x)\right| \le \frac{\sup_{x\in C}\sup_{y\in\mathbb{R}} \left|F^{1,n}\ (x,y) - E(F^{1,n}(x,y))\right|}{\inf_{x\in C} f_n(x)} +$$

$$+\frac{\sup\ _{x\in C}\left|\hat{F}_n(x) - E(\hat{F}_n(x)\right|}{\inf_{x\in C} \hat{f}_n(x)} + \frac{\sup_{x\in C}\sup_{y\in\mathbb{R}} |E(|\ F^{1,n}(x,y) - F(y/x)E(\hat{f}_n(x))\rfloor}{\inf_{x\in C} \hat{f}_n(x)}$$

Remarque: le lemme suivant servira dans la plupart de nos démonstrations (Inégalité de Bernstein)

Lemme 3:

Soit: $\{\triangle_i\}_{i=1,n}$ une suite de v. a. r. qui vérifie:

$$E(\triangle_i) = 0, E(\triangle^2) \le D^2$$

alors:

$$P(\frac{1}{n} \mid \sum_{i=1}^{n} \triangle_i \mid > t) < 2\exp(\frac{nt^2}{4D^2}), 0 \le t \le D^2$$

.

Preuve: voir par exemple Collomb (1976).

Lemme 4: Sous les hypothèses: (2) et (3), on a: Pour n assez grand et pour tout $t > 0$

$$\sup_x \sup_y P(\mid F^{1,n}(x,y) - E(F^{1,n}(x,y)) \mid > t) \leq 2\exp(-bnh^d(n))$$

(cette relation sera appelée (R1) par la suite)

Preuve: C'est une application du lemme 3. Posons :

$$\triangle_i = I_{\{Y_i \leq y\}} K(\frac{x - X_i}{h(n)}) - E(I_{\{Y_i \leq y\}} K(\frac{x - X_i}{h(n)}))$$

$$E({\triangle^2}_i) \leq h^d(n) K_1 \Gamma = D^2$$

$$P(\left|F^{1,n}(x,y) - E(F^{1,n}(x,y))\right| > t) = P(\frac{1}{nh^d(n)} \left|\sum_{i=1}^{n} \Delta_i\right| > t) = P(\frac{1}{n} \left|\sum_{i=1}^{n} \triangle_i\right| > h^d(n)t)$$

on pose:

$$b = \frac{t^2}{4K_1\Gamma}$$

et on applique le lemme 3 pour avoir :

$$P(\left|F^{1,n}(x,y) - E(F^{1,n}(x,y))\right| > t) \ \leq \ 2\exp(-bnh^d(n))$$

cette majoration ne dépend ni de x ni de y et ceci permet d'écrire

$$\sup_x \sup_y P(\mid F^{1,n}(x,y) - E(F^{1,n}(x,y)) \mid > t) \leq 2\exp(-bnh^d(n)) \quad \blacksquare$$

.

Lemme 5: Sous les hypothèses: (2) et (3), si :

(i) le noyau K est lipschitzien.

(ii) La serie $\sum_{n=0}^{\infty} \exp(-\lambda nh^d(n) < \infty, \ \lambda > 0,$

.

Alors:

$$\sup_{x \in C} \sup_{y \in \mathbb{R}} \mid F^{1,n}(x,y) - E(F^{1,n}(x,y)) \mid_{n\to\infty} \to 0 \ p.co$$

Preuve: On suppose que K est de rapport $L < \infty$ et d'ordre $\gamma > 0$, alors :

$$\left|K(\frac{x - X_i}{h(n)}) - K(\frac{x' - X_i}{h(n)})\right| \leq Lh(n)^{-\gamma} \ \left\|x - x'\right\|^{\gamma}, \forall (x,x') \in \mathbb{R}^{2d}.$$

On se fixe un réel α qui vérifie :

$$\alpha > 1 + \frac{1}{\gamma} \quad (I)$$

Pour tout $n \in \mathbb{N}$, on désigne par $\{I_{t_k}, k = 1, ..., l_n\}$un recouvrement de C par des boules de centre t_k de rayon$\leq h(n)^\alpha$. Puisque C est borné, on peut choisir un recouvrement tel que : $\forall n \in \mathbb{N}$,

$$l_n \leq Bh(n)^{-d\alpha} \; (*)$$

où B est une constante strictement positive indépendante de n. On pose :

$$\triangle_n(x, y) = \frac{1}{nh^d(n)} \left[\sum_{i=1}^{n} I_{\{Y_i \leq y\}} K(\frac{x - X_i}{h(n)}) - E(I_{\{Y_i \leq y\}} K(\frac{x - Xi}{h(n)})) \right]$$

$$\triangle_n(x, y) = \triangle_n(t_k, y) + \triangle_n(x, y) - \triangle_n(t_k, y) \; (**)$$

or on a:

$$|\triangle_n(x, y)| \leq |\triangle_n(t_k, y) - \triangle_n(x, y)| + |\triangle_n(t_k, y)| = \left| \hat{\triangle}_n(x, y) \right| + |\triangle_n(t_k, y)|$$

avec :

$$\hat{\Delta}_n \; (x, y) = \frac{1}{h^d(n)} (\frac{1}{n} \sum_{i=1}^{n} I_{\{Y_i \leq y\}} \left[K(\frac{x - X_I}{h(n)}) - K(\frac{t_k - X_I}{h(n)}) \right]$$

$$-E(\frac{1}{n} \sum_{i=1}^{n} I_{\{Y_i \leq y\}} \left[K(\frac{x - X_I}{h(n)}) - K(\frac{t_k - X_I}{h(n)}) \right])$$

En tenant compte du fait que $I_{\{Y_i \leq y\}} \leq 1$ et que K est Lipschitzien, on a

$$\left| \hat{\triangle}_n(x, y) \right| \leq (nh^d(n))^{-1} \sum_{i=1}^{n} (\left| K(\frac{x - X_i}{h(n)}) - K(\frac{t_k - X_i}{h(n)}) \right| + E \; K(\frac{x - X_i}{h(n)}) - K(\frac{t_k - X_i}{h(n)})$$

$$\leq \; 2L \; h(n)^{-(d+\gamma)} \, \|x - t_k\|^\gamma$$

Puisue t_k a été choisi de manière à ce que la boule I_{t_k} contienne $x$alors:

$$\|x - t_k\| \leq h(n)^\alpha$$

ceci implique :

$$\left| \hat{\triangle}_n(x, y) \right| \leq 2L \; h(n)^{\alpha\gamma - (d+\gamma)}.$$

et cette majoration est indépendante de x et y. L'inégalité (I) et l'hypothèse: $(h(n)_{n\to\infty} \to 0)$ permettent de conclure que :

$$\sup_{x\in C}\sup_{y\in\mathbb{R}}\left|\hat{\triangle}_n(x,y)\right| \to 0, n\to\infty$$

convergence certaine. Il reste à démontrer:

$$\sup_{y\in\mathbb{R}}\max_{k=1,\ldots,l_n}\left|\triangle_n(t_k,y)\right| \overset{p.co.}{\to} 0 \quad (***)$$

$\forall\ \varepsilon > 0$, on pose :

$$B_n = \sup_{y\in\mathbb{R}} P\ (\max_{k=1,l_n}\left|\triangle_n(t_k,y)\right| > \varepsilon) \leq \sum \sup_{y\in\mathbb{R}} P(\max_{k=1,l_n}\left|\triangle_n(t_k,y)\right| > \varepsilon)$$

$$\leq l_n \sup_{x\in C}\sup_{y\in\mathbb{R}} P\ (\left|\triangle_n(t_k,y)\right| > \varepsilon)$$

d'où, en utilisant le lemme 4et les inégalités (*) et (I), on a:

$$B_n \leq 2Bh(n)^{-(d+1/\gamma)}\exp(-bnh^d(n))$$

comme

$$nh^d(n)_{n\to\infty} \ \to \infty$$

alors

$$(nh^d(n))^{-\zeta} \to 0, \zeta > 0 \text{ et } n\to\infty$$

on en déduit qu'il existe δ telle que:

$$(nh^d(n))^{-\xi} < \delta.$$

en posant

$$\xi\ = d + 1/\gamma$$

on obtien

$$B_n \leq 2B\delta n^{\xi - b\gamma(n)},\ \gamma(n) = nh^d(n)$$

et alors : La serie de terme génerale B_n est convergent $\sum_{n\geq 1} B_n < \infty$ et donc on a (***), ce qui achève démonstration. ■

Lemme 6:

Sous le hypothèses du lemme 5, on a:

$$\sup_{x\in C}\left|\hat{f}_n(x) - E(\hat{f}_n(x))\right|_{n\to\infty} \ \to_{pco} 0$$

Preuve: Il suffit de remplacer $1_{\{Yi \leq y\}}$ par 1 dans toutes les étapes de la démonstration. ■

Lemme 7: On suppose que $F(y/.)$ est lipschitzienne en x uniformément par rapport à y, si:

$$\int |x| \, K(x)dx < \infty,$$

alors:

$$\sup_{x \in C} \sup_{y \in \mathbb{R}} \left| E(F^{1,n}(x,y)) - F(y/x) \; E(\hat{f}_n(x)) \right|_{n \to \infty} \to 0$$

Preuve: $\forall \; x \in C$, $y \in \mathbb{R}$, on a :

$$E(F^{1,n}(x,y)) - F(y/x) \; E \; (\hat{f}_n(x)) =$$

$$E(\frac{1}{nh^d(n)} \sum_{i=1}^{n} I_{\{Yi \leq y\}} K(\frac{x - X_i}{h(n)})) - F(y/\; x) \frac{1}{nh^d(n)} E \sum_{i=1}^{n} K(\frac{x - X_i}{h(n)}) =$$

$$\frac{1}{n} \sum_{i=1}^{n} (E(h^{-d}(n)(I_{\{Yi \leq y\}} K(\frac{x - X_i}{h(n)})) - F(y/x)E(h^{-d}(n)K(\frac{x - X_i}{h(n)}))) =$$

$$\frac{1}{n} \sum_{i=1}^{n} E \left[I_{\{Yi \leq y\}} - F(y/x) \right] \; h^{-d}(n)K(\frac{x - X_i}{h(n)}) \qquad (**)$$

en utilisant un résultat bien connu sur l'espérance conditionnelle(Théorème de Blackwell : $E(Y) = EE(Y/X))$ où chaque opérateur E a un sens), on a :

$$(**) = \frac{1}{n} \sum_{i=1}^{n} E((E(I_{\{Yi \leq y\}}/X_i) - F(y/x)h^{-d}(n)K(\frac{x - X_i}{h(n)}))$$

d'après (1), on a:

$$E(I_{\{Yi \leq y\}}/X_i) = F(y/X_i),$$

donc

$$(**) \quad = \quad \frac{1}{n} \sum_{i=1}^{n} E((F(y/X_i) - F(y/x)h^{-d}(n)K(\frac{x - X_i}{h(n)})) \leq$$

$$\frac{1}{n} \sum_{i=1}^{n} E(L \quad | \quad x - X_i \mid h^{-d}(n)K(\frac{x - X_i}{h(n)})$$

On pose

$$t \;\; = \frac{x - X_i}{h(n)}$$

l'hypothèse (2) sur la loi marginale de X_i permet d'avoir :

$$(**)\frac{1}{n}\sum_{i=1}^{n} L\Gamma h^d(n)\int |t|\, K(t)dt = L\Gamma h^d(n)\int |t|\, K(t)\; dt$$

comme

$$\int |t|\, K(t)dt < \infty \text{ et } h(n) \to 0, n \to \infty$$

Alors:

$$(**) \to 0, n \to \infty.\blacksquare$$

Lemme 8: Sous l'hypothèse du lemme 5, on a :

$$\exists \delta > 0 \sum_{i=1}^{n} P(\inf_{x\in C} f_n(x) \leq \delta) < \infty$$

Preuve:

$$(*) \inf_{x\in C} f_n(x) \geq \inf_{x\in C} E(f_n(x)) - \sup_{x\in C} |f_n(x) - E(f_n(x))|$$

$\forall x \in C$, on a :

$$E(f_n(x) = (nh^d(n))^{-1}\sum_{i=1}^{n} E(K(\frac{x-X_i}{h(n)}) = n^{-1}\sum_{i=1}^{n}\int (nh^d(n))^{-1} K(\frac{x-t}{h(n)}) P_x(dt)$$

l'hypothèse (2) et la derniére partie de (3) permettent d'affirmer

$$\exists \tau > 0, \exists n_{\circ^{\cdot}} > 0 : E(f_n(x)) \geq \tau \int K(z)dz = \tau, \forall x \in C, \forall n \in \mathbb{N}$$

et $n \geq n$, on pose:

$$\delta = \tau/2$$

et on utilise (*) et le lemme 5 pour conclure. ■

Théorème 3:

Sous les hypothèses1), 2), 3)et U), si:
(i)$F(y/.\ )$ est lipschitzienne en x unifomément en y.
(ii) Le noyau Kest lipschitzien et vérifie $\int |u|\, K(u)du < \infty$.
(iii)La serie: $\sum_{n=0}^{\infty} \exp\,(-\lambda nh^d(n) < \infty,\ \lambda > 0$
Alors:

$$\sup_{x\in C} |m_n(x) - m(x)|_{n\to\infty} \to_{pco} 0$$

Preuve: C'est une conséquence du lemme 6 et des lemmes 8, 7, 6, 5 et 4. ■

Loi limite de l'estimateur m_n(x)

On se restreind à: $d = 1$ et : l'hypothèse(*)

$\forall\ y$, $F(y/.\ )$ est deux fois continuement dérivables dans un voisinage U de x tel que :

$$\sup_{x \in U} \sup_{y} \left| F''(y/x) \right| < \infty \qquad (*)$$

et

$$F'(m(x)/x) = \frac{\partial}{\partial y} F(y/x)_{m(x)} > 0 \qquad (*)$$

Théorème4: Sous l'hypothèse (*), si: le noyau K vérifie $\int uK(u)\ du = 0$ et, si :

$$nh^3(n) \to \infty \text{ et } nh^5(n) \to 0$$

quand $n \to \infty$,

Alors, pour presque tout x:

$$(nh(n)^{1/2}(m_n(x) - m(x)) \underset{n \to \infty}{\overset{L.}{\longrightarrow}} \quad N(0, \sigma^2)$$

où

$$\sigma^2 = \frac{1}{4} \int K^2(x)\ dx\ /(F'(m(x)/x))^2.$$

Preuve: voir Stute (1986 (b)).

Chapter 5

Etude de la Médiane conditionnelle: Cas Multidimensionnel

5.1 Résultats Théoriques

5.1.1 Définitions et Propriétés

Soient les hypothèses suivantes:

$(X,\ Y)$ un couple de vec. a. de $\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$, $P_{X,Y}$ la loi du couple$(X,\ Y)$ dont l'espace d'états est $(\mathbb{R}^{d_1+d_1},\ \mathcal{B}(\mathbb{R}^{d_1+d_1})\ )$, $E(\|Y\|)<+\infty$ où : $\|\cdot\|$ désignant la norme euclidienne de $\mathbb{R}^{d_2}$, et P_Y^X une version régulière de la probabilité conditionnelle de Y par rapport à X telle que:

$$\forall\ (x,\alpha)\ \in\ \mathbb{R}^{d_1}\times\mathbb{R}^{d_2} \text{ on a: } \phi_x(\alpha)=\int_{R^{d_2}}\|Y-\alpha\|\ P_Y^{X=x}\ (dy)\ <+\infty$$

Définition 1:
Une médiane conditionnelle $m(x)$ de$Y$ sachant X est telle que:
$\forall x\in\mathbb{R}^{d_1}$, $m(x)$ vérifie:

$$\phi_x(m(x))=\inf_{\alpha\in\mathbb{R}^{d_2}}\phi_x(\alpha)=\inf_{\alpha\in\mathbb{R}^{d_2}}\int_{R^{d2}}\|Y-\alpha\|\ P_Y^{X=x}\ (dy)$$

Propriétes Générales de $\phi_x(.\ )$:
Proposition 1: $\phi_x(\alpha)$ est lipschitzienne de rapport 1.

Preuve:

$$| \phi_x(\alpha) - \phi_x(\beta) |= \int_{\mathbb{R}^{d_2}} (\|y-\alpha\| - \|y-\beta\|)\ P_Y^{X=x}(dy$$

$$\leq \int_{\mathbb{R}^{d_2}} \|\alpha - \beta\|\ P_Y^{X=x}\ (dy) =$$

$\|\alpha - \beta\|$

Corollaire: $\phi_x(.\ )$ est continue

Preuve: Elle découle immédiatement de la proposition 1

Proposition 2:

$$\lim_{\|\alpha\| \to +\infty} \phi_x(\alpha)\ /\ \|\alpha\| = 1,$$

Preuve: Soit $\alpha(n) \in \mathbb{R}^{d_2}$avec $\|\alpha(n)\| \to +\infty, \quad n \to \infty$,

on a:

$$\frac{\phi_x(\alpha(n))}{\|\alpha(n)\|} - 1 = \int_{\mathbb{R}^{d_2}} \frac{\|y-\alpha(n)\| - \|\alpha(n)\|}{\|\alpha(n)\|} P_Y^{X=x}(dy)$$

comme

$$\|y - \alpha(n)\| - \|\alpha(n)\| \leq \|y\|$$

Alors:

$$\int_{\mathbb{R}^{d_2}} \frac{\|y-\alpha(n)\| - \|\alpha\ (n)\|}{\|\alpha(n)\|} P_Y^{X=x}(dy) \leq \int_{\mathbb{R}^{d_2}} \frac{\|y\|}{\|\alpha(n)\|} P_Y^{X=x}(dy)$$

Le derniér majorant tend vers 0 quand $n \to \infty$

car par hypothèse $E(\|Y\|) < \infty$. On a en plus:

$$\frac{\|y-\alpha(n)\| - \|\alpha(n)\|}{\|\alpha(n)\|} \leq \frac{\|y\|}{\inf \|\alpha(n)\|}$$

et on utilise le théorème de Lebesgue(convergence dominée)pour conclure. ■

Proposition 3: $\phi_x(.\ )$ est convexe

Preuve: Soit $(\lambda_1,\ \lambda_2) \in \mathbb{R}^+ \times \mathbb{R}^+$tels que: $(\lambda_1 + \lambda_2) = 1$, on a:

$$\int_{\mathbb{R}^{d_2}} \|y - \lambda_1\alpha - \lambda_2\beta\|\, P_Y^{X=x}(dy) \leq \int_{\mathbb{R}^{d_2}} \|(\lambda_1+\lambda_2)y - \lambda_1\alpha - \lambda_2\beta\|\, P_Y^{X=x}(dy)$$

$$= \int_{\mathbb{R}^{d_2}} \|\lambda_1(y-\alpha) + \lambda_2(y-\beta)\|\, P_Y^{X=x}(dy)$$

$$\leq \lambda_1 \int_{\mathbb{R}^{d_2}} \|y-\alpha\|\, P_Y^{X=x}(dy) + \lambda_2 \int_{\mathbb{R}^{d_2}} \|y-\beta\|\, P_Y^{X=x}(dy)$$

donc:

$$\phi_x(\lambda_1\alpha + \lambda_2\beta) \leq \lambda_1\phi_x(\alpha) + \lambda_2\phi_x(\beta).\ ■$$

5.1.2 Etude de L'existence et de L'unicité

Existence et Unicité de la Médiane conditionnelle:

A: Existence:

On pose: $h(x) = E(\|Y\| \ /X = x) < \ \infty$, $\phi_x^* \ = \ \inf \left\{\phi_x(\alpha), \ \alpha \in \mathbb{R}^{d_2}\right\}$ on a:

$$-\infty < \phi_x^* < \phi_x(0) = \ h(x)$$

Soit:

$$r = \sup \left\{\|\alpha\| : \phi_x(\alpha) \leq h(x)\right\}, \|\alpha\| > r \ \Rightarrow \phi_x(\alpha) > \phi_x^*$$

ceci implique que l'ensemble des médianes conditionnelles est un sous ensemble de $B(r)$où:

$$B(r) = \left\{\alpha \in \mathbb{R}^{d_2} \ : \|\alpha\| \leq r\right\},$$

donc:

$$\phi_x^* = \inf \left\{\phi_x(\alpha), \alpha \in B(r)\right\}$$

$B(r)$ est un compact on a montré (corollaire 1) que $\phi_x(.\ )$ est continue sur $\mathbb{R}^{d_2}$ et par conséquent $\phi_{x|B(x)}$ l'est aussi, elle atteint donc son minimum sur ce compact. ceci assure l'existence de la médiane conditionnelle. ■

Toute médiane conditionnelle vérifie la propriété suivante :

Poposition 4:

$$\|m(x)\| \leq 2 \int_{\mathbb{R}^{d_2}} \|y\| \, P_Y^{X=x}(dy)$$

Preuve: On a

$$\|m(x)\| = \int_{\mathbb{R}^{d_2}} \|m(x)\| \, P_Y^{X=x}(dy) \leq \int_{\mathbb{R}^{d_2}} \|y - m(x)\| \, P_Y^{X=x}(dy) + \int_{\mathbb{R}^{d_2}} \|y\| \, P_Y^{X=x}(dy)$$

$= \ \phi_x(m(x)) + \phi_x(0) \ \leq \ 2\phi_x(0 = \ \ 2 \int \|y\|$ $P_Y^{X=x}(dy)$. ■

B: Uncité de la médiane conditionnelle multidimensionnelle

Ceci résulte de la:

Proposition 5: Si le support de la loi conditionnelle n'est pas une droite, alors: la médiane conditionnelle est unique.

Preuve: $\mathbb{R}^{d_2}$muni de la norme euclidienne est strictement convexe et par conséquent : la fonction $\phi_x(.\ )$ l'est aussi, le théorème de Kemperman(1987) indique que l 'ensemble des médianes (conditionnelles) est au plus réduit à un point. ■

Remarque:

L' unicité peut être démontrée en utilisant le théorème de Milasevic Ducharme(1987) en remplaçant $\phi(\alpha)$ par $\phi_x(\alpha)$.

Estimation de la Médiane conditionnelle dans le cas Multivarié

5.2 Méthode du Médianogramme

5.2.1 Construction de l'estimateur:

Soit $(X_i,\ Y_i)_{i=1,n}$ une suite de vec. a. i. i. d. et de même loi que $(X,\ Y)$ (La loi étant inconnue)

$$J = \left\{ I_{n,z}\ z \in \mathbb{Z}^{d_1} \right\}$$

le pavage de $\mathbb{R}^{d_1}$ défini par: $z = (z_1,\ \ldots,\ z_{d_1}) \in\ \mathbb{Z}^{d_1}$ et

$$I_{n,z}\ =\ \prod_{k=1}^{d_1} [z_k\ h(n),\ (z_k+1)\ h(n)[$$

où $h(n)$ est un suite: $\lim_{n\to\infty} h(n) = 0$. $\forall\ x \in \mathbb{R}^{d_1}$, on désigne par $I_n(x)$ le pavé de J_n qui contient x, et par

$q_n(x)$ le nombre des $(X_i)_{i=1,n}$ qui appartiennent à $I_n(x)$.

On considère l'estimateur $(P_Y^{X\in I_{n\ (x)}})_n$ défini par :

$$(P_Y^{X\in I_{n\ (x)}})_n = \frac{\overset{n}{\sum\limits_{i=1}^{n}} \delta_{Y_i}.1_{X_i\ \in I_n(n)}}{\sum\limits_{i=1}^{n} 1_{X_{i\in I_n(x)}}} \text{ si } \sum_{i=1}^{n} 1_{X_{i\in I_n(x)}} \neq 0$$

$= 0$ sinon. Un estimateur de $\phi_x(\alpha)$ qu'on note: $\hat{\phi}_{n,x}(\alpha)$, est tel que:

$$\hat{\phi}_{n,x}(\alpha) = \int_{\mathbb{R}^{d_2}} \|Y-\alpha\|\ (P_Y^{X\in I_{N\ (x)}})_n\ (dy)$$

Soit:

$$\hat{\phi}_{n,x}(\alpha) = \frac{1}{\sum\limits_{i=1}^{n} 1_{I_n(x)}(X_i)} \sum_{i=1}^{n} \|Y_i-\alpha\| \times 1_{I_n(x)}(X_i) \text{ si } \sum_{i=1}^{n} 1_{I_n(x)}(X_i)\ \neq 0$$

0 sinon
Soit l'estimateur empirique de $m(x)$ qu'on note $\hat{m}_n(x)$est tel que

$$\hat{\phi}_{n,x}(\hat{m}_n(x)) = \inf_{\alpha} \hat{\phi}_{n,x}(\alpha)$$

5.2.2 Etude de la convergence

Etude de la Convergence simple de $m_n$vers $m(x)$

Soit $F(.\ /x)$ une version continue de la f. r. c. de Y sachant $X = x$:

$$F(y/\ x) = \int_{\{Y \leq y\}} P_Y^{X=x}\ (dy).$$

Notons: $\hat{F}_n(y\ /\ x)$ son estimateur empirique:

$$\hat{F}_n(y\ /\ x) = \int_{\{Y\ \leq\ y\}} (P_Y^{X \in I_n(x)_n})_n(dy)$$

Soit:

$$\hat{F}_n(y\ /\ x) = \frac{1}{\sum_{i=1}^{n} 1_{I_n(x)}(X_i)} \sum_{i=1}^{n} 1_{\{Y_i \leq y\}} 1_{I_n(x)}(X_i) \text{ si } \sum_{i=1}^{n} 1_{I_n(x)}(X_i) \neq 0 \text{ et } 0 \text{ sinon.}$$

Commençons par des lemmes techniques:

Lemme 1:

$$\forall\ \varepsilon > 0, P(\sup_{y \in \mathbb{R}^{d_2}} \left| \hat{F}_n(y)\ - F(y) \right|\ > \varepsilon) \leq 4e^{(4\varepsilon + 4\varepsilon^2)}(1+n^2)^{d_2} \exp -2n\varepsilon^2$$

Preuve: on pose:
$\hat{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,y]}(Y_i), \quad y = (y_1,\ ..\ ,\ y_{d_2})$ et $]-\infty, y] = \prod_{i=1}^{d_2}]-\infty, y_i]$
et on utilise Devroy (1982).

Lemme 2:

On suppose que la loi marginale de X a une densité g continue et strictement positive. Si:

$$\lim_{n\to\infty} nh(n)^{d_1} = \infty \text{ et } h_{n\to\infty}(n) \to 0$$

Alors :

$$\sup_{y \in \mathbb{R}^{d_2}} \mid \hat{F}_n(y\ /x)\ - F\ (y\ /\ x) \mid_{n\to\infty} \to^p 0.$$

i. e: pour x fixé, $\forall y \in \mathbb{R}^{d_2}$

$\hat{F}_n(\cdot/x)_{n\to\infty} \to^p F(\cdot/x)$ Soit: $\forall\varepsilon > 0$ on a: $P(d(\hat{F}_n(\cdot/x), F(\cdot/x)) > \varepsilon)_{n\to\infty} \to 0$

Preuve: Posons: $d(\hat{F}_n(\cdot/x),\ F(x/\cdot) = \sup_{y\in\mathbb{R}^{d_2}} \mid \hat{F}_n(y\ /x) - F\ (y\ /\ x) \mid$

. Soit $p(x,\ h(n))$ la probabilité pour que$(X,\ Y) \in I_n(x)$. en utilisant la formule de probabilité totale

$$P(d(\hat{F}(\cdot/x), F(\cdot/x) \quad > \quad \varepsilon) = \sum_{i=1}^{n} C_n^q\ [p(x, h(n))]^q\, [1 - p\ (x, h(n)]^{n-q}$$

$\times P(d(\hat{F}_{q(x,h(n))}(\cdot/x), F(\cdot/x)) \quad > \quad \varepsilon/\ q(x, h(n) = q)$

$q(x,\ h(n)) =$ le nombre de points X_i qui tombent dans $I_n(x)$

$P(d(\hat{F}_{q(x,h(n))}(\cdot/x), F(\cdot/x)) > \varepsilon/\ q(x, h(n) = q) = P(d(\hat{F}_q(\cdot/x), F(\cdot/x)) > \varepsilon)$.

l'utilisation du lemme 1 permet d'écrire: $P(d(\hat{F}_n(\cdot/x),\ F(\cdot/x)) > \varepsilon) =$

$$= 4e^{(4\varepsilon+4\varepsilon^2)} \sum_{i=1}^{n} C_n^q\ (1+q^2)^{d_2}\, [p(x, h(n))]^q \times\ [1 - p\ (x, h(n)]^{n-q}\ \exp -2q\varepsilon^2$$

$$\leq\ 4e^{(4\varepsilon+4\varepsilon^2)}(1+n^2)^{d_2} \sum_{i=1}^{n} C_n^q\ [p(x, h(n))]^q \times\ \ [1 - p(x, h(n)]^{n-q}\ \exp -2q\varepsilon^2$$

$$\leq\ 4e^{(4\varepsilon+4\varepsilon^2)}(1 + n^2)^{d_2} \sum_{i=1}^{n} C_n^q\ \left[p(x, h(n)) \exp -2\varepsilon^2\right]^q \times [1 - p\ (x, h(n)]^{n-q}$$

$$\leq\ 4e^{(4\varepsilon+4\varepsilon^2)}(1 + n^2)^{d_2} \left[1 - p(x, h(n))\ (1 - \exp -2\varepsilon^2)\right]^n$$

Le théorème de la moyenne et la condition sur la loi marginale de X permettent d'affirmer:

$\exists\xi \in I_n(x)$ tel que:

$$p(x, h(n)) = h(n)^{d_1} g(\xi).$$

comme:

$$\left[1 - p(x, h(n))\ (1 - \exp -2\varepsilon^2)\right]^n =\ \exp\ \left[n \log(1 - p(x, h(n))\ (1 - \exp -2\varepsilon^2))\right]$$

et pour $x\ <<\ 1$, on a:

$$\log(1 - x) \leq -x$$

donc:

$$n \log(1 - p(x, h(n))(1 - \exp -2\varepsilon^2) \leq -nh(n)^{d_1} g(\xi)(1 - \exp -2\varepsilon^2)\quad (*)$$

par ailleurs

$$1 - \exp -2\varepsilon^2 \leq 2\varepsilon^2,$$

Alors:

$$(*) \quad \leq -nh(n)^{d_1} g(\xi) 2\varepsilon^2$$

Donc:

$$P(d(\hat{F}_n(\cdot/x), F(\cdot/x)) > \varepsilon) \leq 4e^{(4\varepsilon+4\varepsilon^2)}(1+n^2)^{d_2} \exp -nh(n)^{d_1} g(\xi) 2\varepsilon^2 \qquad \text{(A)}$$

le deuxième membre de cette inégalité tend vers 0 quand $n \to \infty$. ■

Lemme 3:

Sous les hypothèses du lemme (2) si:

$\sum_{i=1}^{n} (1+n^2)^{d_2} \exp -\lambda nh(n)^{d_1} < \infty$ pour tout $\lambda > 0$, Alors pour x fixé:

$$\sup_{y \in \mathbb{R}^{d_2}} \left| \hat{F}_n(y/x) - F(y/x) \right|_{n \to \infty} \to^{p.co} 0$$

c-à-d: Pour x fixé, $\forall y \in \mathbb{R}^{d_2}$

$$\hat{F}_n(\cdot/x) \overset{p.co}{\to} F\ (\cdot/\ x)$$

Preuve: On pose: $\lambda = 2\varepsilon^2 g(\xi)$

on utilise (A) de la démonstration du lemme (2) pour obtenir

$$\sum_{n=1}^{\infty} P(d(\hat{F}_n(\cdot/x), F(\cdot/x)) > \varepsilon) \leq 4e^{(4\varepsilon+4\varepsilon^2)} \sum_{n=1}^{\infty} (1+n^2)^{d_2} \exp -nh(n)^{d_1} g(\xi) 2\varepsilon^2$$

la condition du lemme permet de finir la démonstration. ■

Kemperman a démontré le lemme suivant:

Lemme 4:

Soient P une mesure de probabilité sur $\mathbb{R}^{d_2}$ dont le support n'est pas une droite et $P_n$une suite de mesure de probabilité.

Si P_n converge faiblement vers P, alors: La médiane de la mesure P_n converge presque sûrement vers la médiane unique de P.

Preuve: voir Kemperman(1987).

Théorème 1:

On suppose que le support de la loi conditionnelle n'est pas une droite et que la loi marginale de X a une densité g strictement positive, si :

$$\sum_{n \geq 1} (1+n^2)^{d_2} \exp -(\lambda nh(n)^{d_1}) < \infty$$

Alors: $\forall\ x$ fixé

$$\hat{m}_n(x)_{n\to\infty} \overset{p.s.}{\to} m(x)$$

Preuve:

les lemmes 1, 2 et 3ont servi à démontrer la convergence faible de la loi conditionnelle empirique vers la loi conditionnelle. on utilise le lemme 4 pour finir la démonstration du théorème. ■

5.3 La Méthode du Noyau

5.3.1 Construction de l'estimateur

Soient $(X_i,\ Y_i)$ un échantillon de vec. a. i. i. d. de même loi que $(X,\ Y)$ et $\mathbb{K}$ un noyau strictement positif de $\mathbb{R}^{d_1}$ tel que: $\mathbb{K}(x) = \prod_{i=1}^{d_1} K(x_i)$, $x = (x_1,\ \ldots\ldots,\ x_{d_1})$, K étant un noyau de $\mathbb{R}$ vérifiant :

$|K(.)| < k_1\ <\ \infty,\ \ |z|\, K(z) \to 0,\ |z| \to \infty$ et $\int K(z)\, dz\ = 1.$

K est à support et à variation bornés.

$\forall x \in \mathbb{R}^{d_1}$, considérons l'estimateur $(P_Y^{X=x})_n$ défini par :

$$(P_Y^{X=x})_n = \frac{\sum\limits_{i=1}^{n} \delta_{Y_i}\ \mathbb{K}\ (\frac{x-X_i}{h(n)})}{\sum\limits_{i=1}^{n} \mathbb{K}\ (\frac{x-X_i}{h(n)})}$$

Un estimateur de $\phi_{n,x}$, de $\phi_x(\alpha)$ est défini par :

$$\phi_{n,x}(\alpha) = \int_{\mathbb{R}^{d_2}} \|y-\alpha\|\, (P_Y^{X=x})_n(dy) = \frac{\sum\limits_{i=1}^{n}\ \|Y_i-\alpha\|\quad \mathbb{K}\ (\frac{x-X_i}{h(n)})}{\sum\limits_{i=1}^{n} \mathbb{K}\ (\frac{x-X_i}{h(n)})}$$

L'estimateur de la médiane conditionnelle $m_n(x)$ est tel que:

$$\phi_{n,x}(m_n(x)) = \inf_{\alpha\in\mathbb{R}^{d_2}}\ \phi_{n,x}\ (\alpha)$$

5.3.2 Etude de la convergence

Convergence simple de $m_n\ (x)$vers $m(x)$

$\hat{F}_n(y/\ x)$ est un estimateur de la fonction répartition conditionnelle défini par :

$$\hat{F}_n(y/\ x) = \int_{\{Y\leq y\}}(P_Y^{X=x})_n(dy) = \frac{\sum\limits_{i=1}^{n} \ 1_{\{Y_i\leq y\}}\ \mathbb{K}\left(\frac{x-X_i}{h(n)}\right)}{\sum\limits_{i=1}^{n} \mathbb{K}\left(\frac{x-X_i}{h(n)}\right)}$$

Stute (1986 (b)) a démontré le lemme suivant:

Lemme 5: si:

$\sum\limits_{n=0}^{\infty} \exp\,(-\lambda nh(n)^{d_1}) < \infty$, pour $\lambda > 0$, alors pour x fixé on a:

$$F_n(\cdot\ /x)_{n\to\infty} \ \to^{p.s}\ F\ (\cdot/x)$$

Soit: $\sup\limits_{y\in\mathbb{R}^{d_2}} \mid F_n(y\ /x)\ - F\ (y/x)\ \mid \to^{p.s} 0$

Théorème 2:

On suppose que le support de la loi conditionnelle n'est pas une droite et que le noyau $\mathbb{K}$

est à support et à variation bornés. si:

$\sum\limits_{n=0}^{\infty} \exp\,(-\lambda nh(n)^{d_1}) < \infty$, pour $\lambda > 0$,

alors : Pour tout x fixé on a :

$$m_n(x)_{n\to\infty} \ \overset{p.s.}{\to}\ m(x)$$

Preuve: le lemme 5 permet de montrer que $\phi_{n,x}(.\ )$ converge $p.\ s.$ vers $\phi_x(.\ )$. on utilise le lemme 4 pour finir la démonstration. ■

Chapter 6

Application de l'Estimation Non Paramétrique à la Prédiction

Introduction.

les données: Soit $(Z_n)_n$ un processus strictement stationnaire, Markovien d'ordre k et α -mélangeant dont

l'éspace d'état est une partie mesurable E de $\mathbb{R}$. On considère le problème de la prédiction de Z_{n+1} à l'aide des variables observées $\{Z_i; i = 1, ..., n\}$, pour se faire en cherche à évaluer:

$$m(Z_{n-k+1}, ..., Z_n) = m(Z_{n+1}/ [Z_{n-k+1}, ..., Z_n])$$

avec m la médiane conditionnelle. Cette variable aléatoire réelle est le meilleur predicteur probabiliste de Z_{n+1}. dans le sens suivant (realisateur de l'inf au sens de L^1) c-à-d:

$$\|Z_{n+1} - m(Z_{n-k+1}, ..., Z_n)\|_{L^1} = \inf \left\{ \|Z_{n+1} - h\| , h \in L^1(\mathcal{B}(Z_{n-k+1}, ..., Z_n)) \right\}$$

On utilise deux predicteurs : Le Medianogramme et Le predicteur à noyau pour l'estimation de ce predicteur probabiliste.

6.1 Prédicteur: Le Médianogramme

Le Médianogramme: IL a la même conception que le regressogramme-estimateur de l'ésperance conditionnelle- et l'histogramme -éstimateur de la densité. On se place dans le cas ou le processus est à valeur dans $[0,1]$(ou à support compact). Pour n fixé soit $\mathbf{P}_n$ une partition de $[0,1]$. Un estimateur de la $f.\ r.\ c$ (fonction de répartition conditionnelle de Z_{n+1} sachant $Z_{n-k+1},\ \ldots,\ Z_n$est égale : à la moyenne des $1_{\{Z_{i+1}\leq y\}} i=1,\ \ldots,\ n-1$ tel que: le vecteur de $\mathbb{R}^k$: $(Z_{i-k+1},\ \ldots,\ Z_i)$ appartient à l'élement de $\mathbf{P}_n$ contenant $(Z_{n-k+1},\ \ldots,\ Z_n)$ et à 0 si aucun des vecteurs$(Z_{i-k+1},\ \ldots,\ Z_i)$ n'appartient pas a cet element de $\mathbf{P}_n$ ce qui se traduit par:

$$\hat{F}_n(y) = \frac{1}{q(n)} \sum_{i=1}^{n-k+1} 1_{\{\{Z_{i+1}\leq y\}\}} \times 1_{(Z_{i-k+1},\ldots,Z_i)\in J_n} \text{si} \quad q(n) \neq 0$$

et $\hat{F}_n(y) = 0$ sinon, avec $J_n \in \mathbf{P}_n$ tel que:

$$(Z_{n-k+1}, ..., Z_n) \in J_n$$

et

$q(n) = card\,\{Z_i : (Z_{i-k+1}, ..., Z_i) \in J_n, i = 1, ..; n-1\}$

$\hat{Z}_{n+1} = m_n(Z_{n-k+1},\ \ldots,\ Z_n) = \begin{cases} Z_{([iq(n)/2]+1)} \text{ si } q(n) \text{ est impair} \\ (Z_{(iq(n)/2)} + Z_{((iq(n)/2)+1)})/2 \text{ si } q(n)\text{est pair} \end{cases}$

(Z_i est tel que$(Z_{i-k+1},\ \ldots,\ Z_i) \in J_n,\ i = k,\ \ldots,\ n-1)$.

Pour toute partition $\mathbf{P}_n$ on désigne par $h(n)$ le plus petit réel positif tel que tout élément de $\mathbf{P}_n$ soit contenu dans un cube de côté $h(n)$, $h(n)$ vérifie l'hypothèse suivante:

$$\lim_{n\to\infty} h(n) = 0 \ \text{ et } \lim_{n\to\infty} n^k h(n) = \infty$$

Posons: $X_i = (Z_{i-k+1},\ \ldots,\ Z_i)$ et $Y_i = Z_{i+s}$

$(X_i,\ Y_i)$ est aussi un processus α-mélageant, voir Billingsly(1968), on suppose que $\alpha(n)$ vérifie avec $h(n)$:

$$\lim_{n\to\infty} nh(n)^k/(k(n)\log n) = \infty$$

où $k(n)$ est une suite entière croissante vérifiant

$$\lim_{n\to\infty} \sup n\alpha(k(n))^{\frac{2k(n)}{3n}} \ / \ k(n) \ < \infty, \text{avec } 1 \leq \ k(n) < n.$$

$\lim_{n\to\infty} k(n) = \infty$ et $\lim_{n\to\infty} k(n)/n = 0$

Corollaire:

On a:

$$\left|\hat{Z}_{n+1} - m(Z_{n-k+1}, ..., Z_n)\right| 1_{(z_{n-k+1},...,z_n) \in I^k} \rightarrow^{p.s} 0$$

Preuve: $(\hat{Z}_{n+1} - m(Z_{n-k+1}, \ . \ . \ . \ , \ Z_n)) =$
$|m_n(Z_{n-k+1}, ..., Z_n) - m(Z_{n-k+1}, ..., Z_n)| \leq \sup_{x \in I^k} |m_n(x) \ - m(x)|$. La dernière expression converge
vers 0 . ■

6.2 Etude du Modèle Paramétrique

Prédiction d'un Processus Stationnaire. Décomposition de Wold.

Soit $(X_t, t \in \mathbb{Z})$ un processus du second ordre $(\|X_t\|^2_{L^2(p)} = E(X_t^2) < +\infty)$; on suppose qu'on a observé les (X_s) jusqu'a l'instant t : $(X_s, s \leq t)$.

Le but: Prévoir $g(X_{t+h}) \in L^2(p)$ avec g une fonction de X_{t+h}(futur de X_t àl'horizon h).Pour cela on cherche la meilleure approximation au sens de $L^2(p)$(en $m.q$)de $g(X_{t+h})$ par une fonction des $(X_s, s \leq t)$ (i.e) on cherche à:

$$\text{minimiser } E(\varphi - g(X_{t+h})^2$$

où φ est réelle est fonction des $(X_s, s \leq t)$ avec φ de carré intégrable.

la solution est connue.

Elle est unique et c'est:

$$E(g(X_{t+h})/\mathcal{B}_t)$$

Espérance conditionnelle de $g(X_{t+h})$ sachant $\mathcal{B}_t = \sigma(X_s, \ s \leq t)$ tribu engendré par le passé de X_t i. e **La projection** orthogonale de $g(X_{t+h})$ sur $L^2(\mathcal{B}_t)$.

Remarque: On a $\varphi \in L^2(\Omega, \ \mathcal{B}_t, \ P) \subset L^2(\Omega, \ \mathcal{A}, \ P)$

Erreur de prédiction(Probabiliste):

Notée Δ:

$$\Delta = \|g(X_{t+h})\|^2_{L^2(p)} - \|E(g(X_{t+h})/\mathcal{B}_t)\|^2_{L^2(p)}$$

et

$$\Delta \ = E((g(X_{t+h}) - E(g(X_{t+h})/\mathcal{B}_t))^2$$

en appliquant la relation de Pythagor on obtient

$$\Delta = \|g(X_{t+h}) - E(g(X_{t+h})/\mathcal{B}_t)\|^2_{L^2(p)}$$

Remarque: Si le processus est strictement stationnaire. Alors Δ ne dépend pas de t et ne dépend que de l'horizon h (Plus h est loin, plus la prédiction est mauvaise).

Prédiction Linéaire:

Même données avec φ linéaire par rapport à $X_s,\ s \leq t$, et de carré intégrable, on désigne par $\mathcal{M}_t = e(X_s,\ s \leq t)$ espace fermé engendré par les $X_s,\ s \leq t$. D'aprés le théorème de la projection orthogonale:

La solution:

est la projection orthogonale de $g(X_{t+h})$sur $\mathcal{M}_t$, notée $\Pr^{\mathcal{M}_t}(g(X_{t+h})$.

Erreure de prédiction(probabiliste):

$\Delta_l = E((g(X_{t+h}) - \Pr^{\mathcal{M}_t}(g(X_{t+h}))^2 = \|g(X_{t+h})\|^2_{L^2(p)} - \left\|\Pr^{\mathcal{M}_t}(g(X_{t+h})\right\|^2_{L^2(p)}.$

Remarque

Si le processus X_t est faiblement stationnaire;alors Δ_l ne dépend pas de t, dépend uniquement de

l'horizon h, pour $h = 1$(Hannan(1970)):

$$\Delta_l = 2\pi \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda) d\lambda\right]$$

où f désigne la densité spectraledu processus X_t.

Décomposition de Wolde d'un processus faiblement stationnaire:

Définition:

Un processus faiblement stationnaire $(X_t,\ t \in \mathbb{Z})$est dit régulier, si:

Δ_l à l'horizon $h = 1$(la premiére valeure future du processus) est strictement positive.

Remarque:

Processus régulier $\Rightarrow \sigma^2 > 0$ (où σ^2désigne l'erreur)

Théorème:

Soit X_t un processus faiblement stationnaire, centré régulier, alors:

On a la la décomposition suivante :

$$X_t = \sum_{j=0}^{\infty} \lambda j U_{t-j} + V_t\ ,\ t \in \mathbb{Z}$$

V_t est appelé la partie réguliere de X_t.

où $\lambda_0 = 1$, $\sum_j \lambda_j^2 < \infty$ et U_t bruit blanc de variance σ^2.

$U_t \in \mathcal{M}_t$, $U_t \in \mathcal{M}_{t-1}^{\perp}$(orthogonale de $\mathcal{M}_{t-1} \subset \mathcal{M}_t$).

$V_t \in \bigcap_{j=0}^{\infty} \mathcal{M}_{t-j}$ et $U_t \perp V_s\ t,\ s \in \mathbb{Z}$ et cette décomposition est unique.

Processus Linéaire(modèle ARIMA)

Objectif:

Soit $(X_t,\ t \in \mathbb{Z})$ un processus linéaire, ayant observer $X_1,\ \ldots,\ X_t$ on veut prévoir X_{t+1}.

On présente **la démarche de Box et Jenkins**:

Elle a pour objectif de définir des modèles statisiques qui sont à la fois simple, ayant le minimum de paramètres à éstimer et un maximum de fidelité et d'adéquation. En effet: Cette méthode permet de faire des prévision à partir des series chronologiques sans disposer au préalable d'une spésification du processus à l'origine de la série, et la détermination de celui-ci est résolue vu l'existence d'une classe générale de modèles:

$$\mathcal{C} = \{AR\,[p]\,, MA\,[q]\,, ARMA\,[p,q]\,, ARIMA\,[p,q,d]\}$$

à laquelle peut se rattacher n'importe quelle serie chronologique.

D'ailleurs l'identification et la *mod*élisation fait partie de cette méthode.

Le principe:

Une démarche en quatre étapes:

1)L'identification:

On modelise $(X_t,\ t \in \mathbb{Z})$. Il sagit de choisir un ou plusieurs élements de la classe $\mathcal{C}$, les plus appropriés à la serie étudiée(i. e identifier p ou q ou d), ce choix est lié à la fonction d'autocorrélation ρ_k et la fonction d'autocorrélation partielle r_k. $k = 0,\ 1,\ 2, \ldots$

2)Lestimation:

On estime les paramètres du modèle identifié, p. ex pour un $ARMA(p,\ q)$ on a $(p+q+1)$ paramètres: En effet un $ARMA(p,\ q)$ avec $p,\ q$ connus s'ecrit sous la forme:

$$\Phi_p(B)X(t) = \Theta_q(B)U(t)$$

les paramètres à estimer sont $(\phi_1,\ \ldots,\ \phi_p,\ \theta_1,\ \ldots,\ \theta_q,\ \sigma^2)$. Pour les $AR(p)$ fonction liniéaire des paramètres on utlise la méthode des moindre carrés;on obtient des estimateurs trés proches de la méthode de la régression multiple.

3)La validation du diagnostique:

Vérifier si le modèle est adéquat avec la série donnée. Ce contrôle est basé principalement sur l'analyse résiduel. En effet, dans la mesure ou le modèle proposé est correcte, les résiduts doivent suivre un processus de bruit blanc. Les tests de signification des autocorrélations résiduelles permette de le vérifier parmi ces tests celui de Durbin-Waston et un test utilisant la stasistique de Box-Pierce, valable pour tout processus $ARIMA(p,\ q,\ d)$.

4)Prévision des valeurs futures: Construire un prédicteur(fonction des observations)c'est l'objet de notre préocupation.

Identification, Estimation, Prévision d'un processus $ARIMA(p,\ q,\ d)$

Introduction:

Un processus stationnaire peut en génerale être approché par un $ARMA(p,\ q)$et comme en réalité la plupart des séries chronologiques ne sont pas stationnaires, et les différence premiéres ou plus géneralement les différences d'orde d (noté$\nabla^d X_t = (I-B)^d X_t$) d'une séri chronologique rendent l'hypothése de stationnarité plus vraissemblable alors l'idée est de considérer la classe des processus dont la différence d'un certain ordre est un $ARMA(p,\ q)$.

Définition:

Soit $(X_t,\ t\in\mathbb{Z})$ un processus du second d'ordre, on dit qu'il un $ARIMA(p,\ q,\ d)$ ou un $ARMA$ intégré d'ordre d s'il s'ecrit sous la forme:

$$\Phi_p(B)(I-B)^d X(t) = \Theta_q(B)U(t) \quad (*)$$

avec B opérateur de retard ($B^k X_t = X_{t-k}$)et Φ_p polynôme de degré p, Θ_q polynôme de degré q ($\phi_p \neq 0,\ \theta_q \neq 0$)et$\phi_0 = \theta_0 = 1$avec les racines de$\Phi$ etde Θ de module supérieur à 1.

Remarque: Pour compléter la définition(*), il, faut introduire le mécanisme de démarage du processus

Exemple:

Soit le cas particulier suivant(marche aléatoire):

$$X_t = U_1 + ... + U_t \quad t \geq 1$$

avec U_t des bruits blanc X_t est un $ARIMA(0,\ 0,\ 1)$, en effet: $X_t - X_{t-1} = U_t$ se met sous la forme:

$$\Phi_0(B)(I-B)^1 X(t) = \Theta_0(B)U(t)$$

Pour compléter la définition de $X_t$on introduit une valeur initiale du processus qu'on note $X_{-S+1}(S\in\mathbb{N})$ X_t peut alors s'écrir:

$$X_t = X_{-S+1} + \sum_{i=0}^{t+S-2} U_{t-i}$$

si $t \geq -S+2$;plus géneralement pour un $ARIMA(p,\ q,\ d)$ on se donne $p' = p+d$ valeurs initiales: $x_{-S+1},\ \ldots,\ x_{-S+p}$qu'on suppose non aléatoires; la définition est valable pour $t > -S+p'$ avec $U_t = 0\ \forall t \leq -S+p'$ par conséquent la relation (*) aura atteint son régime de croisiére au sens où toutes les $U_t$qui interviennent sont tous non nul à partir de la date $-S + p' + q + 1$.

Identification:

1) Pour identifier d, la méthode est la suivante.

On commence par calculer:

$$\hat{\rho}_k = \sum_{j=1}^{n-k}(X_j - \bar{X}) / \sum_{j=1}^{n}(X_j - \bar{X})$$

(Empirique basé sur les observations) Estimateur de ρ_k fonction d'autocorrélation $\rho_k = \frac{\gamma_k}{\gamma_0}$

avec γ_k fonction d'autocovarince. Si on constate que :

$\hat{\rho}_k \nrightarrow 0$ on calcule : $X_t - X_{t-1}$(on différentie le processus) puis on continue.

2) Pour identifier $p,\ q$, on calcule $\hat{r}_k$(estimateur de r_k fonction d'autocorrélation partielle)définie comme suit:

$$r_k = \rho(X_t - \tilde{X}_t, X_{t-k} - \tilde{X}_{t-k})$$

avec $\tilde{X}_t$: projection orthogonale des X_t sur $e = e(X_{t-1},\ .\ .\ .\ ,\ X_{t-k+1})$ l'éspace fermé engendré par le passé de X_t et $\tilde{X}_{t-k}$: projection orthogonale deX_{t-k} sur e (On a toujours $r_1 = \rho_1$), on a:

$$\hat{r}_k = \varphi_k(\hat{\rho}_1, ..., \hat{\rho}_k)$$

Plusieurs possibilités à envisagées. Le principe empirique est le suivant. Si on constate que:

$\hat{\rho}_k$ est trés petit à partir de $k > q$ on a affaire à $MA\,[q]$.

Si $\hat{r}_k$ est trés petit à partir de $k > p$ on a affaire à un $AR\,[p]$;

et si on constate que $\hat{\rho}_k$ et $\hat{r}_k$ petits on a affaire à un $ARMA\,[p, q]$.

Pour déterminer p et q on utilise une méthode dite méthode du coin décrite dans Gourieroux, Monfort(1980), en cas ou il ya plusieurs coins on applique le principe de parcimonie qui consiste au choix de $(p,\ q)$ qui minimise une certaine fonction de p et q soit p. ex $p + q$.

Estimation:

L'idée génerale est la méthode: Estimateur du maximum de vraissemblance(EMV), un des avantages de cette méthode elle permet d'obtenir des estimateurs asymptotiquement optimale (la taille est grande et l'erreur quadratique petite). Par contre EMV n'est pas robuste(Si le modèle est légerment faux l'estimateur peut devenir mauvais).

1: a. Cas : $MA\,[q]$ **Gaussien**:

$$X_t = \sum_{j=1}^{q} \psi_j U_{t-j}$$

avec U_t bruit blanc Gaussien $U_t \sim N(0,\ \sigma^2)$ si on observe $X = (X_1,\ \ldots,\ X_n)$ et si on considére $(U_{1-q},\ \ldots,\ U_0,\ U_1,\ \ldots U_n)$ sa loi est $N(0,\ \sigma^2)^{\bigotimes(n+q)}$ et comme $X = l(U)$ avec l fonction linéaire, on déduit la loi de X et ainsi sa matrice de variance covariance.

b. cas: $MA(q)$ **non Gaussien**.

Box et Jenkins ont proposé une méthode d'estimation non linéaire par les moindres carrés(vu la non linéarité des paramètres dans les $MA(q)$)

2: a. cas: $ARMA(p,\ q)$ **Gaussien:**

$\Phi_p(B)X(t) = \Theta_q(B)U(t)$. on écrit

$$X(t) = \Phi^{-1}(B)\Theta_q(B)U(t) = \sum_{j=0}^{\infty} \psi_j U_{t-j} X_t$$

est une transformation linéaire des U_t du type $MA(\infty)$ donc on se ramène à $MA(Q)$et on choisit un $ARMA(p,\ q)$ qui ressemble à $MA(Q)$;ainsi

$$X(t) \cong \sum_{j=1}^{Q} \psi_j U_{t-j}$$

b. cas: $ARMA(p,\ q)$ **non Gaussien:**

La premiére méthode consiste à le considérer comme s'il est Gaussien;

la deuxiéme méthode est celle du maximum de vraissemblance conditionnelle(EMVC); on se place dans le cas d'un $AR(p)$ sinon on fait comme s'il était en effet : supposons $X(t) = \sum_{j=1}^{p} \pi_j X_{t-j} + U_t$ avec les hypothéses usuelles: $N = n + p$ observations $(X_{1-p},\ \ldots,\ X_0,\ X_1,\ \ldots,\ X_n)$.

La densité est incalculable! On considére alors le vecteur: $(\underbrace{X_{1-p}, ..., X_0}, \underbrace{U_1, ...U_n})$sa densité $g(x_{1-p},\ \ldots,\ x_0,\ u_1,\ \ldots,\ u_n)$ est $f(x_{1-p},\ \ldots,\ x_0) \times (\frac{1}{\sigma^2\sqrt{2\pi}})^n \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{n} u_t\right]$, par le changement de variables suivants:

$$\left\{\begin{array}{c} x_{1-p} = x'_{1-p} \\ . \\ x_0 = x'_0 \\ . \\ u_t = x_t - \sum_{j=1}^{p} \pi_j x_{t-j} \end{array}\right\}$$

Le Jacobien de la transformation est égalle à 1. On a:

$$g(x_{1-p},...,x_0,u_1,...,u_n) = f(x_{1-p},...,x_0)f(u_1,...,u_n)$$

On a l'indépendance car des U_t sont des innovations de la forme

$$X_t - X_t^*$$

ils sont orthogonales à $\mathcal{M}_{t-1}$espace fermé engendré par le passé de $X_{t-1:}e(X_s,\ s \leq t-1)$,

et **dans le cas Gaussien on a**: **faiblement stationnaire** $\Leftrightarrow$**stictement stationnaire.**

La densité de $(X_{1-p},\ \ldots,\ X_0,\ X_1,\ \ldots,\ X_n)$ est:

$$f(x_{1-p},...,x_0)(\frac{1}{\sigma^2\sqrt{2\pi 1}})^n \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(x_t - \sum_{j=1}^{p}\pi_j x_{t-j})\right]$$

Si on cherche la densité de $X_1,\ \ldots,\ X_n/X_{1-p},\ \ldots,\ X_0$, on a:

$$f(x_1,...;x_n/x_{1-p},...,x_0) = (\frac{1}{\sigma^2\sqrt{2\pi 1}})^n \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(x_t - \sum_{j=1}^{p}\pi_j x_{t-j})\right]$$

Estimer $(\pi_1,\ \ldots,\ \pi_p,\ \sigma)$ par la méthode MV, on obtient le système le deux équations suivants:

$$\sum_{t=1}^{n} X_t X_{t-j} - \sum_{t=1}^{p}\hat{\pi}_i \sum_{t=1}^{n} X_{t-i} = 0 \quad \text{où} \quad j = 1,...,p$$

et:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{\pi}_1 X_{t-1} - .... - \hat{\pi}_p X_{t-p})$$

Remarque: Les estimateurs obtenus sont presque du MV.

6.3 Remarque Générale:

Les méthodes non paramétriques s'adaptent mieux à la prévision que les méthodes paramétriques par leurs simplicité de calcul : On a pas besoin d'evaluer des densités conditionnelles souvent difficiles à évaluer,et par leurs rigueur: On a pas besoin de l'étape d'identification du modèle comme dans le cas paramétrique, donc moins de doute.

Par contre un des inconvénient, on a des estimateurs qui ne sont pas stricts (tous les estimateurs cités sont des fonctions en éscaliers) et toute l'estimation se fait asymptotique d'où le choix de la fenêtre et du noyau qui se pose; quoique pour une certaine classe de noyau de convolution le problème de l'optimalité est réglé par le choix du noyau d'Epanechnikov et de Konakov.

Conclusions:.

Ce travail a été consacré à l'étude de la convergence des estimateurs des paramètres fonctionnels à savoir la densité conditionnelle, l'espérance conditionnelle, la médiane et la médiane conditionnelle en utilisant:

– Les propriétés des paramètres.

– Les résultats aquis.

– En s'inspirant de l'étude concernant l'estimation NP de l'espérance conditionnelle(synthèse de plusieurs articles).

On a pu aboutir à des résultats sur les différents modes de convergence (consistence faible et fortes) en suivant les mêmes étapes de démonstration utilisées pour l'estimation NP de l'espérance conditionnelle.

L'utilisation des données de simulation de Kaigh (1983) a permis de confirmer certains résultats théorique, concernant la performance des estimateurs de la médiane.

On pense qu'en utilisant les deux prédicteurs celui de l'espérance conditionnelle et de la médiane conditionnelle simultanément sur les mêmes données (séries réelles ou simulées) on peut avoir meilleurs résultats dans la prévision.

cela se justifie par le fait que le paramètre médiane conditionnelle est peut influencé par les valeurs abérantes relativement à l'espérance conditionnelle, des simulations sont souhaitables pour le vérifier.

A souligner que les méthodes NP sont mieux adaptées que les méthode paramétriques dans la prédiction par leurs facilités dans la pratique.

Outre toute l'étude est faite dans le cas d'un échantillon (i. i. d), et comme l'hypothèse de l'indépendance des variables statistiques est difficile à satisfaire dans la pratique, une étude dans le cas des processus est entrain de se faire avec l'hypothèse de mélangeance dans toutes ses formes, on a entamé le cas fortement mélangeant reste les autres cas.

Entre autre, parmi mes préocupations, c'est de refaire l'étude dans le cas d'un espace de dimension infini: l'estimation NP avec les observations à valeurs dans un espace de Banach tel que: $(C\,[0,1], \|\|_u\ )$ et ainsi retrouver les résultats sur les processus (en utilisant les notion de probabilités dans le espaces de Banach)

Il est à préciser que l'utilisation de la compacité de la boule unité de $\mathbb{R}^s$ n'est plus valable dans un tel cas car la boule unité de$C\left[0,1\right]$ n'est pas compact; une méthode d'estimation est adaptée dans ce cas, c'est la méthode par projection, qui consiste à estimer les coefficient de Fourrier du paramètre projeté.

Annexe:

Intégrale au sens de Bochner

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé ,$\mathcal{X}$ un espace vectoriel normé et $\mathcal{B}$ sa tribu borélienne.

Définition de l'intégale au sens de Bochner :

une $\mathcal{X}-$ $v.$ $a.$ X admet une espérance mathématique, notée $E(X) = \int_\Omega X dP$, au sens de Bochner, si X est intégrable au sens de Bochner (ou fortement intégrable) ce qui est équivalent à dire que X est fortement mesurable et que $\|X\|$ est intégrable au sens de Lebesgue ($E\|X\| = \int_\Omega \|X\| \, dP < \infty$) .

L'intégrale de Bochner d'une fonction étagée

$$X = \sum_{i \in I} x_i 1_{A_i}$$

$(\forall i \in I),\ x_i \in \mathcal{X},\ A_i \in \mathcal{A}$ sur tout ensemble $A \in \mathcal{A}$ étant définie par:

$$E(X) = \int_A X dP = \sum_{i \in l} x_i P(A \cap A_i)$$

Appelons $L^l(\Omega,\ \mathcal{A},\ P)$ l'espace vectoriel des classes de $\mathcal{X}$ $-v.$ $a.$ fortement mesurables (deux fonctions étant équivalentes si elles ne différent que sur ensemble de probabilité nulle) et Bochner - intégrables.

L'espace vectoriel des classes de fonctions étagées est dense dans L^l :

Soit X une $\mathcal{X}$ $-v.$ a, il existe une suite de $\mathcal{X} - v.$ $a.$ étagées (X_n) telle que: $X_n = \sum_{i_n \in I_n} x_{i_n} 1_{A_{i_n}}$, $(\forall i_n \in I_n)$, $x_{i_n} \in \mathcal{X}$, $A_{i_n} \in \mathcal{A}$ et $X_n \longrightarrow^{p.s.} X$

X sera Bochner intégrable si : $\lim_{n \to \infty} \int_\Omega \|X_n - X\| \, dP = 0$, et nous aurons par définition :

$(\forall A \in \mathcal{A})$, $\int_A X dP = \lim_{n \to \infty} \int_A X_n dP$.

Si $X \in L^l$, nous avons les principaux résultats suivants : $\|E(X)\| \leq E\|X\|$;

Si T est une application linéaire continue de $\mathcal{X}$ dans un $e.$ $v.$ normé $\mathcal{Y}$, alors :

$T(X)$ est Bochner -intégrable avec $E[T(X)] = T[E(X)]$; si $Y \in L^l$:

$(\forall a$ et $b \in \mathbb{R})$, $E(aX + bY) = aE(X) + bE(Y)$.

Si $\mathcal{X}$ est un espace de Banach, alors L^l est aussi un espace de Banach, muni de la norme $\|X(.)\| = \int_\Omega \|X(\omega)\| \, P(d\omega)$.

Inégalité de Holder :

Si pour certain $p \geq 1$, $E\|X\|^P < \infty$ alors : $\|E(X)\| \leq (E\|X\|^P)^{1/P}$.

Définition de l'intégrale au sens de Pettis :

une $\mathcal{X}. -$ $v.a$ X admet une espérance mathématique, notée $E(X)$, au sens de Pettis, si X est intégrable au sens de Pettis, ce qui signifie pour tout $x^* \in \mathcal{X}^*$(dual de $\mathcal{X}$) $x^*(X)$ est intégrable au sens de Lebesgue, on écrit:

$E(X) = \int_\Omega XdP$, où $E(X)$ est l'élément (unique s'il existe) de X tel que :

$(\forall x^* \in \mathcal{X}^*),\ E[x^*(X)] = x^*[E(X)]$

$i.\ e.$ $\quad x^*[E(X)] = \int_\Omega x^*(X)dP$ (au sens de Lebesgue).

par exemple si: $\mathcal{X} = C[0,1]$, $\mathcal{X}^* = \mathcal{M}_+[0,1]$: l'ensemble des mesures positives de $[0,1]$

Comme l'intégrale de Bochner, l'intégrale de Pettis est linéaire et commute avec toute application linéaire continue , si $E(X)$ et $E(Y)$ existent (X et Y étant des $\mathcal{X} -$. $v.\ a$), nous avons en effet les principaux résultats suivants :

-$(\forall a$ et $b \in \mathbb{R}),\ E(aX + bY) = aE(X) + bE(Y)$;

Si T est une application linéaire continue de$\mathcal{X}$ dans un $e.\ v.$ normé $\mathcal{Y}$, alors :

- $E[T(X)] = T[E(X)]$

- $\|E(X)\| \leq E\|X\|$ (mais il se peut que $E\|X\| = +\infty$).

dans le cas où $\mathcal{X}$ est un espace de Banach, toute $\mathcal{X} - v.\ a.$ l'intégrable au sens de Bochner est l'intégrable au sens de Pettis et les deux intégrales ont la mème valeur . Par contre , si $\mathcal{X}$ n'est pas complet , il est possible de trouver des contre -exemples où $E\|X\| < \infty$ et où $E(X)$, au sens de Pettis , n'existe pas .

Même dans le cas où $\mathcal{X}$ est un espace de Banach , la réciproque n'est pas vraie : il existe des $\mathcal{X}$ -v.a. intégrables au sens de Pettis mais pas au sens de Bochner .lorsque nous ne préciserons pas $E(X)$ désignera l'espérance mathématique au sens de Pettis

si X est intégrable , on appelle variance de X:

$V(X) = \int_\Omega \|X - E(X)\|^2 dP.$

$E(X)$ existe et que $(\forall x^* \in \mathcal{X}^*)$, $E(x^*[X - E(X)])^2 < \infty$, alors la forme bilinéaire symétrique, non négative, définie sur $\mathcal{X}^*$ par :

$$(covX)(x_1^*, x_2^*) = E(x_1^*[X - E(X)])(x_2^*[X - E(X)])$$

est appelée la covariance de X.

X et Ysont indépendantes: cov X+cov Y = cov $(X + Y)$.

Théorème :

Soit $X_1, ..., X_n$ des $\mathcal{X}$-v.a. ($\mathcal{X}$ Banach) indépendantes, symétriques .

Posons $S_k = \sum_{i=l}^k X_i$.

Alors , $\forall t > 0 : P\{\max_{l \leq k \leq n} \|S_k\| > t\} \leq 2P\{\|S_n\| > t\}$, et $P\left\{\max_{l\leq k\leq n}\|X_k\|>t\right\}\leq 2P\left\{\|S_n\|>t\right\}.$

Espérance conditionnelle :

soit $X\in L^l$ et $\mathcal{C}$ une sous tribu de $\mathcal{A}$. il existe une $\mathcal{X}$-v.a. Z, $\mathcal{C}$-mesurable , intégrable , telle que :

$\forall A\in\mathcal{C}\quad \int_A ZdP=\int_A XdP;$

Z est unique , à une équivalence près (pour la relation d'équivalence d'égalité presque sûre) et sa classe d'équivalence est appelée espérance mathémätique conditionnelle de X en $\mathcal{C}$, et notée :

$E(X/\mathcal{C})$ ou $E^{\mathcal{C}}(X)$.

l'application de L^l dans $L^l(\Omega,\mathcal{C},P;\mathcal{A})$ qui à X fait correspondre $E(X/\mathcal{C})$, est linéaire et a comme

propriétés : $(\forall X\in L^l),(\forall h\in L^\infty(\Omega,\mathcal{C},P;X)),\ E(h\circ X/\mathcal{C})=h\circ E(X/\mathcal{C})$, où L^l est l'espace vectoriel des classes de $\mathcal{X}$-v.a. essentiellement bornées ; $(\forall X\in L^l)$, $\|E(X/\mathcal{C})\|\leq\|X\|$.

Bibliographies

[1]. Bartlett M. S. (1963): Statistical estimation of density functions. *Sankya, A**25**, 245-254.*

[2] . Benedetti, J. (1975): Kernel estimation of regression function. *U. C. L. A. 405-408.*

[3]. Bhattacharya, P. K et Parthasarthy, K. R. (1961): Some limit theorems in regression theory. *Sankya, Ser. A **23** 91-102.*

[4]. Bickel, P. J et Lehman, E. L. (1969): Unbiased estimation in convex families. *Ann. Math. Statist. 40, 1523-1525.*

[5]. Billingsley, P. (1968): *Convergence of Probabilitey Measures.* Wiley (1968)

[6]. Bochner, S. (1955): *Harmonic Analysis and the Theory of Probability.* University of california Press.

[7] . Borisov, I. S. (1980): Abstracts of the colloquium on non parametric statistical inference*(Budapest).*

[8]. Bosq, D. (1969): Estimation de la densite conditionnelle et de la régression. *C. R. A. S. Paris, t. 269 p 661-664*

[9]. Bosq, D. (1973): Estimation de la densité d'un processus stationnaire mélangeant. *C. R. A. S. Paris, t 277.*

[10] . Bosq, D. (1989): L'estimation et prévision non paramétrique d'un processus non stationnaire. *C. R. A. S. Paris(1989)*

[11]. Cacoulos, T. (1966): Estimation of a multivariate density. *Ann. Ins. Statist. Math. **18**, 179-189.*

[12]. Carbon, M. (1983): Inegalité de Bernstein pour les Processus Fortement mélangeants. *C. R. A. S. Paris, t. 297.*

[13]. Cencov, N. N(1962): Evaluation of an unknown distribution density from observations. *Sov. Math. Vol.* ***3****, 1559-1562.*

[14]. Collomb, G. (1976): Estimation NP de la régresion par la méthode du noyau. *Toulouse.*

[15]. Collomb, G(1979): Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la régression. *C. R. A. S. PARIS A,* ***288****, 161-163.*

[16]. Deheuvel, P(1974): Conditions nécessaires et suffisantes de convergence ponctuelle presque sûr et uniforme presque sûr des estimateurs de la densité. *C. R. A. S. Paris. Vol.* ***278,*** *1217-1220.*

[17]. Deheuvels, P. (1977): Estimation non paramétrique de la densité par les histogrammes généralisés. *Rev. Statist. Appl.* ***35****, 5-42.*

[18]. Devroye, L. P(1978): Consistency of a sequential nearest neighbour regression function estimate. *J. M. Analysis*

[19]. Devroye, L. P(1978a): The uniforme convergence of nearest neighbour regression function estimators and their application in optimisation. *I. E. E. E. Trans. Info. Theory,* ***24,*** *142-151.*

[20]. Devroye, L. P(1982): Bound of the uniforme deviation of empirical measures. *J. M Analysis.*

[21]. Devroye, L et Gyôrfi, L(1985): *Non parametrique Density Estimation. The L^1view.* Wiley. 1985.

[22]. Ducharme, G et Milasevic, P. (1987): Uniqueness of the spatial median. *Ann. of Stat. , Vol.* ***15****, N°***3***, 1332-1333*

[23]. Duhamel, C(1978): Estimation d'une courbe de régression de la moyenne. *Revue de Statistique et d'analyse des données, 47-52.*

[24]. Dvoretzky, A. Kiefer, JetWolfitz, J. (1956): Asymptotique minimax character of the sample distribution function and the classical multinomial estimator. . *Ann. Math. Statist. Vol.* ***33****, 642-669.*

[25]. Epanechnikov, V. A. (1969): Non parametric estimation of a multidimensional probability density.

Theory Probab. Appl. ***14****, 153-158.*

[26]. Feller, W. (1968): *An introduction to Probability Theory and its Application*, 2nd. Wiley.

[27]. Ferguson, T. (1973): A bayesian analysis of some non parametric problems. *Ann. Statist.* ***1, 2****, 209-230.*

[28]. Fix, E et Hodges, J. L, (1951): Discriminatory analysis, non paramétric discrimination, consistency properties. *Texas, Report n°***4**

[29]. Geffroy, J. (1973): Sur la convergence uniforme des estimateurs d'une densité deprobabilité. *I. S. U. P. Paris6.*

[30]. Geffroy, J. (1975): Sur l'estimation de la densité dans un éspace métrique. *C. R. A. S. Paris, t. 278.*

[31]. Geffroy, J. (1980): Etude de la convergence du Régressogramme. *I. S. U. P. Jussieux, Paris6.*

[32]. Gourieroux, C. Monfort, A. (1983): *Cours de series temporelles.* Economica(1983)

[33] . Grenander, Ulf. (1981): *Abstract inference.* Wiley(1981)

[34]. Haldane, J. B. S. (1948): A note on the Median of a Multivariate Distribution. *Biometrika Vol.* ***35****, 414-415.*

[35]. Hall, P. (1984): Integrated square error properties of kernel estimators of regression functions. *Annals ofstatistics, 1984, Vol* **2, n°1**, *241-260.*

[36]. Hannan, E. J: *Multiple time series.* Wiley(1970)

[37]. Hardle, W. et Tsybakov, A. B(1988): Robust non parametrique regression with simultaneous scale curve estimation. *Ann of Stat Vol***16**

[38] . Hoeffding, W. (1963): Probability inequalities for sums of bounded random variables. *J. am. stat. assoc,* ***48,*** *p 13-30.*

[39] . Hoffmann-Jorgensen, J. (1976): *Probability in Banach spaces.* Saint-Flour. (1976)

[40]. Isogai, T. (1985): Some Extension of Haldane Multivariate Median and its Application. *Ann. Ins. Stat. Math. Vol.* ***37****, Part. A, 289-301.*

[41] .Kaigh, W. D. (1983) Quantiles Interval estimation. *Comm. Stat. Theo. Math. Vol. 12; N 21; 2427-2443.*

[42] Kempermann, J. H. D. (1987): The median of finite measure on Banach space. *Y. Rodge Ed. North Holland Amestrdam.*

[43] Komlos, J. Major, P. et Tusnady, G. (1975): An approximation of partial sums of independent r. v. and the sample d. f. *I. Z. Wahr. verw. geb.* ***32****, 111-131. .*

[44]Konakov, V. D. (1973): Asymptotic properties of some functions of non-parametric estimate of density function. *J. Multi. Anal.* ***5****, 454-468.*

[45] Lecoutre, J. P. (1975)(1983): Convergence et optimisation de certains estimateurs des densités de probabilité. *I. S. U. P. Paris 6.*

[46] . Major, P. (1973): On non-parametric estimation of the regression function. *Studia Scientiarum Mathematicarum Hungarica,* ***8,*** *347-361*

[47] Mood, A. M. (1941): On the joint Distribution of the Median in Sample from Multivariate Population. *Ann. Math. Stat. Vol.* ***12****, 268-278*

[48] Nadaraya, E. A(1964): On estimating regression. *TheoryProb. Applic. Vol.* ***9****, 141-142.*

[49]. Nadaraya. E. A(1965): On non-parametrique estimates of density functions and regression curves. *Theory. Prob. Applic. Vol.* ***10****, 186-190.*

[50] Nadaraya. E. A(1970): Remarks on non-paramrtric estimates for density functions and regression curves. *Theory. Prob. Applic. Vol.* ***15,*** *134-137*

[51] Noda, K. (1976): Estimation of a regression function by the Parzen kernel-type density estimators. *Ann. Inst. Math. Statist.* ***28,*** *221-234.*

[52] Parzen, E(1962): On estimate of probability density function and mode. *Ann. Math. Stat, Vol.* ***33,*** *1065-1076.*

[53] Pearson, K. et Lee, A. (1903): On the laws of inheritance in a man. *Biometrika* ***2,*** *357-462.*

[54] Prakaza Rao, BLS. (1983): *Non parametric functional Estimation.* Academic Press. 1983.

[55] Rousas, G. (1987): Non parametrique Estimation in mixing sequences of random variables. *U. C. L. A. (1987).*

[56] Rosenblatt, M. (1956): Remarks on some non -parametric estimates of a density function. *Ann. Math. Statist. Vol.* ***27,*** *642-669.*

[57] Rosenblatt, M. (1969): Conditional probability density and regression estimators. *In Multivariate Analysis* ***II,*** *pp. 25-31. New York: Academic Press.*

[58] Rosenblatt, M. (1971): Curve estimates. *Ann. Math. Statist.* ***42,*** *1815-1842.*

[59] Sabry, H. (1978): Sur l'estimation non-paramétrique des fonctions de régression. *C. R. A. S. Paris 286A, 941-944.*

[60]Schlee, W. (1982): Estimation Non paramétrique du α-quantile conditionnel. *Stat. et analyse des données Vol.* ***1,*** *32-47.*

[61]. Schuster, E. F. (1969): Estimation of probability density function an its derivatives. *Ann. Math. Stat. 40p. 1187-1195.*

[62]Schuster, E. F. (1972): Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Stat. Vol.* ***43,*** *84-88.*

[63].Shorack, G. R. et Wellner, J. A. (1986): *Empirical Processs With Appliation to Statistics.* Wiley (1986)

[64] .Siddiqui, M. M. (1960): Distribution of Quantiles in samples from a bivariate Population. *J. Res. NBS 64B N°***3***, 145-150.*

[65] Stone, C. J. (1976): Nearest neighbour estimators of a non linear regression function. *Proc. Comp. Sci. Statist. U. C. L. A 413-418.*

[66] Stone, C. (1977): Consistent Non parametric Regression. *Ann. of Stat, Vol.* ***5,*** *n°***4** *594-645.*

[67] Stone, C. (1980): Optimal rates of convergence for non parametric estimators. *Ann. Statist.* ***8, 6,*** *1348-1360.*

[68] Stone, C(1982): Optimal global rate of convergence for non-parametric regression. *Ann. of stat, Vol.* 10, *n*°**4**, *1040-1053.*

[69] Stute, W. (1986a): Conditional Empirical process. *Ann. of Stat. vol. 14, n*°**2,** *638-647.*

[70] Stute, W. (1986b): On almost sure convergence of conditionnal empirique distribution

function. *Ann of Prob. Vol. 14, n*°**3**, *891-901*

[71] Tomkins, R, J(1976): On conditional Median. *Ann of Prob. Vol.* ***3,*** *375-379.*

[72].Truong, Y. R. (1989): Asymptotique properties of kernel estimators based on local median. *Ann. of. Stat.*

[73] Tukey, J. W. (1961): Curves as parameters, and touch estimation. *Proc.* ***4th.*** *symp. math. stat. Prob. Berkeley, p. 681-694*

[74] Valadier, M. (1984): La multi-application Mediane Conditionnel. *Z. W. G. Vol.* ***67,*** *279-282*

[75] Watson G. , Leadbetter M. , (1963): On the estimation of probability density, *Ann. Math. Statist.* ***34,*** *475-491.*

[76] .Watson, G. S. (1964): Smooth régression analysis. *Sankhya A**26,*** *359-372.*