



Faculté des Sciences Exactes et Informatique
Département de Mathématiques

N° d'ordre :

N° de séries :

Mémoire de fin d'études

Présenté pour l'obtention du diplôme de

Master

Spécialité : Mathématiques.

Option : Probabilités et Statistique.

Thème

Étude de la Convergence Presque Complète des Estimateurs à Noyaux pour les Fonctions de Densité et de Régression

Le : 08/07/2019

Présenté par :

- Boukahoul Hassina.

- Sioual Rima.

Devant le jury :

Président : Chraitia Hassan M. C. B Université de Jijel

Encadreur : Madi Meriem M. A. A Université de Jijel

Examineur : Gherda Mebrouk M. A .A Université de Jijel

Table des matières

Introduction	vi
1 Estimation non paramétrique	1
1.1 Estimation de la fonction de répartition	2
1.1.1 Estimation par la fonction de répartition empirique	2
1.1.2 Estimation à noyau de la fonction de répartition	5
1.2 Estimation de la densité	8
1.2.1 Histogramme	8
1.2.2 Estimateur simple	10
1.2.3 Méthode du noyau	12
1.2.4 Estimation de la densité de probabilité par un système trigonomé- trique	14
1.3 Estimation de la fonction de régression	17
1.3.1 Estimateurs à noyau de la fonction de régression	18
1.3.2 La méthode des k-plus proches voisins	24
1.3.3 la méthode des polynômes locaux	27
2 la convergence presque complète des estimateur à noyaux	29
2.1 Définitions et propriétés	29

2.1.1	Liens entre la convergence complète et la convergence presque sûre et en probabilité	30
2.1.2	Taux de convergence	32
2.1.3	Opération de calcul	33
2.2	Exemples	36
2.3	Convergence presque complète de l'estimateur de la fonction de répartition	37
2.4	Convergence presque complète de l'estimateur de la densité	40
2.5	Convergence presque complète de l'estimateur de la fonction de régression .	46
	Conclusion	55
	Bibliographie	56

Remerciements

Nous remercions le Dieu pour le courage, la patience et la volonté qui nous ont été utiles tout au long de notre parcours.

Nous tenons à remercier *M^{me}* **M. Madi** pour la proposition du thème, l'encadrement de ce travail, pour ses précieux conseils et orientations.

Nous remercions également les membres du jury *M^r* **M. Gharda** & *M^r* **H. Chraitia** pour avoir accepté d'examiner et d'évaluer notre travail.

Nos sincères remerciements s'adressent enfin à tous ceux qui nous ont soutenu de près ou de loin.

Hassina & Rima

Dédicace

Je dédie ce travail :

À

mes très **chers parents**. Vous vous êtes dépensés pour moi sans compter. En reconnaissance de tous les sacrifices consentis par vous, pour me permettre d'atteindre cette étape de ma vie, vous avez toute ma tendresse.

À

mes frères et sœurs : **Mohammed, yacine, Hicham, Hanane, Nassima, Siham, Khadidja** Pour votre profonde passion et soutien tout au long de mes études. Affectueuse reconnaissance.

À

Les enfants de mes frères et Les enfants de mes sœurs qui leur souhaitent une carrière d'apprentissage réussie.

À

mes amis : **Najiba, Iman, Rania, Ghada, Hafida, Amina, Karima** ceux avec qui j'ai parcouru un long chemin avec tant de peine et de joie, inquiétude et espoir.

À

mon ami proche Ammar pour son soutien continu et ses conseils.

tous ceux dont les noms n'y figurent pas pour une raison ou une autre trouve l'expression de ma profonde gratitude.

Hassina

Dédicace

Je dédie ce travail :

À ceux qui m'ont tout donné sans rien en retour A ceux qui m'ont encouragée et soutenue dans les moments les plus difficiles A vous **mes chers parents**, le plus beau cadeau que Dieu puissent faire à un enfant, pour leur amour et leur support continu. Que ce travail soit le témoignage sincère et affectueux de ma profonde reconnaissance pour tout ce que vous avez fait pour moi.

À mes chers grands-parents et grand-mères.

À mes chers frères **tadjinne** et **hichame**.

À mes soeurs **Nadjma**, **Zineb** et **Afaf**.

À mon mari **Ishak** et sa famille.

À tous les membres de ma famille, en particulier mes oncles

À mon ami **Newal**, que a toujours été avec moi.

À tous mes amis et collègues.

Rima

Introduction

La théorie de l'estimation est une des branches les plus basiques de la statistique. Cette théorie est habituellement divisée en deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non paramétrique. Le problème de l'estimation non paramétrique consiste, dans la majeure partie des cas, à estimer, à partir des observations, une fonction inconnue, élément d'une certaine classe fonctionnelle. Plus particulièrement, on parle d'estimation non paramétrique lorsque celle-ci ne se ramène pas à l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon. Un des plus vieux problèmes de la statistique non paramétrique consiste à estimer la densité de probabilité $f(\cdot)$, la fonction de répartition $F(\cdot)$ et la régression $r(\cdot)$ à partir d'un échantillon des variables aléatoires indépendantes et identiquement distribuées X_1, X_2, \dots, X_n . Il s'agit d'un problème fondamental qui a connu, durant ces dernières années, des développements théoriques et pratiques à la fois rapides et nombreux. Le problème de l'estimation de la densité de probabilité, la fonction de répartition et la fonction de régression est important pour plusieurs raisons :

1. La densité de probabilité permet d'avoir un aperçu très rapide des principales caractéristiques de la distribution (pics, creux, symétries,...), ce qui explique le volume important de littérature qui lui est consacré.
2. C'est en terme de comportement local de la fonction de répartition que s'explique plus facilement le comportement des estimateurs fonctionnels (vitesse de convergence, normalité asymptotique) et c'est par un estimateur de la fonction de répartition que l'on passe pour estimer des probabilités d'ensembles : la probabilité qu'une variable soit dans un intervalle donné ou qu'une observation au moins d'un nouvel échantillon dépasse un seuil fixé.
3. La régression non paramétrique est devenue une méthode populaire pour analyser la relation entre une variable dépendante Y et une variable indépendante X . Son objet, est d'estimer cette relation de dépendance sans faire d'hypothèses paramé-

triques sur la forme de cette dépendance.

Il existe plusieurs méthodes non paramétriques pour l'estimation de la densité, la fonction de répartition et de la régression. Dans ce mémoire nous étudierons quelques méthodes comme : la méthode d'estimation par histogramme, la méthode d'estimation par noyau, La méthode des k-plus proches voisins et la méthode des Polynômes locaux ,..., mais nous concentrons particulièrement sur la méthode d'estimation par noyau. Par définition la méthode du noyau est une méthode statistique non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. En ce sens, cette méthode généralise astucieusement la méthode d'estimation par un histogramme. Les estimateurs obtenus par la méthode à noyau (ainsi que les estimateurs des autres) sont consistants, et notre objectif est d'étudier leur convergence presque complète. Cette dernière est un mode de convergence de variables aléatoires plus forte que la convergence presque sûre, introduite par H. Robbins en 1947.

Ce mémoire est composé d'une introduction, de deux chapitres et d'une conclusion.

Dans le chapitre un nous concentrons sur les méthodes d'estimation, de la fonction de répartition (fonction de répartition empirique, Estimation à noyau), de la densité (Histogramme, Estimateur simple, Méthode du noyau et par un système trigonométrique) et de la régression (Estimateurs à noyau, méthode des k-plus proches voisins et méthode des polynômes locaux), et nous présentons également, les propriétés statistiques des estimateurs (variance, biais, erreur moyenne quadratique...) et les propriétés asymptotique(convergence en moyenne quadratique et la convergence en moyenne quadratique intégrée) pour chaque méthode d'estimation.

Dans le deuxième chapitre, nous étudions la convergence presque complète sa définition, relation entre ce mode de convergence et d'autres modes et propriétés de base. On s'intéresse à l'étude de la convergence presque complète des estimateurs non paramétrique à noyau.

Chapitre 1

Estimation non paramétrique

Dans ce chapitre, nous rappelons quelques méthodes d'estimation non paramétrique, et nous concentrons sur les méthodes d'estimation de la fonction de répartition, de densité et de régression, en donnant les propriétés statistiques des estimateurs de chaque méthode d'estimation.

on s'intéresse à l'estimation non paramétrique à noyaux, et surtout à la convergence de ses estimateurs.

Le problème de l'estimation non paramétrique consiste à estimer, à partir des observations, une fonction inconnue, élément d'une certaine classe fonctionnelle assez large, notons cette classe fonctionnelle par \mathcal{F} , la fonction inconnue d'échantillon par f et l'échantillon des observations par (x_1, x_2, \dots, x_n) . On peut poser le problème de l'estimation de f comme suit : soit (Ω, \mathcal{A}, P) un espace de probabilité sur lequel est définie une famille de variables aléatoires (X_1, X_2, \dots, X_n) indépendantes suivant la même loi.

La question est : Que peut on dire de f à partir de (x_1, x_2, \dots, x_n) avec (x_1, x_2, \dots, x_n) une réalisation de (X_1, X_2, \dots, X_n) ?

Définition 1.1.

Un estimateur de la fonction f est une fonction :

$$\hat{f}_n : x \longrightarrow f_n(x) = f_n(x_1, x_2, \dots, x_n),$$

mesurable par rapport à l'observation (X_1, X_2, \dots, X_n) .

Définition 1.2.

Le biais d'un estimateur \hat{f}_n de la fonction f est la quantité définie par :

$$\text{Biais}(\hat{f}_n) = E(\hat{f}_n) - f.$$

1. On dit qu'un estimateur \hat{f}_n de f est sans biais si :

$$\text{Biais}(\hat{f}_n) = 0.$$

c-à-d :

$$E(\hat{f}_n) = f.$$

2. On dit qu'un estimateur \hat{f}_n de f est asymptotiquement sans biais si :

$$\lim_{n \rightarrow +\infty} E(\hat{f}_n) = f.$$

Définition 1.3.

L'erreur moyenne quadratique d'un estimateur \hat{f}_n , notée MSE , est définie comme suit :

$$MSE(\hat{f}_n(x)) = E(f(x) - \hat{f}_n(x))^2.$$

Remarque.

On peut écrire MSE en fonction du biais et de la variance comme suit :

$$MSE(\hat{f}_n(x)) = \text{Biais}(\hat{f}_n(x))^2 + \text{Var}(\hat{f}_n(x)).$$

Définition 1.4.

L'erreur moyenne quadratique intégrée d'un estimateur \hat{f}_n , qui est notée $MISE$, est donné par l'expression suivante :

$$MISE(\hat{f}_n(x)) = \int MSE(\hat{f}_n(x))dx = \int \text{Biais}(\hat{f}_n(x))^2 dx + \int \text{Var}(\hat{f}_n(x))dx.$$

1.1 Estimation de la fonction de répartition

L'objectif de cette section est d'estimer la fonction de répartition F qui caractérise la loi de probabilité d'une variable aléatoire, d'abord par la fonction de répartition empirique, en suite par la méthode du noyau.

1.1.1 Estimation par la fonction de répartition empirique

On observe X_1, \dots, X_n variables aléatoires réelles indépendantes et identiquement distribuées de fonction de répartition F , définie comme suit :

$$F(x) = P(X \leq x).$$

Définition 1.5.

Un estimateur de la fonction de répartition F est la fonction empirique \hat{F}_n définie par :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}, \forall x \in \mathbb{R}.$$

$$= \begin{cases} 0, & x \leq X_{(1)} \\ \frac{k}{n}, & X_{(k)} \leq x \leq X_{(k+1)} \text{ si } k = 1, \dots, n-1 \\ 1, & x \geq X_{(n)}, \end{cases}$$

avec :

$$X_{(1)} \leq \dots \leq X_{(n)},$$

et 1_A est la fonction indicatrice de l'événement A .

Voici quelques propriétés de la \hat{F}_n :

Propriété 1.6.

L'estimateur \hat{F}_n possède les propriétés suivantes :

- 1) $\forall x \in \mathbb{R}$, $n\hat{F}_n(x)$ est de loi binomiale $B(n, F(x))$.
- 2) $\forall x \in \mathbb{R}$, $\hat{F}_n(x)$ est un estimateur sans biais et convergent en moyenne quadratique.
- 3) $\forall x \in \mathbb{R}$, $\hat{F}_n(x)$ converge presque sûrement vers $F(x)$.

$$\forall x \in \mathbb{R}, \hat{F}_n(x) \xrightarrow[n \rightarrow +\infty]{P. S} F(x).$$

- 4) $\forall x \in \mathbb{R}$, $\hat{F}_n(x)$ converge en probabilité vers $F(x)$.

$$\forall x \in \mathbb{R}, \hat{F}_n(x) \xrightarrow[n \rightarrow +\infty]{P} F(x).$$

Démonstration.

- 1) On a :

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

c'est la somme de n variables aléatoires indépendantes de même loi de Bernoulli de paramètre $P(X_i \leq x) = F(x)$, alors

$$n\hat{F}_n(x) \rightsquigarrow B(n, F(x)).$$

2) Calculons le biais et la variance :

$$\begin{aligned} E(\hat{F}_n(x)) &= E\left(\frac{1}{n}(n\hat{F}_n(x))\right) \\ &= \frac{1}{n}E(n\hat{F}_n(x)) \\ &= \frac{1}{n}nF(x) \\ &= F(x). \end{aligned}$$

Donc $\hat{F}_n(x)$ est un estimateur sans biais.

$$\begin{aligned} Var(\hat{F}_n(x)) &= Var\left(\frac{1}{n}(n\hat{F}_n(x))\right) \\ &= \left(\frac{1}{n}\right)^2 Var(n\hat{F}_n(x)) \\ &= \frac{1}{n^2}(nF(x)(1-F(x))) \\ &= \frac{F(x)(1-F(x))}{n}, \end{aligned}$$

en remarque que $\lim_{n \rightarrow \infty} var(\hat{F}_n(x)) = 0$ et puisque $\hat{F}_n(x)$ est non biaisé, alors

$$\hat{F}_n(x) \xrightarrow[n \rightarrow +\infty]{MQ} F(x), \forall x.$$

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

En appliquant la loi forte de grand nombres sur la suite indépendante de variables aléatoires de Bernoulli $1_{\{X_i \leq x\}}$, on obtient :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} \xrightarrow[n \rightarrow +\infty]{P. S} E(1_{\{X_i \leq x\}}) = P(X_i \leq x) = F(x).$$

D'où la convergence presque sûrement.

4) D'après l'inégalité de Bienaymé-Tchebychev.

$$\begin{aligned} \forall \varepsilon > 0, P(|\hat{F}_n(x) - F(x)| \geq \varepsilon) &\leq \frac{Var(\hat{F}_n(x))}{\varepsilon^2} \\ &\leq \frac{F(x)(1-F(x))}{n\varepsilon^2} \xrightarrow[n \rightarrow +\infty]{} 0. \end{aligned}$$

Alors on a : $\hat{F}_n(x) \xrightarrow[n \rightarrow +\infty]{P} F(x)$.

■

1.1.2 Estimation à noyau de la fonction de répartition

Dans ce partie, on cherche à estimer la fonction de répartition F par la méthode du noyau qui a été introduite par Rosenblatt en 1956, puis amélioré par Parzen en 1962 (pour l'estimation de la densité de probabilité).

Un noyau est une fonction de pondération utilisée dans l'estimation non paramétrique de la densité de probabilité, la fonction de répartition, la fonction de régression...

Définition 1.7.

Soit k une fonction à valeurs réelle, positive et intégrable, telle que : $\int_{\mathbb{R}} k(u)du = 1$, la fonction k est dite noyau.

Définition 1.8.

Un noyau k est dit symétrique si pour tout u dans l'ensemble de définition de k , on a $k(u) = k(-u)$ ce qui implique

$$\int_{\mathbb{R}} uk(u)du = 0.$$

– k est dit de carré intégrable si

$$\int_{\mathbb{R}} k^2(u)du < +\infty.$$

Exemple 1.9.

Les noyau symétriques les plus utilisées sont :

- Le noyau rectangulaire $k(u) = \frac{1}{2}1_{[-1,1]}(u)$.
- Le noyau parabolique d'Epanechnikov $k(u) = \frac{3}{4\sqrt{5}}(1 - \frac{u^2}{5})1_{]-\sqrt{5},\sqrt{5}[}(u)$.
- Le biweight de tukey $k(u) = \frac{15}{16}(1 - u^2)^2 1_{]-1,1]}(u)$.
- Noyau sinusoidal $k(u) = \frac{\Pi}{4} \cos(\frac{\Pi}{2}u) 1_{]-1,1]}(u)$.
- Le noyau gaussien $k(u) = \frac{1}{2\sqrt{\Pi}} \exp(\frac{-u^2}{2})$.

Définition 1.10.

Un estimateur à noyau H de la fonction de répartition F , noté \hat{F}_{h_n} , est définie par :

$$\hat{F}_{h_n}(x) = \frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h_n}\right),$$

où h_n est le paramètre de lissage appelé fenêtre de l'estimateur.

Remarque.

H est appelé noyau intégré, c'est à dire :

$$H(u) = \int_{-\infty}^u k(t)dt,$$

tel que :

$$\int_{\Omega} k(t)dt = 1.$$

Propriété 1.11.

L'estimateur \hat{F}_{h_n} possède les propriétés suivantes :

- 1) \hat{F}_{h_n} est un estimateur asymptotiquement sans biais.
- 2) L'estimateur à noyau $\hat{F}_{h_n}(x)$ converge en moyenne quadratique vers $F(x)$.

Démonstration.

- 1) L'objectif est de vérifier que $E(\hat{F}_{h_n}(x)) \rightarrow F(x)$. Quand $n \rightarrow \infty$

$$\begin{aligned} E(\hat{F}_{h_n}(x)) &= \frac{1}{n} E \left(\sum_{i=1}^n H \left(\frac{x - X_i}{h_n} \right) \right) \\ &= E \left(H \left(\frac{x - X}{h_n} \right) \right) \quad (\text{car } X_1, \dots, X_n \text{ sont de même loi que } X) \\ &= \int_{-\infty}^{+\infty} H \left(\frac{x - z}{h_n} \right) f(z) dz, \end{aligned}$$

Une intégration par parties, nous donne :

$$\begin{aligned} E(\hat{F}_{h_n}(x)) &= \left[H \left(\frac{x - z}{h} \right) F(z) \right]_{-\infty}^{+\infty} + \frac{1}{h_n} \int_{-\infty}^{+\infty} H' \left(\frac{x - z}{h_n} \right) F(z) dz \\ &= 0 + \frac{1}{h_n} \int_{-\infty}^{+\infty} H' \left(\frac{x - z}{h_n} \right) F(z) dz \\ &= \int_{-\infty}^{+\infty} H'(y) F(x - hy) dy, \quad \text{en posant } y = \frac{x - z}{h}, \end{aligned}$$

en utilisant le développement de Taylor à l'ordre 1 de F au voisinage de x on obtient :

$$F(x - yh_n) = F(x) - h_n y f(x) + O(h_n^2).$$

Alors

$$\begin{aligned} E(\hat{F}_{h_n}(x)) &= \int_{-\infty}^{+\infty} H'(y)[F(x) - h_n y f(x)] dy \\ &= \int_{-\infty}^{+\infty} [H'(y)F(x) - h_n y H'(y) f(x)] dy + O(h_n^2) \\ &= F(x) \int_{-\infty}^{+\infty} H'(y) dy - h_n \int_{-\infty}^{+\infty} y H'(y) f(x) dy + O(h_n^2) \end{aligned}$$

car $H'(u) = k(u)$, alors

$$E(\hat{F}_{h_n}(x)) = F(x) - h_n f(x) \int_{-\infty}^{+\infty} y H'(y) dy + O(h_n^2),$$

car le noyau est symétrique. au passage à la limite quand $n \rightarrow \infty$, $h \rightarrow 0$

$$E(\hat{F}_{h_n}(x)) \rightarrow F(x).$$

D'où \hat{F}_{h_n} est un estimateur asymptotiquement sans biais.

- 2) L'objectif est de vérifier que $\lim_{n \rightarrow \infty} E(\hat{F}_{h_n}(x) - F(x))^2 = 0$.

On a :

$$\text{var}(\hat{F}_{h_n}(x) - F(x)) = E[(\hat{F}_{h_n}(x) - F(x))^2] - (E(\hat{F}_{h_n}(x) - F(x)))^2.$$

Il suffit donc de vérifier : $\text{var}(\hat{F}_{h_n}(x)) \rightarrow 0$ quand $n \rightarrow \infty$.

$$\begin{aligned} \text{var}[\hat{F}_{h_n}(x)] &= \text{var} \left(\frac{1}{n} \sum_{i=1}^n k \left(\frac{x - X_i}{h_n} \right) \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{var} \left(k^2 \left(\frac{x - X_i}{h_n} \right) \right) \\ &= \frac{1}{n} \left[E \left(k^2 \left(\frac{x - X_i}{h_n} \right) \right) - \left(E \left(k \left(\frac{X_i - x}{h_n} \right) \right) \right)^2 \right] \end{aligned}$$

$$\begin{aligned} E \left[k^2 \left(\frac{x - X_i}{h_n} \right) \right] &= \int_{-\infty}^{+\infty} k^2 \left(\frac{x - z}{h_n} \right) f(z) dz. \\ &= \left[k^2 \left(\frac{x - z}{h_n} \right) F(x) \right]_{-\infty}^{+\infty} + \frac{1}{h_n} \int_{-\infty}^{+\infty} \dot{k}^2 \left(\frac{x - z}{h_n} \right) F(z) dz. \\ &= F(x) \int_{-\infty}^{+\infty} \dot{k}^2(y) dy, \end{aligned}$$

donc

$$\text{var}[\hat{F}_{h_n}(x)] = \frac{1}{n} [F(x) \int_{-\infty}^{+\infty} (\dot{k}^2)(y) dy - (F(x) - h_n f(x) (\int_{-\infty}^{+\infty} y \dot{k}(y) dy))^2 + O(h_n^2)] \xrightarrow[n \rightarrow +\infty]{} 0.$$

Puis que $\hat{F}_{h_n}(x)$ est non biaisé alors

$$\hat{F}_{h_n}(x) \xrightarrow[n \rightarrow +\infty]{\text{MQ}} F(x), \forall x.$$



1.2 Estimation de la densité

Dans cette section, on suppose que la loi de l'échantillon (X_1, \dots, X_n) est continue et on cherche à estimer sa densité f .

Par définition f est la dérivée de la fonction de répartition F . Mais la fonction de répartition empirique \hat{F}_{h_n} n'est pas dérivable, car c'est une fonction en escalier, on ne peut pas donc utiliser directement les résultats sur la fonction de répartition empirique pour estimer la densité. On peut se demander quelle est l'utilité d'estimer la densité alors que l'on a déjà un très bon estimateur de la fonction de répartition ? La principale raison est que la forme d'une densité est beaucoup plus facile à interpréter que celle d'une fonction de répartition.

1.2.1 Histogramme

L'estimation de f par histogramme, consiste à estimer f en un point x par la proportion de variables aléatoires (X_1, \dots, X_n) qui se trouve dans un intervalle de longueur h_n , et qui contient x .

Un histogramme est défini à partir d'une suite double de valeurs croissantes

$\{., ., a_{i-1}, \dots, a_{-1}, a_0, a_1, \dots, a_i\}$, consistant un découpage, en intervalles $]a_k, a_{k+1}]$ de la droite réelle.

Soit n_k la fréquence des observations situées dans l'intervalle $]a_k, a_{k+1}]$ pour un échantillon de taille n , alors l'histogramme est la fonction constante par morceaux \hat{f}_n définie pour tout $k \in Z$ par :

$$\hat{f}_n(x) = \frac{n_k/n}{(a_{k+1} - a_k)}, \text{ pour } x \in]a_k, a_{k+1}]$$

qui nous donne la représentation graphique en rectangles.

Pour estimer f par un histogramme, on fixe d'abord une borne inférieure de l'échantillon $a_0 < x_1$ et une borne supérieure $a_k > x_n$, puis on partitionne l'intervalle $]a_0, a_k]$ contenant toutes les observations en k classes $]a_{k-1}, a_k]$, dont la largeur de la classe j est $h_j = a_j - a_{j-1}$.

Définition 1.12.

L'estimateur de f sur $]a_k, a_{k+1}]$, de type histogramme, noté \hat{f}_{h_n} , est définie par :

$$\begin{aligned}\hat{f}_{h_n}(x) &= \frac{n_k}{nh_n}, \text{ pour } x \in]a_k, a_{k+1}] \\ &= \frac{\sum_{k=1}^n 1_{]a_k, a_{k+1}]}(x)}{nh_n} \\ &= \frac{N_k}{nh_n},\end{aligned}$$

où N_k est le nombre aléatoire de valeurs tombant dans $]a_k, a_{k+1}]$, telle que $N_k \rightsquigarrow B(n, p_k)$, où p_k est la probabilité estimer par $\frac{n_k}{n}$.

Propriété 1.13.

L'estimateur \hat{f}_n de la densité par histogramme possède les propriétés suivantes :

1. L'espérance : on a :

$$N_j \rightsquigarrow B(n, p_j),$$

on déduit que :

$$E(\hat{f}_{h_n}(x)) = \frac{E(n_j)}{nh_n} = \frac{p_j}{h_n}.$$

2. La variance :

$$\text{var}(\hat{f}_{h_n}(x)) = \frac{p_j(1-p_j)}{nh_n^2}.$$

L'expression de la variance montre que plus le paramètre h_n est petit, plus $\text{var}(\hat{f}_{h_n}(x))$ est grand c'est à dire moins efficace, inversement, plus h_n est grand, plus $\hat{f}_{h_n}(x)$ est efficace.

Maintenant, le comportement asymptotique du biais, variance, MSE et MISE : en faisant le développement de Taylor à l'ordre 1, on obtient les propriétés suivantes :

– Le biais de l'estimateur $\hat{f}_{h_n}(x)$:

$$\begin{aligned}\text{Biais}(\hat{f}_{h_n}(x)) &= E(\hat{f}_{h_n}(x)) - f(x) \\ &= \frac{1}{2}f'(x)[h_n - 2(x - a_k)] + O(h_n^2).\end{aligned}$$

– La variance de l'estimateur $\hat{f}_{h_n}(x)$:

$$\begin{aligned}\text{Var}(\hat{f}_{h_n}(x)) &= E(\hat{f}_{h_n}(x)^2) - (E(\hat{f}_{h_n}(x)))^2 \\ &= \frac{f(x)}{nh_n} + O(n^{-1}).\end{aligned}$$

Si $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$ alors $\text{Var}(\hat{f}_{h_n}(x)) \rightarrow 0$.

– La convergence en moyenne quadratique de l'estimateur $\hat{f}_{h_n}(x)$:

$$\begin{aligned} MSE(\hat{f}_{h_n}(x)) &= Var(\hat{f}_{h_n}(x)) + Biais(\hat{f}_{h_n}(x))^2 \\ &= \frac{f(x)}{nh_n} + \frac{f'(x)^2}{4}(h_n - 2(x - a_k))^2 + O(n^{-1}) + O(h_n^2). \end{aligned}$$

Si $h \rightarrow 0$ et $nh \rightarrow \infty$ quand $n \rightarrow \infty$ alors $MSE(\hat{f}_h(x)) \rightarrow 0$.

– La convergence en moyenne quadratique intégrée de l'estimateur $\hat{f}_{h_n}(x)$

$$MISE(\hat{f}_{h_n}(x)) = \frac{1}{nh_n} + \frac{h_n^2 \int f'^2(t) dt}{12} + O(n^{-1}) + O(h_n^3).$$

Si $h \rightarrow 0$ et $nh \rightarrow \infty$ quand $n \rightarrow \infty$ alors $MISE(\hat{f}_{h_n}(x)) \rightarrow 0$.

1.2.2 Estimateur simple

Rappelons que la densité de probabilité f est égale à la dérivée de la fonction de répartition F (si cette dérivée existe). On peut donc écrire

$$\begin{aligned} f_{h_n}(x) = F'(x) &= \lim_{h \rightarrow 0} \frac{P(x - h_n \leq X_i \leq x + h_n)}{2h_n} \\ &= \lim_{h \rightarrow 0} \frac{F(x + h_n) - F(x - h_n)}{2h_n} \end{aligned}$$

c'est à dire que

$$f_{h_n}(x) \simeq \frac{F(x + h_n) - F(x - h_n)}{2h_n}.$$

Si on remplace F par l'estimateur \hat{F}_{h_n} , on obtient :

$$\hat{f}_{h_n}(x) = \frac{\hat{F}(x + h_n) - \hat{F}(x - h_n)}{2h_n},$$

où \hat{f}_{h_n} est l'estimateur de f appelé estimateur de Rosenblatt.

Définition 1.14. L'estimateur de Rosenblatt, de la fonction de densité f est définie par :

$$\begin{aligned} \hat{f}_{h_n}(x) &= \frac{\hat{F}(x + h_n) - \hat{F}(x - h_n)}{2h_n} \\ &= \frac{1}{2nh_n} \sum_{i=1}^n 1_{\{-1 \leq \frac{x - X_i}{h_n} < +1\}}. \end{aligned}$$

Remarque.

Notons que cet estimateur peut encore s'écrire comme

$$\hat{f}_{h_n}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} w\left(\frac{x - X_i}{h_n}\right),$$

$$\text{où } w(y) = \begin{cases} 1, & \text{si } y \in [1, -1[\\ 0, & \text{sinon.} \end{cases}$$

Propriété 1.15.

1. *Le biais : Nous savons que*

$$n\hat{F}_{h_n}(x) = \sum_{i=1}^n 1(X_i \leq x) \rightsquigarrow B(n, F(x)),$$

et

$$2nh_n\hat{f}_{h_n}(x) = n\hat{F}(x + h_n) - n\hat{F}(x - h_n) \rightsquigarrow B(n, F(x + h_n) - F(x - h_n)),$$

$$\implies E(2nh_n\hat{f}_{h_n}(x)) = n(F(x + h_n) - F(x - h_n)).$$

Par conséquent :

$$E(\hat{f}_{h_n}(x)) = \frac{1}{2h_n}(F(x + h_n) - F(x - h_n)),$$

si $n \rightarrow \infty$ et $h \rightarrow 0$ alors

$$E(\hat{f}_{h_n}(x)) \rightarrow f_{h_n}(x).$$

2. *La variance :*

$$\text{var}(2nh_n\hat{f}_{h_n}(x)) = n(F(x + h_n) - F(x - h_n))(1 - F(x + h_n) + F(x - h_n))$$

$$\implies \text{var}(\hat{f}_{h_n}(x)) = \frac{1}{4nh_n^2}(F(x + h_n) - F(x - h_n))(1 - F(x + h_n) + F(x - h_n)),$$

et

$$nh_n\text{var}(\hat{f}_{h_n}(x)) \rightarrow \frac{1}{2}f_{h_n}(x).$$

3. *Le MSE :*

si $nh_n \rightarrow \infty$ et $h_n \rightarrow 0$ quand $n \rightarrow \infty$ on a que Le risque quadratique moyen de l'estimateur

$$MSE(\hat{f}_{h_n}(x)) = E(\hat{f}_{h_n}(x) - f_{h_n}(x))^2 \rightarrow 0,$$

pour tout point x .

1.2.3 Méthode du noyau

La méthode du noyau est une généralisation de la méthode d'estimation par histogramme. l'estimateur \hat{f}_n reste une fonction en escalier, pour obtenir quelque chose de plus lisse. On a d'après l'estimateur simple de la densité

$$\begin{aligned}\hat{f}_{h_n}(x) &= \frac{1}{2nh_n} \sum_{i=1}^n 1_{]x-h_n, x+h_n]}(X_i) \\ &= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} 1_{\{x-h_n < X_i < x+h_n\}} \\ &= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} 1_{]-1,1]} \left(\frac{x - X_i}{h_n} \right).\end{aligned}$$

On pose $k(u) = \frac{1}{2} 1_{]-1,1]}(u)$, la méthode du noyau consiste à généraliser cette approche à d'autres fonction k .

Définition 1.16.

Un estimateur à noyau de la densité f est une fonction \hat{f}_{h_n} définie par :

$$\hat{f}_{h_n}(x) = \frac{1}{nh_n} \sum_{i=1}^n k \left(\frac{x - X_i}{h_n} \right),$$

où $\{h_n\}_{n \geq 1}$ est une suite de réels positifs appelées paramètres de lissage ou largeurs de la fenêtre qui tend vers 0 quand n tend vers l'infini et k est une densité de probabilité appelée noyau.

Propriété 1.17.

Soit x fixe dans R :

- 1.) $Biais(\hat{f}_{h_n}(x)) = \frac{h_n^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 k(u) du + O(h_n^2)$.
- 2.) $var(\hat{f}_{h_n}(x)) = \frac{1}{nh_n} f(x) \int_{-\infty}^{+\infty} k^2(u) du + O(\frac{1}{nh_n})$.
- 3.) $MSE(\hat{f}_{h_n}(x)) = \frac{1}{nh_n} f(x) \int_{-\infty}^{+\infty} k^2 du + O(\frac{1}{nh_n}) + \frac{h_n^4}{4} f''(x)^2 [\int_{-\infty}^{+\infty} u^2 k(u) du]^2 + O(h_n^4)$.

Démonstration.

Sous les hypothèses :

- (K. 1) $\int_R k(u) du = 1$,
- (K. 2) $\int_R uk(u) du = 0$,
- (K. 3) $\int_R u^2 |k(u)| du \leq \infty$.

On a :

1)

$$\begin{aligned}
E(\hat{f}_{h_n}(x)) &= \frac{1}{nh_n} E\left(\sum_{i=1}^n k\left(\frac{X_i - x}{h_n}\right)\right) \\
&= \frac{1}{h_n} \int_{-\infty}^{+\infty} k\left(\frac{y-x}{h_n}\right) f(y) dy \\
&= \int_{-\infty}^{+\infty} k(u) f(x + uh_n) du, \text{ en posant } u = \frac{y-x}{h_n}.
\end{aligned}$$

En utilisant le développement de Taylor de f au voisinage de x on obtient :

$$\begin{aligned}
f(x + uh_n) &= f(x) + h_n u f'(x) + \frac{(uh_n)^2 f''(x)}{2} + O(h_n^2) \\
E(\hat{f}_{h_n}(x)) &= \int_{-\infty}^{+\infty} k(u) [f(x) + f'(x)uh + \frac{1}{2}f''(x)(uh_n)^2 + O(h_n^2)] du \\
&= f(x) \int_{-\infty}^{+\infty} k(u) du + h_n f'(x) \int_{-\infty}^{+\infty} uk(u) du + f''(x) \frac{h_n^2}{2} \int_{-\infty}^{+\infty} u^2 k(u) du \\
&= f(x) + \frac{h_n^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 k(u) du + O(h_n^2),
\end{aligned}$$

alors :

$$\begin{aligned}
\text{Biais}(\hat{f}_{h_n}(x)) &= E(\hat{f}_{h_n}(x)) - f(x) \\
&= \frac{h_n^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 k(u) du + O(h_n^2).
\end{aligned}$$

Alors :

$$\lim \text{Biais}(\hat{f}_{h_n}(x)) = 0,$$

lorsque $n \rightarrow \infty$ et $h \rightarrow 0$.

2)

$$\begin{aligned}
\text{var}(\hat{f}_{h_n}(x)) &= \text{var}\left(\frac{1}{nh_n} \sum_{i=1}^n k\left(\frac{X_i - x}{h_n}\right)\right) \\
&= \frac{1}{n} \left\{ \frac{1}{h_n^2} E\left[k\left(\frac{X-x}{h_n}\right)^2\right] - \frac{1}{h_n} E^2\left[k\left(\frac{X-x}{h_n}\right)\right] \right\} \\
&= \frac{1}{nh_n} \int_{-\infty}^{+\infty} (k(u))^2 f(x + uh_n) du - \frac{1}{n} \left[\int_{-\infty}^{+\infty} k(u) f(x + uh_n) du \right]^2,
\end{aligned}$$

le terme $\left[\frac{1}{n} \int_{-\infty}^{+\infty} k(u) f(x + uh_n) du \right]^2 \rightarrow 0$ lorsque $n \rightarrow +\infty$.

Donc :

$$\text{var}(\hat{f}_{h_n}(x)) = \frac{1}{nh_n} f(x) \int_{-\infty}^{+\infty} k^2(u) du + O\left(\frac{1}{nh_n}\right).$$

D'où :

$$\lim_{n \rightarrow \infty} \text{var}(\hat{f}_{h_n}(x)) = 0, \text{ quand } nh_n \rightarrow \infty.$$

3)

$$\begin{aligned} \text{MSE}(\hat{f}_{h_n}(x)) &= E(\hat{f}_{h_n}(x) - f(x))^2 \\ &= E(\hat{f}_{h_n}(x) - f_{h_n}(x) + f_{h_n}(x) - f(x))^2 \\ &= \text{var}(\hat{f}_{h_n}(x)) + \text{Biais}(\hat{f}_{h_n}(x))^2 \\ &= \frac{1}{nh_n} f(x) \int_{-\infty}^{+\infty} k^2 du + O\left(\frac{1}{nh_n}\right) + \frac{h_n^4}{4} f''(x)^2 \left[\int_{-\infty}^{+\infty} u^2 k(u) du \right]^2 + O(h_n^4). \end{aligned}$$

■

Remarque.

$$\begin{aligned} \text{MISE}(\hat{f}_{h_n}(x)) &= \frac{1}{nh_n} \int_{-\infty}^{+\infty} \left(f(x) \left[\int_{-\infty}^{+\infty} k^2(u) du \right] + O\left(\frac{1}{nh_n}\right) \right) dx \\ &\quad + \frac{h_n^4}{4} \int_{-\infty}^{+\infty} \left((f''(x))^2 \left[\int_{-\infty}^{+\infty} u^2 k(u) du \right]^2 + O(h_n^4) \right) dx. \end{aligned}$$

$h = h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$ alors $\text{MSE}(\hat{f}_{h_n}(x)) \rightarrow 0$. On dit que $\hat{f}_{h_n}(x)$ converge en probabilité vers $f(x)$.

1.2.4 Estimation de la densité de probabilité par un système trigonométrique

Soit (X_1, \dots, X_n) une suite de variables aléatoires indépendantes et identiquement distribuées de densité de probabilité inconnue f sur $[-\pi, +\pi]$. Il s'agit d'estimer $f(x)$ à partir des observations (x_1, \dots, x_n) , Pour cela on considère la base donnée par :

$$e_k(x) = \frac{1}{\sqrt{2\pi}} (\cos(kx) + \sin(kx)) 1_{[-\pi, +\pi]}, k = 0, 1, 2, \dots, \quad (1.1)$$

la base donnée par (1. 1) est orthogonale dans $[-\pi, +\pi]$, c'est-à-dire :

$$\int_0^1 e_k(x) e_j(x) dx = \begin{cases} 1 & j = k \\ 0 & \text{sinon.} \end{cases}$$

La densité de probabilité $f(x)$ peut se mettre sous la forme :

$$f(x) = \sum_{k=0}^{\infty} a_k e_k(x), x \in [-\pi, +\pi],$$

où

$$a_k = \int_0^1 e_k(x) f(x) dx = E(e_k(x)).$$

Les coefficients a_k peuvent être estimés par :

$$\hat{a}_k = \frac{1}{n} \sum_{k=0}^{\infty} e_k(X_i) = \frac{1}{\sqrt{2\pi 2}} \sum_{i=1}^n (\cos k(X_i) + \sin k(X_i)).$$

Définition 1.18.

L'estimateur $\hat{f}_{d_n}(x)$ de $f(x)$ est donnée par :

$$\begin{aligned} \hat{f}_{d_n}(x) &= \sum_{k=0}^{d_n} \hat{a}_k e_k(x) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{k=0}^{d_n} a_k(X_i) e_k(x) \\ &= \frac{1}{4\pi n} \sum_{i=1}^n \left[\frac{\sin\left(\frac{(2d_n+1)(X_i+x)}{2}\right)}{\sin\left(\frac{(X_i+x)}{2}\right)} \right] + \sum_{i=1}^n \left[\frac{\sin\left(\frac{(2d_n+1)(X_i-x)}{2}\right)}{\sin\left(\frac{(X_i-x)}{2}\right)} \right], \end{aligned}$$

où d_n est une suite des nombres réels positifs tels que $d_n \rightarrow \infty$ quand $n \rightarrow \infty$.

Propriété 1.19.

1. $\int_{-\pi}^{\pi} \hat{f}_{d_n}(x) dx = 1$ et $|\hat{f}_{d_n}(x)| \leq \frac{2d_n+1}{2\pi}$.
2. $\text{Biais}(\hat{f}_{d_n}(x)) = \sum_{k=d_n+1}^{\infty} a_k e_k(x)$.
- 3.

$$\begin{aligned} \text{var}(\hat{f}_{d_n}(x)) &= \frac{d_n+1}{4\pi^2 n} + \frac{1}{2\pi\sqrt{2\pi n}} \sum_{k=0}^{d_n} \beta_{2k} - \frac{1}{2\pi n} \sum_{k=0}^{d_n} a_k^2 + \frac{1}{4\pi^2 n} \sum_{k=0}^{d_n} \sin(2kx) \\ &+ \frac{1}{2n\pi\sqrt{2\pi}} \sum_{k=0}^{d_n} \sin(2kx) \beta_{2k} - \frac{1}{2\pi n} \sum_{k=0}^{d_n} a_k^2 \sin(2kx) + \\ &\sum_{k=0}^{d_n} \sum_{j=0}^{d_n} \frac{1}{n} \left[\frac{1}{\sqrt{2\pi}} \gamma_{k-j} + \frac{1}{\sqrt{2\pi}} \beta_{k+j} - a_k a_j \right] e_k(x) e_j(x), \end{aligned}$$

avec

$$\gamma_k = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) \cos(kx) dx = \frac{1}{\sqrt{2\pi}} E(\cos(kX)),$$

et

$$\beta_k = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) \sin(kx) dx = \frac{1}{\sqrt{2\pi}} E(\sin(kX)).$$

4.

$$\begin{aligned}
MSE(\hat{f}_{d_n}(x)) &= \frac{d_n + 1}{4\pi^2 n} + \frac{1}{2\pi\sqrt{2\pi n}} \sum_{k=0}^{d_n} \beta_{2k} - \frac{1}{2\pi n} \sum_{k=0}^{d_n} a_k^2 + \frac{1}{4\pi^2 n} \sum_{k=0}^{d_n} \sin(2kx) \\
&+ \frac{1}{2n\pi\sqrt{2\pi}} \sum_{k=0}^{d_n} \sin(2kx)\beta_{2k} - \frac{1}{2\pi n} \sum_{k=0}^{d_n} a_k^2 \sin(2kx) + \\
&\sum_{k=0}^{d_n} \sum_{j=0}^{d_n} \frac{1}{n} \left[\frac{1}{\sqrt{2\pi}} \gamma_{k-j} + \frac{1}{\sqrt{2\pi}} \beta_{k+j} - a_k a_j \right] e_k(x) e_j(x) + \left[\sum_{k=d_n+1}^{\infty} a_k e_k(x) \right]^2.
\end{aligned}$$

$$5. MSIE(\hat{f}_{d_n}(x)) = \int_{-\pi}^{\pi} f^2(x) dx + \frac{d_n+1}{2\pi n} + \frac{1}{\sqrt{2\pi n}} \sum_{k=0}^{d_n} \beta_{2k} - \frac{n+1}{n} \sum_{k=0}^{d_n} a_k^2.$$

Démonstration.voir [4] ■**propriété asymptotique****Théorème 1.20.**1. Si $d_n \rightarrow \infty$, $\hat{f}_{d_n}(x)$ est asymptotiquement sans biais :

$$\lim_{n \rightarrow \infty} E(\hat{f}_{d_n}(x)) = f(x).$$

2. Si $d_n = o(\sqrt{n})$ et $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, alors

$$\lim_{n \rightarrow \infty} \text{var}(\hat{f}_{d_n}(x)) = 0.$$

3. Si $d_n = o(\sqrt{n})$ et $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, alors

$$\lim_{n \rightarrow \infty} MSE(\hat{f}_{d_n}(x)) = 0.$$

4. Si $d_n = o(\sqrt{n})$ et $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, alors

$$\lim_{n \rightarrow \infty} MSIE(\hat{f}_{d_n}(x)) = 0.$$

Démonstration.Voir [4] ■

1.3 Estimation de la fonction de régression

Supposons que l'on dispose de n couples de variables aléatoires $(X_1, Y_1), \dots, (X_n, Y_n)$ indépendantes et de même loi que le couple (X, Y) telles que :

$$Y = r(X) + \varepsilon.$$

où la variable aléatoire ε est la variation de Y autour de X , vérifie $E(\varepsilon) = 0$ et indépendante de X . r est une fonction de $[0, 1]$ à valeurs réelle, dite fonction de régression, est inconnue. Le problème de régression non paramétrique est d'estimer r , lorsque l'on sait a priori que cette fonction appartient à un ensemble non paramétrique (de dimension infinie) \mathcal{F} , qui peut être l'ensemble de toutes les fonctions continues sur $[0, 1]$ où l'ensemble des fonctions convexe,...

On cherche dans cette famille de fonctions \mathcal{F} , quelle est celle pour laquelle les Y sont les plus proches de $r(X)$, c'est à dire on déterminera la fonction $r(X)$ qui rendra l'erreur quadratique moyenne la plus petite possible i. e :

$$E(r(X) - Y)^2 = \min_r E(r(X) - Y)^2,$$

ce minimum est donnée par :

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

la fonction r exprime la valeur moyenne de la variable aléatoire à expliquer Y en fonction de la variable explicative X .

Le couple (X, Y) est à valeur dans R^2 , donc il admet une densité jointe $f(x, y)$ sur R^2 et une densité marginale $f(x) > 0$ par rapport à la mesure de Lebesgue sur R .

On suppose que la variable aléatoire Y est intégrable par conséquent on peut définir la fonction de régression $r(x)$ par :

$$r(x) = E(Y/X = x) = \int_R y f_{Y/X}(y/X = x) dy \quad (1.2)$$

$$= \frac{\int_R y f_{X,Y}(x, y) dy}{f(x)}. \quad (1.3)$$

$$(1.4)$$

Remarque.

$r(x)$ est la fonction qui donne la meilleure approximation de Y sachant $X = x$ au sens des moindres carrés.

1.3.1 Estimateurs à noyau de la fonction de régression

Dans ce paragraphe, on représente la construction de l'estimateur à noyau de Nadaraya Watson (1964) basé sur l'idée de Tukey (1961), qui a introduit un estimateur à noyau de type régressogramme de la fonction de régression définie par :

$$r_{h_n}(x) = \frac{\sum_{i=1}^n Y_i 1_{[t_k, t_{k+1}[}(X_i)}{\sum_{i=1}^n 1_{[t_k, t_{k+1}[}(X_i)} \text{ pour } x \in [t_k, t_{k+1}[,$$

où $[t_k, t_{k+1}[$, $k \in N$ est une partition de support de x . Bosq (1969) fût le premier à donner une étude des propriétés statistiques de cet estimateur, il a montré la convergence uniforme presque sûre du régressogramme sur l'intervalle $[a, b]$, quand

$$h_n = O(n^{-\alpha}), 0 < \alpha < 1.$$

Sabry (1978) a montré la convergence uniforme presque sûre sur $[0, \sqrt{\frac{\log n}{\log \log n}}]$ et Lecoutre (1982) a procédé à l'extension de tous ces résultats à R . (pour une étude plus approfondie de cet estimateur on se réfère au livre de Bosq et Lecoutre (1987)). Pour éviter le problème de positionnement des bornes dans intervalles de la partition, un autre estimateur a été construit comme suit :

$$\forall x, \hat{r}_h(x) = \frac{\sum_{i=1}^n Y_i 1_{[x-h, x+h[}(X_i)}{\sum_{i=1}^n 1_{[x-h, x+h[}(X_i)}, \text{ où } h \in R_*^+.$$

L'inconvénient de ce nouveau estimateur est sa discontinuité. Sa généralisation a été introduite et étudiée par Nadaraya (1964) et Watson (1964).

Une façon de construire l'estimateur de Nadaraya-Watson et l'utiliser les estimateurs de $f(x, y)$ et de $f(x)$ dans (1. 3) comme suit :

$$r(x) = \frac{\int_R y f_{X,Y}(x, y) dy}{f(x)}.$$

un estimateur à noyau de la densité joint $f_{(X,Y)}$, noté $\hat{f}_{(X,Y)}$ peut être obtenu en suivant les mêmes étapes utilisés pour obtenir $\hat{f}(x)$ estimateur de $f(x)$, et on obtient :

$$\hat{f}_{X,Y}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n k\left(\frac{X_i - x}{h}\right) k\left(\frac{Y_i - y}{h}\right),$$

d'où l'estimateur de Nadaraya-Watson (1964) de la fonction r est :

$$\hat{r}_{h_n}(x) = \frac{\frac{1}{nh_n} \sum_{i=0}^n Y_i k\left(\frac{X-x}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=0}^n k\left(\frac{X-x}{h_n}\right)}.$$

Définition 1.21.

L'estimateur à noyau, de Nadaraya-Watson (1964) notée $\hat{r}_{h_n}(x)$ de la fonction de régression r est donnée par :

$$\hat{r}_{h_n}(x) = \begin{cases} \frac{\sum_{i=0}^n Y_i k\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=0}^n k\left(\frac{x-X_i}{h_n}\right)} = \frac{\hat{\phi}_{h_n}(x)}{\hat{f}_{h_n}(x)}, & \text{si } 1\left(\sum_{i=0}^n k\left(\frac{x-X_i}{h_n}\right)\right) \neq 0 \\ \frac{1}{n} \sum_{i=0}^n Y_i, & \text{sinon,} \end{cases}$$

où

$$\hat{\phi}_{h_n}(x) = \frac{1}{nh_n} \sum_{i=0}^n Y_i k\left(\frac{x-X_i}{h_n}\right).$$

Remarque.

L'estimateur \hat{r}_{h_n} dépend de deux paramètres le paramètre de lissage h_n et noyau k

Définition 1.22.

Un estimateur $\hat{r}_{h_n}(x)$ de $r(x)$ est dit estimateur linéaire de la régression non-paramétrique si :

$$\hat{r}_{h_n}(x) = \sum_{i=0}^n Y_i w_i,$$

où la fonction de poids $w_i(\cdot)$ ne dépend pas des observations Y_i .

Définition 1.23.

L'estimateur à noyau introduit par Watson Nadaraya, de la fonction de régression évaluée au point x_0 , noté $\hat{r}_{h_n}(x_0)$, est défini par :

$$\hat{r}_{h_n}(x) = \sum_i^n Y_i w_i(x_0),$$

avec

$$w_i(x_0) = \frac{k\left(\frac{x_0-X_i}{h_n}\right)}{\sum_{i=0}^n k\left(\frac{x_0-X_i}{h_n}\right)},$$

où $k(\cdot)$ désigne une fonction noyau, $h_n > 0$ un paramètre de lissage.

Les propriétés de l'estimateur de la fonction de régression

1. Étude asymptotique de la variance :

Proposition 1.24.

Sous les hypothèses :

(k. 1)

(k. 4) k est bornée, c'est à dire $\exists M \geq 0, \forall u \in R, |k(u)| < M$,

(k. 5) $\lim |u|k(u) = 0$, quand $|x| \rightarrow \infty$,

(k. 6) $k \in L^1(R)$, c'est à dire $\int_R |k(u)| du < \infty$,

et si $E(Y^2) < \infty$, alors en chaque point de continuité des fonctions $r(x)$, $f(x)$ et $\sigma^2(x) = \text{Var}(Y/X = x)$. On a

$$\text{var}(\hat{r}_{h_n}(x)) = \frac{1}{nh_n} \left\{ \frac{\sigma^2(x)}{f(x)} \int_R k^2(u) du \right\} (o(1) + 1),$$

où

$$f(x) > 0.$$

Démonstration.

Soit la fonction $\psi(x) = \int_R y^2 f(x, y) dy$, en se basant sur le lemme de Bochner suivant :

Lemme 1.25. {Bochner}

Soit $K : (R^m, \beta^m) \mapsto (R, \beta)$ une fonction mesurable, où

β^p est la tribu borélienne de R^p , vérifiant :

$$\exists M \text{ (constante) telle que, } \forall z \in R^m, |k(z)| \leq M,$$

$$\int_{R^m} |k(z)| dz < \infty,$$

et

$$\|z\|^m |k(z)| \mapsto 0 \text{ quand } \|z\| \mapsto \infty.$$

Par ailleurs, soit

$g : (R^m, \beta^m) \mapsto (R, \beta)$ une fonction telle que

$$\int_{R^m} |g(z)| dz < \infty,$$

Si g est continue, et si $0 < h_n \mapsto 0$, quand $n \mapsto \infty$ alors :

$$\lim_{n \rightarrow \infty} \frac{1}{h_n^m} \int_{R^m} k\left(\frac{z}{h_n}\right) g(x - z) dz = g(x) \int_{R^m} k(z) dz.$$

on a

$$\begin{aligned} \text{var}(\hat{\phi}_{h_n}(x)) &= \frac{1}{nh_n^2} \left\{ E \left[Y^2 k^2 \left(\frac{x-X}{h_n} \right) \right] - E \left[Y k \left(\frac{x-X}{h_n} \right) \right]^2 \right\} \\ &= \frac{1}{nh_n} \left\{ \int_R k^2(u) \psi(x-h_n u) du - h_n \left(\int_R k(u) f(x-h_n u) (x-h_n u) du \right)^2 \right\} \\ &= \frac{1}{nh_n} \psi(x) \int_R k^2(u) du (1 + o(1)). \end{aligned}$$

on a

$$\begin{aligned} \text{cov}(\hat{f}_{h_n}(x), \hat{\phi}_{h_n}(x)) &= E[\{\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))\} \{\hat{\phi}_{h_n}(x) - E(\hat{\phi}_{h_n}(x))\}] \\ &= \frac{1}{nh_n} \phi(x) \int_R k^2(u) du (1 + o(1)), \end{aligned}$$

et

$$\text{var}(\hat{f}_{h_n}(x)) = \frac{1}{nh_n} f(x) \int_R k^2(u) du (1 + o(1)).$$

Posons

$$B_n(x) = \begin{pmatrix} \hat{f}_{h_n}(x) \\ \hat{\phi}_{h_n}(x) \end{pmatrix}$$

et

$$A(x) = \begin{pmatrix} -r(x) \\ \frac{1}{f(x)} \end{pmatrix}$$

La matrice de variance covariance de $B_n(x)$ est alors donnée par l'expression suivante :

$$\Sigma := \frac{1}{nh_n} \begin{pmatrix} f(x) & \phi(x) \\ \phi(x) & \psi(x) \end{pmatrix} \int_R k^2(u) du (1 + o(1))$$

En remarquant, que :

$$\text{var}(\hat{f}_{h_n}(x)) = A \Sigma A^t = \frac{1}{nh_n} f(x) \left(\frac{\psi(x)}{(f(x)^2)} - \frac{(\phi(x))^2}{(f(x)^3)} \right) \int_R k^2(u) du (1 + o(1)),$$

où A^t désigne la transposée de A , on obtient alors :

$$\text{var}(\hat{f}_{h_n}(x)) = \frac{1}{nh_n} \left\{ \frac{\sigma^2(x)}{f(x)} \int_R k^2(u) du \right\} (1 + o(1)).$$

■

2. **Étude asymptotique du biais** : l'étude asymptotique du biais repose sur la proposition suivante :

Proposition 1.26.

Sous les hypothèses de la proposition (1. 16) et

a) Si $|Y| \leq C_1 < \infty$ P.S et si $nh_n \rightarrow \infty$, quand $n \rightarrow \infty$, alors :

$$E(\hat{r}_{h_n}(x)) = \frac{E(\hat{\phi}_{h_n}(x))}{E(\hat{f}_{h_n}(x))} + O\left(\frac{1}{nh_n}\right).$$

b) Si $E(Y^2) < \infty$, $nh_n^2 \rightarrow \infty$, quand $n \rightarrow \infty$, alors :

$$E(\hat{r}_{h_n}(x)) = \frac{E(\hat{\phi}_{h_n}(x))}{E(\hat{f}_{h_n}(x))} + O\left(\frac{1}{\sqrt{nh_n}}\right).$$

Démonstration. En multipliant les deux expressions de l'identité suivante :

$$\frac{1}{\hat{f}_{h_n}(x)} = \frac{1}{E(\hat{f}_{h_n}(x))} - \frac{\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))}{E(\hat{f}_{h_n}(x))^2} + \frac{[\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))]^2}{\hat{f}_{h_n}(x)E(\hat{f}_{h_n}(x))^2}.$$

Par $\hat{\phi}_{h_n}(x)$ et en passant à l'espérance nous obtenons :

$$E(\hat{r}_{h_n}(x)) = \frac{E(\hat{\phi}_{h_n}(x))}{E(\hat{f}_{h_n}(x))} - (E(\hat{f}_{h_n}(x)))^{-2} E(\hat{\phi}_{h_n}(x)) - E(\hat{\phi}_{h_n}(x))(\hat{f}_{h_n}(x))E(\hat{f}_{h_n}(x)) \quad (1.5)$$

$$+ (E(\hat{f}_{h_n}(x)))^{-1} (E(\hat{f}_{h_n}(x)))^{-2} E(\hat{\phi}_{h_n}(x)) [\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))] \quad (1.6)$$

$$= \frac{E(\hat{\phi}_{h_n}(x))}{E(\hat{f}_{h_n}(x))} + [c_n^1(x) - c_n^2(x)] (E(\hat{f}_{h_n}(x)))^{-2}, \quad (1.7)$$

où

$$c_n^{(1)}(x) = E[(\hat{\phi}_{h_n}(x) - E(\hat{\phi}_{h_n}(x)))(\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)))], \quad (1.8)$$

$$c_n^{(2)}(x) = E[(\hat{f}_{h_n}(x)^{-1} - E(\hat{f}_{h_n}(x))) (\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)))]. \quad (1.9)$$

Comme $var(\hat{f}_{h_n}(x)) \sim \frac{1}{nh_n} f(x) \int_R k^2(t) dt$ et

$var(\hat{\phi}_{h_n}(x)) \sim \frac{1}{nh_n} \int_R y^2 f(x, y) dy \int_R k^2(t) dt$, alors

$$|c_n^1(x)| \leq var(\hat{\phi}_{h_n}(x))^{\frac{1}{2}} var(\hat{f}_{h_n}(x))^{\frac{1}{2}} = O\left(\frac{1}{nh_n}\right). \quad (1.10)$$

Nous constatons aussi que l'hypothèse a) implique l'inégalité suivante

$$|c_n^2(x)| \leq C_1 var(\hat{f}_{h_n}(x)) \sim \frac{1}{nh_n} f(x) \int_R k^2(t) dt = O\left(\frac{1}{nh_n}\right). \quad (1.11)$$

En combinant les relations (1. 5), (1. 10), (1. 11) nous obtenons le résultat a). Pour montrer le cas b) il suffit de remarquer que la relation (1. 10) est toujours valable

mais la relation (1. 11) devient

$$|c_n^2(x)| \leq E \left(\max_{1 \leq i \leq n} |Y_i| \right) E \left[\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)) \right]^2 \quad (1.12)$$

$$\leq \left[\sum_{i=1}^n E(Y_i^2) \right]^{\frac{1}{2}} \left[E[\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))]^4 \right]^{\frac{1}{2}} \quad (1.13)$$

$$= \sqrt{n} E(Y^2)^{\frac{1}{2}} O \left(\frac{1}{nh_n} \right) \quad (1.14)$$

$$= O \left(\frac{1}{\sqrt{nh_n}} \right) \quad (1.15)$$

Les relations, (1. 10), (1. 11) et (1. 12) donnent le résultat b). ■

Maintenant nous sommes en mesure d'énoncer le résultat suivant.

Proposition 1.27.

Si les hypothèses (K. 1), (K. 2) et (K. 3) sont vérifiées et si $f(\cdot)$ et $r(\cdot)$ sont de classe $C^2(R)$ et si $|y|$ est borné.

Alors :

$$E(\hat{r}_{h_n}(x)) - r(x) = \frac{h_n^2}{2} \left\{ \left\{ r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right\} \int_R u^2 k(u) du \right\} (1 + o(1)). \quad (1.16)$$

Remarque.

(a) (K. 1), (K. 2) et (K. 3) peuvent être remplacées par le noyau k est d'ordre 2 au sens de Gasser et Müller.

(b) $o(1)$ dans la relation (1. 16) est égale à $O(h)O((nh)^{-1})$

Démonstration.

On a

$$\begin{aligned} \frac{E(\hat{\phi}_{h_n}(x))}{E(\hat{f}_{h_n}(x))} - r(x) &= [E(k \left(\frac{x-t}{h_n} \right))]^{-1} \left\{ \int_R \frac{1}{h_n} k \left(\frac{x-t}{h_n} \right) \phi(t) dt - r(x) \int_R \frac{1}{h_n} k \left(\frac{x-t}{h_n} \right) f(t) dt \right\} \\ &= \left\{ (f(x))^{-1} \left\{ \frac{h_n}{2} \phi''(x) - \frac{h_n}{2} f''(x) \right\} \int_R k(u) du + \phi(x) - r(x)f(x) \right\} (1 + o(1)) \end{aligned}$$

Comme $\phi(x) = r(x)f(x)$. L'équation précédente peut s'écrire :

$$E(\hat{r}_{h_n}(x)) - r(x) = \left\{ \frac{h_n}{2} \left\{ r''(x) - 2r'(x) \frac{f'(x)}{f(x)} \right\} \int_R k(u) du \right\} (1 + o(1)),$$

D'où

$$\lim_{n \rightarrow \infty} E(\hat{r}_{h_n}(x)) = r(x). \quad \blacksquare$$

1.3.2 La méthode des k-plus proches voisins

La construction des estimateurs k-plus proches voisins diffère de l'estimateur à noyau. L'estimateur à noyau $\hat{r}(x)$ est défini comme une moyenne pondérée des variables réponses dans un voisinage fixé autour de x , déterminé dans la forme par le noyau k et la largeur de bande h . L'estimateur k plus proche voisin est une moyenne pondérée dans un voisinage variable. Ce voisinage est défini par ces variables X qui sont parmi les k-plus proches voisins de x par rapport à la distance euclidienne. La suite des poids des k-plus proches voisins a été introduite par [Loftsgaarden and Quesenberry] dans le domaine lié à l'estimation de la densité et utilisée par [Cover and Hart] dans le but de classification

Définition 1.28.

Dans le cadre présent de la régression, le lisseur k-plus proche voisin est défini par

$$\hat{r}_k(x) = n^{-1} \sum_{i=1}^n \omega_{ki}(x) y_i$$

$\{\omega_{ki}(x)\}_{i=1}^n$ est une suite de poids définie par l'ensemble d'indices

$J_x = \{i : x_i \text{ est l'une des } k\text{-plus proches observations de } x\}$, avec cet ensemble d'indices des observations voisines, la suite des poids des k-plus proches voisins est construite comme suit

$$\omega_{ki}(x) = \begin{cases} \frac{n}{k} & i \in J_x \\ 0 & \text{sinon.} \end{cases}$$

Exemple 1.29.

Considérons l'exemple suivant.

Soit $\{(x_i, y_i)\}_{i=1}^5 = \{(1, 5), (7, 12), (3, 1), (2, 0), (5, 4)\}$ et calculons les k-plus proches voisins $\hat{r}_k(x)$ pour $x = 4$ et $k = 3$. Les observations proches de x sont les 3 derniers points des données. Par conséquent $J_x = J_4 = \{3, 4, 5\}$ donc

$$w_{k1}(4) = w_{k2}(4) = 0, w_{k3}(4) = w_{k4}(4) = w_{k5}(4) = \frac{5}{3}.$$

Ce qui résulte

$$\hat{r}_3(4) = \frac{1}{5}(1 + 0 + 4) \frac{5}{3}.$$

Dans une expérience où la variable X est choisie dans une grille équidistante, les poids des k-plus proches voisins sont équivalents au poids du noyau. Soit $k = 2nh$ et comparons $\{w_{ki}(x)\}$ avec $\{w_{hi}(x)\}$ pour un noyau uniforme $k(u) = \frac{1}{2}1(|u| \leq 1)$ pour un x pas très

proche de la limite.

Ainsi pour $i \in J_x$:

$$w_{ki}(x) = \frac{n}{2nh} = \frac{1}{2}h^{-1} = w_{hi}(x).$$

Le paramètre de lissage k règle le degré de lissage de la courbe estimée. Il joue un rôle similaire à la largeur de bande pour le noyau. Quand on fait varier k l'influence sur les caractéristiques qualitatives de la courbe estimée est similaire à celle observée par l'estimation à noyau avec le noyau uniforme.

Considérons pour n fixé, le cas où k devient plus grand que n . Le lisseur k -plus proche voisin est alors égal à la moyenne des variables réponses. L'autre cas limite est $k = 1$ dans lequel les observations se répètent à x_i , et pour un x entre deux variables adjacentes, une fonction seuil est obtenue avec un saut au milieu des deux observations. Un autre problème de sélection du paramètre de lissage est observé : k doit être choisi comme une fonction de n ou même des données. Comme 1^{er} objectif, c'est de pouvoir réduire le bruit, posons $k = k_n$ tend vers l'infini comme fonction de la taille de l'échantillon. Le second objectif est de garder l'erreur approximative (le biais) petite. Le second objectif est réalisé si le voisinage autour de x se rétrécit asymptotiquement vers zéro, ceci peut être fait par définir $k = k_n$ tel que $k_n \rightarrow 0$. Malheureusement, cette condition contredit le 1^{er} objectif. Pour garder la variance petite autant que possible il faut choisir k grand que possible.

Pour cela, nous envisageons encore un compromis entre "trouver une bonne approximation" pour la fonction de régression et une bonne réduction du bruit observé. Ce problème de compromis peut être formulé par un développement de l'erreur moyenne quadratique de l'estimateur k -plus proches voisins.

Proposition 1.30. (Lai, 1977)

Soit $k \rightarrow \infty$, $\frac{k}{n} \rightarrow \infty$, le biais et la variance de l'estimateur k -plus proches voisins $\hat{r}_k(x)$ sont donnés par :

$$E(\hat{r}_k(x) - r(x)) \approx \frac{1}{24f(x)^3} [(r''f_X + 2r'f'_x)(x)] \left(\frac{k}{n}\right)^2,$$

$$\text{var}(\hat{r}_k(x)) \approx \frac{\sigma^2(x)}{k}.$$

Le compromis entre le biais-variance est ainsi réalisée dans un sens asymptotique en posant $k \sim n^{\frac{4}{5}}$. Une conséquence est que l'erreur moyenne quadratique elle-même converge vers 0 au taux de $k^{-1} \sim n^{-\frac{4}{5}}$. Autre suite de poids $\omega_{ki}(x)$ ajoutant au poids uniforme il a défini les poids triangulaires et quadratiques du k -plus proches voisins, Généralement, les poids

peuvent être pris comme étant générés par des fonctions noyaux K .

$$\omega_{Ri} = \frac{k \left(\frac{x-x_i}{R} \right)}{\sum_{i=1}^n k \left(\frac{x-x_i}{R} \right)}$$

où $R = R_n$ est la distance entre x et ses k plus proches voisins.

Proposition 1.31. (Mack, 1981)

Soit $k \rightarrow \infty$, $\frac{k}{n} \rightarrow 0$, et soit c_k et d_k définis comme suit : $c_k = \int k^2(u)du$ et $d_k = \int u^2 k(u)du$ Alors on a :

$$E(\hat{r}_R(x) - r(x)) \approx \frac{k}{n} \frac{r'' f + 2r' f'}{8f(x)^3} d_k$$

$$\text{var}(\hat{r}_R(x)) \approx 2 \frac{\sigma^2(x)}{k} c_k$$

Une conséquence de cette proposition est qu'une contribution d'équilibre entre le biais² et la variance de l'erreur moyenne quadratique est pour le poids uniforme du k - plus proche voisin est réalisée en posant $k = k_n$ est proportionnel à $n^{\frac{4}{5}}$ Pensant, que la largeur de bande h des lisseurs noyau peuvent être équivalent à $\frac{1}{2}kn^{-1}$ grossièrement, il est vu que les taux du biais et de la variance de \hat{r}_R sont complètement équivalents à ceux des lisseurs noyau, uniquement les constantes diffèrent Le biais de \hat{r}_R tend d'être grand (à la fin de la distribution marginale). L'estimateur à noyau présente un comportement différent : la variance est un multiple de $f(x)^{-1}$ et le biais s'avérer ne pas d'être une fonction de $f(x)^3$. Une comparaison des propriétés de l'erreur moyenne quadratique des lisseurs noyau et k -plus proches voisins peuvent être trouvées dans le tableau suivant :

	noyau	k-plus proche voisin
Biais	$h^2 \frac{(r'' f + 2r' f')(x)}{2f(x)} d_k$	$\left(\frac{k}{2}\right)^2 \frac{r'' f + 2r' f'}{8f^3(x)} d_k$
Variance	$\frac{\sigma^2(x)}{nhf(x)} c_k$	$2 \frac{\sigma^2(x)}{k} c_k$

TABLE 1.1 – Biais et variance des lisseurs noyau et k-plus proches voisins.

Les entrées du tableau montrent une dépendance en f , h et k . L'équivalence essentielle dans la ligne du tableau peut être vue par l'utilisation de la relation $k = 2nhf(x)$ L'utilisation de k conduit (asymptotiquement) à la même erreur moyenne quadratique en x pour les k -plus proches voisins et le noyau.

1.3.3 la méthode des polynômes locaux

La régression par polynômes locaux est très similaire à l'estimation du noyau. Par contre, les valeurs obtenues sont produites par régression polynomiale pondérée par la distance au lieu de la moyenne pondérée par la distance. En fait, ce qui est différent de la méthode du noyau est que l'estimation de $r(x)$ est obtenue par régression polynomiale de y sur x . Un polynôme de degré p en accordant une pondération $w_i(x)$ à chacune des observations x_i de la fenêtre de lissage $((x - h(x)), x + h(x))$. Ainsi, soit $P(u)$ le polynôme de degré p ayant la forme suivante :

$$P(u) = \beta_0 + \beta_1(u - x) + \beta_2 \frac{(u - x)^2}{2!} + \dots + \beta_p \frac{(u - x)^p}{p!}$$

lorsque $|u - x| < h(x)$. Alors, le vecteur de paramètres : $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$ peut être obtenu en appliquant localement la méthode des moindres carrés pondérés, ce qui consiste à obtenir le vecteur de paramètres $\hat{\beta}$ qui minimise

$$\sum_{i=1}^n \omega_i(x) \left\{ y_i - \beta_0 - \beta_1(x_i - x) - \beta_2 \frac{(x_i - x)^2}{2!} - \dots - \beta_p \frac{(x_i - x)^p}{p!} \right\}^2 \dots *$$

où :

$$\omega_i(x) = W \left(\frac{x_i - x}{h(x)} \right),$$

où $W(u)$ est une fonction noyau symétrique de support $[-1, 1]$ qui satisfait les conditions suivantes de Loader (1999) : $W(0) = 1$, $W(1) = 0$ et W décroît sur l'intervalle $[0, 1]$ (et croît sur $[-1, 0]$).

En faisant l'hypothèse que la matrice $X_x^T W_x X_x$, est inversible (dans le cas contraire, on peut faire en sorte qu'elle le devienne en modifiant h), on peut montrer que la solution de (*) donnée par la théorie classique des moindres carrés pondérés est (Wande et Jones, 1990) :

$$\hat{\beta} = (X_x^T W_x X_x)^{-1} X_x^T W_x Y$$

où $Y = (y_1, \dots, y_n)^T$ est le vecteur des réponses .

$$X_x = \begin{pmatrix} 1 & \dots & x_1 - x & \dots & \frac{(x_1 - x)^p}{p!} \\ \vdots & \ddots & \vdots & \dots & \vdots \\ 1 & \dots & x - x_n & \dots & \frac{(x_n - x)^p}{p!} \end{pmatrix} \text{ est la matrice de conception de dimension } n \times (p+1)$$

et $W_x = \text{diag}\{\omega_1(x), \dots, \omega_n(x)\}$ est la matrice diagonale de dimension $n \times n$ des pondérations.

Ensuite, on pose $\hat{r}_p(x) = \hat{\beta}_x^0$.

Il est important de souligner le fait que la méthode du noyau est en réalité un cas particulier de la méthode par polynômes locaux. En effet, pour un choix de $p = 0$, l'équation à minimiser devient exactement celle qui donnait la méthode du noyau :

$$\sum_{i=1}^n W_i(x)(Y_i - \beta_0)^2.$$

En général, hormis le cas où $p = 0$, le choix de l'ordre du polynôme local est de $p = 1$. Dans cette situation, on parle d'une régression localement linéaire. Tout d'abord, il faut savoir que l'estimateur localement linéaire ($p = 1$) de $r(x)$ est

$$\hat{r}_1(x) = \frac{1}{nh} \sum_{i=1}^n \frac{[\hat{s}_2(x, h) - \hat{s}_1(x, h)(x - x_i)](W(\frac{x-x_i}{h}))y_i}{\hat{s}_2(x, h)\hat{s}_0(x, h) - \hat{s}_1(x, h)^2},$$

où

$$\hat{s}_r(x, h) = \frac{1}{nh} \sum_{i=1}^n (x - x_i)^r W\left(\frac{x - x_i}{h}\right).$$

Comme pour la méthode du noyau, il y a deux types de formules pour le biais et la variance de $\hat{r}_1(x)$: celles pour les estimateurs évalués à des points x qui se situent à l'intérieur des extrémités des données et celles pour les estimateurs évalués à des points x situés près des extrémités. Celles pour les estimateurs évalués à des points se trouvant à l'intérieur des bornes sont tout d'abord données. La formule pour le biais est donc

$$\text{Biais}(\hat{r}_1(x)|x_1, \dots, x_n) = \frac{h^2 r''(x) u_2(W_{(1)})}{2} + o_p(h^2),$$

où $u_q(K(p))$ est défini

$$u_q(K(p)) = \int u^q W_p(u) du.$$

La variance vaut pour sa part

$$\text{var}(\hat{r}_1(x)|x_1, \dots, x_n) = \frac{R(W_{(1)})\sigma^2(x)}{nhf(x)} + o_p(nh^{-1}),$$

où :

$$R(W_{(p)}) = \int W_{(p)}(u)^2 du.$$

Puisque $W_{(0)} = W_{(1)}$, on remarque que cette variance est exactement la même que celle de la méthode du noyau, mais que le biais est plus faible. En effet, en comparant ce biais avec celui obtenu par la méthode du noyau, on voit que ce dernier comporte un terme supplémentaire en facteur de h^2 ,

$$\frac{r'(x)f'(x)}{f(x)} u_2(k_{(0)}).$$

Ainsi, contrairement au biais trouvé par la méthode du noyau, le biais obtenu par la méthode localement linéaire ne dépend ni de la fonction de densité des valeurs exogènes x_i , ni de sa dérivée $f'(x)$.

Chapitre 2

la convergence presque complète des estimateur à noyaux

Sur l'espace de probabilité (Ω, A, P) , il ya beaucoup de type de convergence de suite de variable aléatoires : la convergence en loi, en probabilité, presque sure et dans L^p , ($p \geq 1$) et la convergence presque complète. Ce dernier est type de convergence le plus fort que les autres. Ci-dessous on donne la définition de la convergence presque complète, les propositions et quelques exemples, et on étudié le lien entre la convergence presque complète et la convergence en probabilité et presque *sûre*, afin de faire étude de convergence presque complète des estimateurs à noyaux pour les fonctions de répartition, de densité et de régression.

2.1 Définitions et propriétés

Définition 2.1.

On dit que $(X_n)_{n \in \mathbb{N}}$ converge presque complète vers une variable aléatoire X , si et seulement si :

$$\forall \varepsilon > 0, \sum_{n \in \mathbb{N}} P(|X_n - X| > \varepsilon) < \infty$$

et notée par : $\lim_{n \rightarrow \infty} X_n = X$, p. co, cette notion que l'on peut parfois appeler plus simplement une convergence complète.

La convergence presque complète est plus forte que la convergence en probabilité puisque cette dernière nécessite seulement que pour chaque : $\varepsilon > 0, P(|X_n - X| > \varepsilon)$ converge

vers 0 quand $n \rightarrow \infty$, mais la convergence presque complète exige que pour tout $n \geq 1$, $P(|X_n - X| > \varepsilon)$ tend vers 0 lorsque $n \rightarrow \infty$, c'est à dire qu'elle exige une vitesse de convergence suffisante pour assurer que la somme infini $\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon)$ converge. Par exemple, si $P(|X_n - X| > \varepsilon) = n^{-1}$ alors la suite de variables aléatoires $(X_n)_{n \geq 1}$ converge en probabilité vers X , mais pas complètement, lorsque $n \rightarrow \infty$. Si $P(|X_n - X| > \varepsilon) = n^{-2}$ alors la suite de variables aléatoires $(X_n)_{n \geq 1}$ converge en probabilité et complètement à X , lorsque $n \rightarrow \infty$.

2.1.1 Liens entre la convergence complète et la convergence presque sure et en probabilité

Proposition 2.2.

Si $\lim_{n \rightarrow \infty} X_n = X$, p. co, alors nous avons :

1. $\lim_{n \rightarrow \infty} X_n = X$, p.
2. $\lim_{n \rightarrow \infty} X_n = X$, p. s.

Démonstration.

1. On suppose que

$$\lim_{n \rightarrow \infty} X_n = X, p.co \Leftrightarrow \forall \varepsilon > 0, \sum_{n \in \mathbb{N}} P(|X_n - X| > \varepsilon) < \infty. \quad (2.1)$$

On a : $\forall \varepsilon > 0, \sum_{n \in \mathbb{N}} P(|X_n - X| > \varepsilon)$ c'est une somme infinie des probabilités.

Alors si cette somme est finie, il faut que :

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) \rightarrow 0. \quad (2.2)$$

Donc :

$$\lim_{n \rightarrow \infty} X_n = X, p. \quad (2.3)$$

2. On a

$$\lim_{n \rightarrow \infty} X_n = X, p.co \Leftrightarrow \forall \varepsilon > 0, \sum_{n \in \mathbb{N}} P(|X_n - X| > \varepsilon) < \infty.$$

On applique le lemme de Borel Cantelli, on obtient :

$$P(\limsup_{n \rightarrow \infty} \{|X_n| > \varepsilon\}) = 0,$$

qui peut être réécrit comme :

$$P(A_\varepsilon) = 1, \text{ où } A(\varepsilon) = \{\exists n, \forall m > n, |X_m| \leq \varepsilon\},$$

notez que $(A(\varepsilon))_\varepsilon$ est une séquence d'événement emboîtés, donc : $P(A_\varepsilon) = 1$ implique directement la convergence presque sure, à savoir :

$$P(\forall \varepsilon, \exists n, \forall m > n, |X_m| \leq \varepsilon) = 1.$$

■

Théorème 2.3.

soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, et c une constante.
Si $X_n \xrightarrow{p. s.} c$, quand $n \rightarrow \infty$, alors $X_n \xrightarrow{p. co} c$ quand, $n \rightarrow \infty$.

Démonstration.

Supposons que : $X_n \xrightarrow{p. s.} c$ est une constante.

Ensuite, pour chaque $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_m - c| \leq \varepsilon, \forall m \geq n) = 1,$$

équivalent à :

$$\lim_{n \rightarrow \infty} P(|X_m - c| > \varepsilon, \exists m \geq n) = 0,$$

Notez que

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|X_m - c| > \varepsilon, \exists m \geq n) &= \lim_{n \rightarrow \infty} P(\cup_{m=n}^{\infty} |X_m - c| > \varepsilon) \\ &= P(\cap_{n=1}^{\infty} \cup_{m=n}^{\infty} |X_m - c| > \varepsilon) \\ &= P(\limsup_{n \rightarrow \infty} \{|X_m - c| > \varepsilon\}) \end{aligned}$$

où le fait que :

$$\{(\cup_{m=n}^{\infty} |X_m - c| > \varepsilon)\}_{n=1}^{\infty}.$$

est une suite d'événements monotones et décroissants. Notez maintenant que depuis $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes, il s'en suit que $(|X_n - c| > \varepsilon)_{n \geq 1}$ est une suite de variables aléatoires indépendantes pour chaque $\varepsilon > 0$. Alors :

$$P(\limsup_{n \rightarrow \infty} P\{|X_m - c| > \varepsilon\}) = 0$$

pour chaque $\varepsilon > 0$, alors :

$$\sum_{n=1}^{\infty} P(|X_n - c| > \varepsilon) < \infty$$

pour tout $\varepsilon > 0$ alors $X_n \xrightarrow{p. co} c$, quand, $n \rightarrow \infty$.

■

Théorème 2.4.

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires convergentes en probabilité vers une variable aléatoire X . Alors, il existe une suite non-décroissante d'entiers positifs $(n_k)_{k \geq 1}$ tel que $X_{n_k} \xrightarrow{p. co} X$ et $X_{n_k} \xrightarrow{p} X$ quand, $k \rightarrow \infty$.

Démonstration.

Voir [1] ■

2.1.2 Taux de convergence

le taux de convergence presque sure à 0 pour une suite des variable aléatoire est définie par :

$$\begin{aligned} X_n - X = O_{p.s.}(u_n) &\Leftrightarrow P((X_n - X) = O_{u_n}) = 1. \\ &\Leftrightarrow P(\exists c < \infty, \exists n, \forall m > n, |X_n - X| \leq cu_m) = 1. \end{aligned}$$

et le taux de convergence en probabilité définie par :

$$X_n - X = O_p(u_n) \Leftrightarrow \lim_m \limsup_{n \rightarrow \infty} P(|X_n - X| < mu_n) = 1.$$

Maintenant on définit le taux de la convergence presque complète :

Définition 2.5.

On dit que le taux de convergence presque complète des $(X_n)_{n \in \mathbb{N}}$ vers X est l'ordre u_n , si et seulement si :

$$\forall \varepsilon_0 > 0, \sum_{n \in \mathbb{N}} P(|X_n - X| > \varepsilon_0 u_n) < \infty.$$

et on écrit : $X_n - X = O_{p.co}(u_n)$

Proposition 2.6.

supposons que : $X_n - X = O_{p.co}(u_n)$, alors :

1. $X_n - X = O_p(u_n)$.
2. $X_n - X = O_{p.s.}(u_n)$.

Démonstration.

1. la définition de $O_{p.co}$ permet d'écrire que :

$$\exists m_0, \forall m > m_0, \sum_{n \in \mathbb{N}} P(|X_n| > mu_n) < \infty.$$

D'après le lemme de Borel Cantelli :

$$\exists m_0, \forall m > m_0, P(\limsup_{n \rightarrow \infty} \{|X_n| > mu_n\}) = 0.$$

En appliquant maintenant le lemme de Fatou :

On obtient :

$$\begin{aligned} & \exists m_0, \forall m > m_0, \limsup_{n \rightarrow \infty} P(|X_n| > mu_n) = 0 \\ \Leftrightarrow & \exists m_0, \forall m > m_0, \liminf_{n \rightarrow \infty} P(|X_n| \leq mu_n) = 1. \end{aligned}$$

ce qui implique directement que : $X_n = O_p(u_n)$

2. Comme précédemment :

en appliquant le lemme de Borelle Cantelli, on obtient que :

$$P(\limsup_{n \rightarrow \infty} |X_n| > \varepsilon_0 u_n) = 0$$

qui peut être réécrit comme :

$$P(B(\varepsilon)) = 1 \text{ où } B(\varepsilon) = \{\exists n, \forall m > n, |X_m| \leq \varepsilon_0 u_m\}$$

avec : $(B(\varepsilon))_\varepsilon$ est une séquence d'événement emboîtés, et donc :

$$\exists \varepsilon > 0, P(B(\varepsilon_0)) = 1$$

implique :

$$\begin{aligned} \forall \varepsilon_0 > 0, P(\exists n, \forall m > n, |X_m| \leq \varepsilon_0 u_m) = 1 & \Leftrightarrow P(X_n = O(u_n)) = 1 \\ & \Leftrightarrow X_n = O_{p.s.}(u_n). \end{aligned}$$

■

2.1.3 Opération de calcul

Proposition 2.7.

Supposons que $\lim_{n \rightarrow \infty} u_n = 0$, $\lim_{n \rightarrow \infty} X_n = L_x$ p. co et $\lim_{n \rightarrow \infty} Y_n = L_y$, p. co, où L_x et L_y sont deux nombres réels :

1. Nous avons :

- a) $\lim_{n \rightarrow \infty} (X_n + Y_n) = L_x + L_y$ p. co
- b) $\lim_{n \rightarrow \infty} (X_n Y_n) = L_x L_y$ p. co
- c) $\lim_{n \rightarrow \infty} \frac{1}{Y_n} = \frac{1}{L_y}$, p. co, tant que : $L_y \neq 0$.
2. Si $X_n - L_x = O_{p.co}(u_n)$ et $Y_n - L_y = O_{p.co}(u_n)$ nous avons :
- a) $(X_n + Y_n) - (L_x + L_y) = O_{p.co}(u_n)$
- b) $X_n Y_n - L_x L_y = O_{p.co}(u_n)$
- c) $\frac{1}{y_n} - \frac{1}{L_y} = O_{p.co}(u_n)$, tant que $L_y \neq 0$.

Démonstration.

1.a)

$$\begin{aligned} P(|(X_n + Y_n) - (L_x + L_y)| > \varepsilon) &= P(|(X_n - L_x) + (Y_n - L_y)| > \varepsilon) \\ &\leq P(|X_n - L_x| > \frac{\varepsilon}{2}) + P(|Y_n - L_y| > \frac{\varepsilon}{2}) \\ &= 0. \end{aligned}$$

donc : $\lim_{n \rightarrow \infty} (X_n + Y_n) = L_x + L_y$

2.a)

$$\begin{aligned} P(|(X_n + Y_n) - (L_x + L_y)| > \varepsilon_0 u_n) &\leq P(|X_n - L_x| > \frac{\varepsilon_0}{2} u_n) + P(|Y_n - L_y| > \frac{\varepsilon_0}{2} u_n) \\ &= \frac{1}{2} P(|X_n - L_x| > \varepsilon_0 u_n) + \frac{1}{2} P(|Y_n - L_y| > \varepsilon_0 u_n) \\ &= \frac{1}{2} (O_{p.co}(u_n) + O_{p.co}(u_n)) \\ &= O_{p.co}(u_n). \end{aligned}$$

Donc : $(X_n + Y_n) - (L_x + L_y) = O_{p.co}(u_n)$ 1.b) Sans perte de généralité, on pose $l_x = 0$. La décomposition suivante

$$X_n Y_n = X_n (Y_n - l_y) + X_n l_y,$$

nous donne

$$\begin{aligned} P(|(X_n Y_n)| > \varepsilon) &\leq P(|(Y_n - L_y)| |X_n| > \frac{\varepsilon}{2}) + P(|L_y X_n| > \frac{\varepsilon}{2}) \\ &\leq P(|(Y_n - L_y)| > \sqrt{\frac{\varepsilon}{2}}) + P(|X_n| > \sqrt{\frac{\varepsilon}{2}}) + P(|L_y X_n| > \frac{\varepsilon}{2}) \\ &\leq P(|(Y_n - L_y)| > \frac{\varepsilon}{2}) + P(|X_n| > \frac{\varepsilon}{2}) + P(|L_y X_n| > \frac{\varepsilon}{2}). \end{aligned}$$

L'inégalité précédente et la convergence presque complète de X_n et Y_n permettent d'écrire

$$\sum_{n \in \mathbb{N}} P(|X_n Y_n| > \varepsilon) < \infty$$

2.b) Il suffit d'appliquer le résultat précédent à $\varepsilon = \varepsilon_0 U_n$

1.c) La convergence presque complète de Y_n à $L_y \neq 0$, implique qu'il existe quelques $\delta > 0$ (choisissez par exemple $\delta = \frac{L_y}{2}$) tels que : $\sum_{n \in \mathbb{N}} p(|Y_n| \leq \delta) < \infty$.

$$\begin{aligned} P\left(\left|\frac{1}{Y_n} - \frac{1}{L_y}\right| > \varepsilon\right) &= P(|Y_n - L_y| > \varepsilon |L_y Y_n|). \\ &\leq P(|Y_n - L_y| > \varepsilon |L_y Y_n| \text{ et } |Y_n| > \delta) + p(|Y_n| < \delta). \\ &\leq P(|Y_n - L_y| > \varepsilon \delta |L_y|) + p(|Y_n| < \delta). \\ &\leq 0 + 0. \\ &= 0. \end{aligned}$$

Alors : $\lim_{n \rightarrow \infty} \frac{1}{Y_n} = \frac{1}{L_y}$ p. co, $L_y \neq 0$.

2.c)

$$\begin{aligned} P\left(\left|\frac{1}{Y_n} - \frac{1}{L_y}\right| > \varepsilon_0 u_n\right) &= P(|Y_n - L_y| > \varepsilon_0 u_n |L_y Y_n|). \\ &\leq P(|Y_n - L_y| > \varepsilon_0 u_n |L_y Y_n| \text{ et } |Y_n| > \delta) + P(|Y_n| \leq \delta). \\ &\leq P(|Y_n - L_y| > \varepsilon_0 \delta u_n |L_y|) + P(|Y_n| \leq \delta). \\ &\leq 0 + 0. \\ &= 0. \end{aligned}$$

Alors : $\frac{1}{Y_n} - \frac{1}{L_y} = O_{p.co}(u_n)$, telle que : $L_y \neq 0$.

■

Proposition 2.8.

Supposons que $\lim_{n \rightarrow \infty} u_n = 0$, $X_n = O_{p.co} u_n$ et $\lim_{n \rightarrow \infty} Y_n = L_y$ p. co, où : L_y est une nombre réel déterministe.

1. On a :

$$X_n Y_n = O_{p.co}(u_n) \tag{2.4}$$

2. on a :

$$\frac{X_n}{Y_n} = O_{p.co}(u_n), \text{ avec : } L_y \neq 0. \tag{2.5}$$

Démonstration.

1. La convergence presque complète de Y_n à $L_y \neq 0$ implique qu'il existe quelque $\delta > 0$, tels que : $\sum_{n \in \mathbb{N}} P(|Y_n| > \delta) < \infty$ alors :

$$\begin{aligned} P(|Y_n X_n| > \varepsilon u_n) &= P(|Y_n X_n| > \varepsilon u_n, \text{ et } |Y_n| \leq \delta) + p(|Y_n X_n| > \varepsilon u_n \text{ et } |Y_n| > \delta) \\ &\leq P(|X_n| > \frac{\varepsilon u_n}{|Y_n|}, \text{ et } |Y_n| \leq \delta) + p(|Y_n X_n| > \varepsilon u_n \text{ et } |Y_n| > \delta) \\ &\leq P(|X_n| > \varepsilon \frac{1}{\delta} u_n) + P(|Y_n| > \delta) \end{aligned}$$

d'après les deux inégalités précédemment ainsi que l'hypothèse que $O_{p.co}(u_n)$, suffit pour montrer que : $X_n Y_n = O_{p.co}(u_n)$.

2. On a : $\frac{X_n}{Y_n} = X_n \frac{1}{Y_n}$ d'autre partie, on a :

$$\lim_{n \rightarrow \infty} \frac{1}{Y_n} = \frac{1}{L_y} \quad (2.6)$$

p. co, telle que : $L_y \neq 0$ et $X_n = O_{p.co}(u_n)$ alors d'après (1) de proposition précédent, en déduire que : $\frac{X_n}{Y_n} = O_{p.co}(u_n)$, telle que : $L_y \neq 0$.

■

2.2 Exemples

Exemple 2.9.

Soit u un variable aléatoire uniforme sur $[0, 1]$ et définit une suite des variables aléatoires $(X_n)_{n \geq 1}$ telle que : $X_n = \delta\{u, (0, \frac{1}{n^2})\}$

où δ est la fonction indicatrice définie par : $\delta\{u, (0, n^{-2})\} = \begin{cases} 1 & u \in (0, n^{-2}) \\ 0 & \text{sinon.} \end{cases}$

Soit $\varepsilon > 0$, alors : $P(X_n > \varepsilon) = \begin{cases} 0 & \varepsilon \geq 1. \\ n^{-2} & \varepsilon < 1. \end{cases}$

par conséquent, pour chaque $\varepsilon > 0$:

$$\sum_{n=0}^{\infty} p(X_n > \varepsilon) \leq \sum_{n=0}^{\infty} n^{-2} < \infty.$$

car : $\sum_{n=0}^{\infty} n^{-2} = \sum_{n=0}^{\infty} \frac{1}{n^2}$, est une série de Riemann ($2 > 1$) d'où $X_n \xrightarrow{P.co} 0$ quand $n \rightarrow \infty$.

Exemple 2.10.

Soit $(X_n)_{n \geq 1}$ une suite des variables aléatoires telle que $E(X_n) = C$, pour tout $n \in \mathbb{N}^*$, où $C \in \mathbb{R}$ est une constante qui ne dépend pas de n .

Supposons en outre que : $V(X_n) = n^{-2}\Gamma$, où Γ est une constante positive fini ne dépend pas de n , dans ces conditions notez que pour tout $\varepsilon > 0$:

$$p(|X - C| > \varepsilon) \leq \varepsilon^{-2}V(X_n) = n^{-2}\varepsilon^{-2}\Gamma$$

par conséquent :

$$\forall \varepsilon > 0 : \varepsilon > 0 : \sum_{n=1}^{\infty} n^{-2}\varepsilon^{-2}\Gamma \leq \varepsilon^{-2}\Gamma \sum_{n=1}^{\infty} n^{-2} < \infty$$

car :

$$\sum_{n=1}^{\infty} n^{-2} = \sum_{n=1}^{\infty} \frac{1}{n^2}, \quad (2.7)$$

on a ($2 > 1$), alors cette série est une série de Riemann converge et $\varepsilon^{-2}\Gamma$ est une constant, donc : $X_n \xrightarrow{P. co} C$, lorsque $n \rightarrow \infty$.

2.3 Convergence presque complète de l'estimateur de la fonction de répartition

Dans cette section on étudie la convergence presque complète de la fonction empirique vers la fonction de répartition.

Théorème 2.11.

Soit X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F , alors

$$\hat{F}_n(x) - F(x) = O\left(\sqrt{\frac{\log(n)}{n}}\right) \text{ p. co}$$

Démonstration.

Par définition, il suffit de montrer que :

$$\exists \varepsilon > 0, \sum_{n=0}^{\infty} P\left[|\hat{F}_n(x) - F(x)| > \varepsilon \left(\sqrt{\frac{\log(n)}{n}}\right)\right] < \infty$$

Soit $\varepsilon > 0$ on a :

$$P\left[|\hat{F}_n(x) - F(x)| > \varepsilon \left(\sqrt{\frac{\log(n)}{n}}\right)\right] = P\left[|\hat{F}_n(x) - E(\hat{F}_n(x))| > \varepsilon \left(\sqrt{\frac{\log(n)}{n}}\right)\right]$$

Par définition de l'estimateur de $\hat{F}_n(x)$ on a

$$\begin{aligned} P \left[\left| \frac{1}{n} \sum_{i=1}^n 1_{] -\infty, x]}(X_i) - \frac{1}{n} \sum_{i=1}^n E(1_{] -\infty, x]}(X_i)) \right| > \varepsilon \left(\sqrt{\frac{\log(n)}{n}} \right) \right] = \\ P \left[\frac{1}{n} \sum_{i=1}^n |1_{] -\infty, x]}(X_i) - E(1_{] -\infty, x]}(X_i))| > \varepsilon \left(\sqrt{\frac{\log(n)}{n}} \right) \right] = \\ P \left[\sum_{i=1}^n |1_{] -\infty, x]}(X_i) - E(1_{] -\infty, x]}(X_i))| > \varepsilon(\sqrt{n \log(n)}) \right]. \end{aligned}$$

En appliquant l'inégalité de Brrenstein-Frechet avec $\alpha_i = 0$, $\beta_i = 1$, $X_i = 1_{] -\infty, x]}(X_i)$ et $t = \varepsilon(\sqrt{n \log(n)})$.

il vient

$$\begin{aligned} P \left[|\hat{F}_n(x) - F(x)| > \varepsilon \left(\sqrt{\frac{\log(n)}{n}} \right) \right] &\leq 2 \exp \left(\frac{-2\varepsilon^2 n \log n}{n} \right) \\ &\leq 2 \exp(\log n^{-2\varepsilon^2}) \\ &= n^{-2\varepsilon^2}. \end{aligned}$$

Il suffit de prendre $2\varepsilon^2 > 1$ ■

Nous allons donner deux versions des inégalités exponentielles de type Bernstein qui nous seront utiles pour l'établissement des résultats que nous avons choisi de reprendre. Nous supposons que X_1, X_2, \dots, X_n est une suite de variables aléatoires réelles, indépendantes et centrées.

Corollaire 2.12.

- a) *S'il existe une constante positive $m < 2$, il existe un réel C_m strictement positif et une constante a positive, tels que :*

$$E(|X_1^m|) \leq C_m a^{2(m-1)},$$

alors on a

$$\forall \varepsilon \geq 0, P \left[\left| \sum_{i=1}^n X_i \right| > \varepsilon n \right] \leq 2 \exp \left\{ \frac{-\varepsilon^2 n}{2a^2(1 + \varepsilon)} \right\}.$$

- b) *Supposons que les $(X_i)_{1 \leq i \leq n}$ dépendent de n .*

Si pour tout $m < 2$, il existe un réel C_m strictement positif et une suite (a_n) de réels positifs, tels que :

$$E(|X_1^m|) \leq C_m a_n^{2(m-1)},$$

et si

$$U_n = n^{-1} a_n^2 \log n, \text{ vérifie } \lim_{n \rightarrow \infty} U_n = 0.$$

Alors on a

$$\frac{1}{n} \sum_{i=1}^n X_i = O_{p.co}(\sqrt{U_n}).$$

Tandis que ce résultat s'applique à des variables dont on a majoré les moments d'ordre m , le corollaire suivant est donné pour des variables identiquement distribuées et bornées.

Corollaire 2.13.

a) S'il existe une constante positive $M < \infty$, telle que :

$$|X_1| \leq M.$$

alors on a :

$$\forall \varepsilon \geq 0, P \left[\left| \sum_{i=1}^n X_i \right| > \varepsilon n \right] \leq 2 \exp \left\{ \frac{-\varepsilon^2 n}{2\sigma^2(1 + \frac{M\varepsilon}{\sigma^2})} \right\}$$

où

$$\sigma^2 = E(X_1^2)$$

b) Supposons que les $(X_i)_{1 \leq i \leq n}$ dépendent de n et que $\sigma_n^2 = E(X_i^2)$,
, s'il existe $M = M_n < \infty$ telle que :

$$|X_1| \leq M.$$

et

$$\frac{M}{\sigma_n^2} \leq C < \infty.$$

et si

$$U_n = n^{-1} \sigma_n^2 \log n, \text{ vérifie } \lim_{n \rightarrow \infty} U_n = 0.$$

Alors on a

$$\frac{1}{n} \sum_{i=1}^n X_i = O_{p.co}(\sqrt{U_n}).$$

L'étude de la convergence, presque complète des estimateurs à noyau dans le cadre de la variable aléatoire fonctionnelle introduite par Ferraty et Vieu (2000), s'inspire énormément du cas réel exposé ci dessous. Cette convergence implique la convergence en probabilité et la convergence presque sûre mais n'implique pas la convergence en moyenne quadratique, comme le montre l'exemple suivant : Soit l'espace de probabilité $([0, 1], B([0, 1]), \lambda)$ où $B([0, 1])$ désigne la tribu borélienne de $[0, 1]$ et λ est la restriction de la mesure de Lebesgue à cet ensemble. La suite $f_n = n1_{[0, \frac{1}{n^2}]}$ converge presque complètement vers 0 mais ne converge pas en moyenne quadratique et inversement la suite $g_n = 1_{[0, \frac{1}{n}]}$ converge en moyenne quadratique vers 0 mais ne converge pas presque complètement.

2.4 Convergence presque complète de l'estimateur de la densité

Ici, on étudie la convergence complète de l'estimateur à noyau de la fonction de densité. Soit f , la fonction densité de probabilité supposée inconnue et \hat{f}_{h_n} sont estimateur à noyau. Les hypothèses qui nous donne la convergence complète de cet estimateur sont les suivantes :

hypothèses :

1. f est continue au voisinage de x , un point fixé de R .
2. le paramètre de lissage h_n est tel que :

$$\lim_{n \rightarrow \infty} h_n = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{\log n}{nh_n} = 0$$

3. Le noyau k est tel que :

k est d'ordre i au sens de Gasser c'est à dire :

$$\int t^j k(t) dt = 0, \forall j = 1, 2, \dots, i - 1. \text{ et } 0 < \left| \int t^i k(t) dt \right| < \infty.$$

(K. 7) k est borné, intégrable et à support compact.

Sous les hypothèses précédents, on obtient la convergence presque complète de \hat{f}_{h_n} vers f , comme le donne la théorème suivant :

Théorème 2.14.

Si les hypothèses (K. 7), 1 et 2 sont satisfait, alors :

$$\lim_{n \rightarrow \infty} \hat{f}_{h_n}(x) = f(x) \text{ p. co}$$

Démonstration.

Il suffit démontrer que :

$$\forall \varepsilon > 0, \sum_{n \in \mathbb{N}} P(|\hat{f}_{h_n}(x) - f(x)| > \varepsilon) < \infty$$

La démonstration de ce théorème est basée sur la décomposition suivante :

$$\begin{aligned} \hat{f}_{h_n}(x) - f(x) &= (\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))) - (f(x) - E(\hat{f}_{h_n}(x))). \\ P[|\hat{f}_{h_n}(x) - f(x)|] &= P[|\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)) - (f(x) - E(\hat{f}_{h_n}(x)))|] \\ &\leq P[|\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))|] + P[|f(x) - E(\hat{f}_{h_n}(x))|] \end{aligned}$$

Le résultat du théorème découle alors des deux lemmes suivants.

Lemme 2.15.

Si les hypothèses (K. 7), 1 et 2 sont vérifiées on a

$$\lim_{n \rightarrow \infty} E(\hat{f}_{h_n}(x)) = f(x).$$

Démonstration.

Nous avons :

$$\begin{aligned} E(\hat{f}_{h_n}(x)) &= \frac{1}{h_n} \int_{-\infty}^{+\infty} k\left(\frac{x-t}{h_n}\right) f(t) dt. \\ &= \int_{-\infty}^{+\infty} k(z) f(x - zh_n) dz. \end{aligned}$$

La continuité uniforme de f sur le support compact de K entraîne $f(x - zh_n) \rightarrow f(x)$, uniformément en z .

D'où

$$\lim_{n \rightarrow \infty} E(\hat{f}_{h_n}(x)) = f(x).$$

■

Lemme 2.16.

Sous les hypothèses (K. 7), 1 et 2 on a :

$$\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)) = O_{p.co} \left(\sqrt{\frac{\log(n)}{n}} \right).$$

Démonstration.

$$\begin{aligned}\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} \left[k \left(\frac{x - X_i}{h_n} \right) \right] - E \left(k \left(\frac{x - X_i}{h_n} \right) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \Gamma_i,\end{aligned}$$

où $\Gamma_i = \frac{1}{h_n} \left[k \left(\frac{x - X_i}{h_n} \right) \right] - E \left(k \left(\frac{x - X_i}{h_n} \right) \right)$.

En utilisant l'hypothèse (K. 7), on a

$$\begin{aligned}\left| k \left(\frac{x - X_i}{h_n} \right) \right| &\leq M \\ \frac{\left| k \left(\frac{x - X_i}{h_n} \right) \right| - E \left(k \left(\frac{x - X_i}{h_n} \right) \right)}{h_n} &\leq \frac{M - E \left(k \left(\frac{x - X_i}{h_n} \right) \right)}{h_n},\end{aligned}$$

alors $|\Gamma_i| \leq \frac{c}{h_n}$, telle que

$$c = M - E \left(k \left(\frac{x - X_i}{h_n} \right) \right).$$

D'autre part le changement de variable $z = \frac{x-t}{h_n}$ nous donne

$$\begin{aligned}\frac{1}{h_n} E \left[\frac{1}{h_n} k^2 \left(\frac{x - X}{h_n} \right) \right] &= \frac{1}{h_n^2} \int_{-\infty}^{+\infty} k^2 \left(\frac{x - t}{h_n} \right) f(t) dt \\ &= \frac{1}{h_n} \int k^2(z) f(x - zh_n) dz.\end{aligned}$$

On remarque que :

$$\begin{aligned}E(\Gamma_i^2) &= \text{var} \left(\frac{1}{h} k \left(\frac{x - X_i}{h_n} \right) \right) \\ &\leq E(v_i^2),\end{aligned}$$

avec

$$v_i = \frac{1}{h_n} k \left(\frac{x - X_i}{h_n} \right).$$

Comme k est bornée et f est continue sur le support compact de k .

$$\begin{aligned}\left| k \left(\frac{x - X_i}{h_n} \right) \right| \leq M &\implies \left| k \left(\frac{x - X_i}{h_n} \right) \right|^2 \leq M^2 \\ &\implies \frac{\left| k \left(\frac{x - X_i}{h_n} \right) \right|^2}{h_n^2} \leq \frac{M^2}{h_n^2} \\ &\implies E(v_i^2) \leq \frac{M^2}{h_n^2},\end{aligned}$$

on a l'existence d'une constante C , telle que :

$$E(\Gamma_i^2) \leq \frac{C}{h_n}, C = \frac{M^2}{h_n}.$$

On obtient alors, en appliquant le corollaire (2. 13) de l'inégalité exponentielle de type Bernstein,

$$\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)) = O_{\text{p.co}} \left(\sqrt{\frac{\log(n)}{n}} \right)$$

■

Ce résultat est plus fort que le résultat demandé.

En remplaçant l'hypothèse 1 par l'hypothèses suivante

4. f est k fois continûment dérivable autour du point x .

On obtient une vitesse de convergence presque complète ponctuelle de l'estimateur à noyau.

Théorème 2.17.

Sous les hypothèses (K. 7), (2) et (4) on a :

$$\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)) = O(h_n^k) + O \left(\sqrt{\frac{\log(n)}{n}} \right).$$

Démonstration.

En reprenant la décomposition de la preuve précédente, le résultat du théorème sera établi par les lemmes précédents et le suivant :

Lemme 2.18.

Sous les hypothèses 2, 3 et 4 on a :

$$\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x)) = O(h_n^k). \quad (2.8)$$

Démonstration.

$$\begin{aligned} E(\hat{f}_{h_n}(x)) &= \frac{1}{h_n} \int_{-\infty}^{+\infty} k \left(\frac{x-t}{h_n} \right) f(t) dt. \\ &= \int_{-\infty}^{+\infty} k(z) f(x - zh_n) dz. \end{aligned}$$

L'hypothèse 4 nous permet de développer f au voisinage de x (Taylor-Lagrange).

$$f(x - zh_n) = f(x) + \sum_{i=1}^{k-1} \frac{(-1)^i (zh_n)^i}{i!} f^{(i)}(x) + \frac{(-1)^k (zh_n)^k}{k!} f^{(k)}(\theta_z),$$

où θ_z entre x et $x - zh_n$.

D'autre part l'hypothèse 3 sur l'ordre de k au sens de Gasser et Müller nous donne :

$$E(\hat{f}_{h_n}(x)) = f(x) + \frac{(-1)^k h_n^k}{k!} \int z^k f^{(k)}(\theta_z) dz.$$

La compacité du support de k et la condition de 4 impliquent la convergence uniforme en z de $f^{(k)}(z)$ vers $f^{(k)}(x)$, d'où

$$E(\hat{f}_{h_n}(x)) - f(x) = (-1)^k h_n^k \int z^k k(z) \frac{f^{(k)}}{k!} dz + O(h_n^k).$$

Cette relation permet d'achever la preuve du théorème. ■

Le résultat du théorème précédent peut être établi uniformément. Il suffit de conserver toutes les hypothèses et de donner une autre version de l'hypothèse 4.

5. Choisir un compact S de R sur lequel f est k fois continûment dérivable.

et de rajouter l'hypothèse de type Lipschitz sur le noyau k .

- il existe $C < 1$, $\forall x \in S, \forall y \in S$.

$$|k(x) - k(y)| \leq C|x - y|. \quad (2.9)$$

Nous avons aussi besoin de l'hypothèse suivante sur le paramètre de lissage,

- il existe $\varepsilon < \infty$, telle que :

$$\lim_{n \rightarrow \infty} n^{2\varepsilon-1} h_n = +\infty \quad (2.10)$$

■

Théorème 2.19.

Sous les hypothèses (K. 7), 2, 3, 5, (2. 8), (2. 9), on a :

$$\sup(|\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))|) = O(h_n^k) + O\left(\sqrt{\frac{\log(n)}{n}}\right), \text{ p. co.}$$

Démonstration.

$$\text{Sup}_{x \in S} \hat{f}_{h_n}(x) - f(x) \leq \text{Sup}_{x \in S} [|\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))|] + \text{Sup}_{x \in S} [(f(x) + E(\hat{f}_{h_n}(x)))].$$

Nous remarquons que grâce à l'hypothèse 5 et à la compacité du support du noyau K les étapes de la démonstration utilisée dans le (2. 8) peuvent être faites uniformément en $x \in S$. D'où

$$\text{Sup}_{x \in S} |\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))| = o(h_n^k).$$

Il reste à montrer l'égalité suivante :

$$\text{Sup}_{x \in S} |\hat{f}_{h_n}(x) - E(\hat{f}_{h_n}(x))| = O\left(\sqrt{\frac{\log(n)}{n}}\right).$$

S est un compact de R , il existe donc un recouvrement fini de S tel que :

$$S \subset \cup_{k=1}^{z_n} S_k$$

où

$$S_k =]t_k - l_n, t_k + l_n[,$$

avec

$$l_n = n^{-2\varepsilon}, l_n = Cz_n^{-1},$$

et

$$t_x = \text{argmin}_{t \in t_1, t_2, \dots, t_{z_n}} |x - t|.$$

On a :

$$\begin{aligned} \text{Sup}_{x \in S} |E[\hat{f}_{h_n}(x)] - \hat{f}_{h_n}(x)| &\leq \text{Sup}_{x \in S} [|\hat{f}_{h_n}(x) - \hat{f}_{h_n}(t_x)|] \\ &\quad + \text{Sup}_{x \in S} [|\hat{f}_{h_n}(t_x) - E[\hat{f}_{h_n}(t_x)]|] \\ &\quad + \text{Sup}_{x \in S} |E[\hat{f}_{h_n}(x)] - E[\hat{f}_{h_n}(t_x)]|. \end{aligned}$$

Comme K est Lipschitzien, l'hypothèse (2. 9) nous donne

$$\begin{aligned} \text{Sup}_{x \in S} [|\hat{f}_{h_n}(x) - \hat{f}_{h_n}(t_x)|] &\leq \text{Sup}_{x \in S} \frac{1}{nh_n} \sum_{i=1}^n \left| k\left(\frac{X_i - x}{h_n}\right) - k\left(\frac{X_i - t_x}{h_n}\right) \right| \\ &\leq \text{Sup}_{x \in S} \frac{c}{h_n^2} |x - t_x| \\ &\leq \frac{l_n c}{h_n^2} \\ &\leq \frac{c}{(h_n n^\varepsilon)^2} = o\left(\frac{\log n}{nh_n}\right). \end{aligned}$$

et de manière évidente on a l'existence d'une constante c telle que :

$$\text{Sup}_{x \in S} |E[\hat{f}_{h_n}(x)] - E[\hat{f}_{h_n}(t_x)]| \leq \frac{c}{(h_n n^\varepsilon)^2}.$$

En ce qui concerne le terme $\text{Sup}_{x \in S} |E[\hat{f}_{h_n}(x)] - E[\hat{f}_{h_n}(t_x)]|$ on a pour tout $\varepsilon > 0$:

$$P[\text{Sup}_{x \in S} |[\hat{f}_{h_n}(t_x)] - E[\hat{f}_{h_n}(t_x)]| > \varepsilon] = P[\max_{k=1, \dots, z_n} |\hat{f}_{h_n}(t_k) - E[\hat{f}_{h_n}(t_k)]| > \varepsilon]. \quad (2.11)$$

$$(\text{ car } \sup_{k=1}^{z_n} = \max\{\text{sup}_{s_1}, \dots, \text{sup}_{s_{z_n}}\}) \quad (2.12)$$

$$\leq z_n P\left[\frac{1}{n} \sum_{i=1}^n |u_i - E[u_i]| > \varepsilon\right], \quad (2.13)$$

$$(2.14)$$

où $u_i = \frac{1}{h_n} k \left(\frac{X_i - t_k}{h_n} \right)$.

En suivant les mêmes étapes que la preuve du lemme (45), nous pouvons alors majorer aisément u_i et $E[u_i^2]$, comme suit

$$u_i \leq \frac{c}{h_n} \text{ et } E[u_i^2] \leq \frac{c}{h_n}.$$

Nous somme en position d'appliquer l'inégalité de type exponentielle de Bernstein, énoncer dans le corollaire (2. 13), cette inégalité associée à (2. 13)

$$\begin{aligned} P[\text{Sup}_{x \in S} |[\hat{f}_{h_n}(t_x)] - E[\hat{f}_{h_n}(t_x)]| > \varepsilon] &\leq z_n \exp(-cn\varepsilon^2 h_n). \\ &\leq n^{2\varepsilon} \exp(-cn\varepsilon^2 h_n). \end{aligned}$$

En posant $\varepsilon = \varepsilon_0 \sqrt{\left(\frac{\log n}{nh_n}\right)}$ on obtient

$$\sum P[\text{sup}_{x \in R} |[\hat{f}_{h_n}(t_x)] - E[\hat{f}_{h_n}(t_x)]| > \varepsilon_0 \sqrt{\left(\frac{\log n}{nh_n}\right)}]$$

pour ε_0 choisi suffisamment grand. ■

■

2.5 Convergence presque complète de l'estimateur de la fonction de régression

En se basant sur la preuve donnée dans Ferraty et Vieu (2003), nous traitons dans ce paragraphe la convergence presque complète de l'estimateur à noyau de la fonction de régression, auxquelles nous rajoutons les hypothèses suivantes :

6. f, r sont des fonctions continues au voisinage de x , un point fixé de R .

La densité f et la variable Y sont telles que :

$$f(x) > 0, \quad (2.15)$$

et

$$|Y| < M < \infty, \quad (2.16)$$

où M est une constante réelle positive.

Théorème 2.20.

Sous les hypothèses 2, 6, (2. 15), (2. 16) (K. 1) et (K. 7) on a

$$\lim_{n \rightarrow \infty} \hat{r}_{h_n}(x) = r(x) \text{ p.co}$$

Démonstration.

La démonstration de ce théorème est basée sur la décomposition suivante :

$$\begin{aligned} \hat{r}_{h_n}(x) - r(x) &= \frac{1}{\hat{f}_n(x)} [(\hat{\phi}_{h_n}(x) - E(\hat{\phi}_{h_n}(x))) + (E(\hat{\phi}_{h_n}(x)) - \phi(x))] \\ &\quad + [(f(x) - E(\hat{f}_{h_n}(x))) + (E(\hat{f}_{h_n}(x)) - \hat{f}_{h_n}(x))] \frac{r(x)}{\hat{f}_{h_n}(x)}, \end{aligned}$$

où $\phi(x) = f(x)r(x)$. Le résultat énoncé découle des lemmes suivants : ■

Lemme 2.21.

D'après les hypothèses 6, (2. 16), (K. 1) et (K. 7) on a

$$\lim_{n \rightarrow \infty} E(\hat{\phi}_{h_n}(x)) = \phi(x)$$

Démonstration.

on a :

$$\begin{aligned} E(\hat{\phi}_{h_n}(x)) &= E \left[\frac{1}{nh_n} \sum_{i=1}^n Y_i k \left(\frac{x - X_i}{h_n} \right) \right] \\ &= \frac{1}{h_n} E \left[Y k \left(\frac{x - X}{h_n} \right) \right]. \end{aligned}$$

Le conditionnement par rapport à $X = x$ nous donne :

$$E[\hat{\phi}_{h_n}(x)] = \frac{1}{h_n} E \left[r(x) k \left(\frac{x - X}{h_n} \right) \right],$$

où :

$$\begin{aligned} E(\hat{\phi}_{h_n}(x)) &= \frac{1}{h_n} \int r(t)k\left(\frac{x-t}{h_n}\right) f(t)dt \\ &= \frac{1}{h_n} \int \phi(t)k\left(\frac{x-t}{h_n}\right)dt, \end{aligned}$$

en utilisant le changement de variable $u = \frac{x-t}{h_n}$, on obtient :

$$E(g(x)) = \int \phi(x - uh_n)k(u)du,$$

comme k est à support compact, la continuité uniforme de ϕ et l'hypothèse (k. 1), nous donnent :

$$\lim_{n \rightarrow \infty} E(\hat{\phi}_{h_n}(x)) = \phi(x).$$

■

Lemme 2.22.

D'après les hypothèses 6, (2. 16), (k. 1) et (k. 7) on a :

$$\lim_{n \rightarrow \infty} E(\hat{\phi}_{h_n}(x)) - \hat{\phi}_n(x) = 0 \text{ p. co.}$$

Démonstration.

on a

$$\hat{\phi}_{h_n}(x) - E(\hat{\phi}_{h_n}(x)) = \frac{1}{n} \sum_{i=1}^n Z_i,$$

où

$$Z_i = \frac{1}{h_n} \left[Y_i k\left(\frac{x - X_i}{h_n}\right) - E\left(Y_i k\left(\frac{x - X_i}{h_n}\right)\right) \right].$$

De plus, les hypothèses (2. 16) et (K. 7) nous donnent :

$$\begin{aligned} \left| k\left(\frac{x - X_i}{h_n}\right) \right| \leq M &\Rightarrow Y_i \left| k\left(\frac{x - X_i}{h_n}\right) \right| \leq M^2 \\ &\Rightarrow Y_i \left| k\left(\frac{x - X_i}{h_n}\right) \right| - E\left(Y_i \left| k\left(\frac{x - X_i}{h_n}\right) \right|\right) \leq M^2 - E\left(Y_i \left| k\left(\frac{x - X_i}{h_n}\right) \right|\right) \\ &\Rightarrow \frac{Y_i \left| k\left(\frac{x - X_i}{h_n}\right) \right| - E\left(Y_i \left| k\left(\frac{x - X_i}{h_n}\right) \right|\right)}{h_n} \leq \frac{M^2 - E\left(Y_i \left| k\left(\frac{x - X_i}{h_n}\right) \right|\right)}{h_n} \\ &\Rightarrow |Z_i| \leq \frac{M^2 - E\left(Y_i \left| k\left(\frac{x - X_i}{h_n}\right) \right|\right)}{h_n} \end{aligned}$$

alors :

$$|Z_i| \leq \frac{C}{h_n}$$

où

$$C = M^2 - E \left(Y_i \left| k \left(\frac{x - X_i}{h_n} \right) \right. \right),$$

d'autre part

$$E(Z_i^2) = \text{var} \left(\frac{1}{h_n} Y_i k \left(\frac{x - X_i}{h_n} \right) \right) \leq E(T_i),$$

où :

$$T_i = \frac{1}{h_n} Y_i k \left(\frac{x - X_i}{h_n} \right).$$

En utilisant le conditionnement par rapport à la variable X , on obtient :

$$\begin{aligned} E[T_i^2] &= \frac{1}{h_n^2} E \left[\phi(x) k^2 \left(\frac{x - X_i}{h_n} \right) \right] \\ &= \frac{1}{h_n^2} \int \phi(x) k^2 \left(\frac{x - t}{h_n} \right) f(t) dt, \end{aligned}$$

où :

$$\phi(x) = E(Y^2 / X = x),$$

en utilisant le changement de variable $z = \frac{x-t}{h_n}$ on obtient :

$$E(T_i^2) = \frac{1}{h_n} \int \phi(x - zh_n) k^2(z) f(x - zh_n) dz,$$

La continuité de f sur le support compact k , les hypothèses (2. 16) et (K. 1) impliquent :

$$E(T_i^2) \leq \frac{C}{h_n}.$$

comme les conditions du corollaire (2. 13) étant satisfaites, alors nous déduisons que :

$$\frac{1}{n} \sum_i Z_i = O_{p.co} \left(\sqrt{\frac{\log n}{nh_n}} \right).$$

La convergence presque complète de la densité établie dans le paragraphe précédent assure les convergences suivantes :

$$\lim_{n \rightarrow \infty} E(\hat{f}_{h_n}(x)) = f(x)$$

$$\lim_{n \rightarrow \infty} E(\hat{f}_{h_n}(x)) - f(x) = 0 \text{ p.co.}$$

■

Lemme 2.23.

D'après les hypothèses 2, 6, (K. 1) et (K. 7) on a

$$\exists \delta > 0, \sum_{n \in \mathbb{N}} p(\hat{f}_{h_n}(x) \leq \delta) < \infty.$$

Le théorème (2. 14), entraîne la convergence presque complète de $\hat{f}_{h_n}(x)$ vers $f(x)$, c'est à dire :

$$\forall \varepsilon > 0, \sum_{n \in \mathcal{N}} p(|\hat{f}_{h_n}(x) - f(x)| > \varepsilon) < \infty,$$

on a

$$|\hat{f}_{h_n}(x)| \leq \frac{f(x)}{2} \Rightarrow |\hat{f}_{h_n}(x) - f(x)| > \frac{f(x)}{2}$$

d'où

$$P \left[|\hat{f}_{h_n}(x) \leq \frac{f(x)}{2} \right] \leq P \left[|\hat{f}_{h_n}(x) - f(x)| > \frac{f(x)}{2} \right]$$

Comme $f(x) > 0$, en posant $\delta = \varepsilon = \frac{f(x)}{2}$, on arrive au résultat.

7. r et f est k fois continûment dérivable autour du point x .

Théorème 2.24.

Considérons le modèle 7 avec $k > 0$, et supposons que les hypothèses 2, 3, (2. 16), (2. 15), (K. 7) soient réalisées, alors on a :

$$|\hat{r}_{h_n}(x) - r(x)| = O(h_n^k) + O\left(\sqrt{\frac{\log(n)}{n}}\right) \text{ p. co.}$$

En utilisant la décomposition précédente, Le résultat énoncé découle des lemmes suivants :

Lemme 2.25.

Sous les hypothèses 3, 7 et (K. 7) on a :

$$E(\hat{\phi}_{h_n}(x)) - \phi(x) = O(h_n^k). \quad (2.17)$$

Démonstration.

L'expression de $\hat{\phi}_{h_n}(x)$ est analogue à la précédente. En effet, on a :

$$E(\hat{\phi}_{h_n}(x)) = \int_{-\infty}^{+\infty} k(z)\phi(x - zh_n)dz.$$

Le modèle 7, nous permet de développer ϕ au voisinage de x , ceci nous permet d'écrire :

$$\phi(x - zh_n) = \phi(x) + \sum_{i=1}^{k-1} \frac{(-1)^i (zh_n)^i}{i!} \phi^{(i)}(x) + \frac{(-1)^k (zh_n)^k}{k!} \phi^{(k)}(\theta_z),$$

où θ_z entre x et $x - zh_n$.

L'hypothèse 3 sur k , implique :

$$E(\hat{\phi}_{h_n}(x)) = \phi(x) + (-1)^k h^k \frac{\int z^k \phi^{(k)}(z)(\theta_z) dz}{k!}.$$

La convergence uniforme de $\phi^k(\theta_z)$ vers $\phi^k(x)$ (assurée par le modèle 7) et la condition (K. 7), nous donnent

$$E(\hat{\phi}_{h_n}(x)) - \phi(x) = O(h_n^k).$$

■

Les lemmes (2. 16), (2. 22), (2. 23) et (2. 25) nous assurent que :

$$E(\hat{f}_{h_n}(x) - f(x)) = O_{p.co} \left(\sqrt{\frac{\log n}{nh_n}} \right).$$

$$E(\hat{g}_{h_n}(x) - g(x)) = O_{p.co} \left(\sqrt{\frac{\log n}{nh_n}} \right).$$

Et :

$$\exists \delta > 0, \sum_{n \in \mathbb{N}} p(\hat{f}_{h_n}(x) \leq \delta) < \infty.$$

En combinant tous les résultats cités précédemment, on arrive au résultat cherché. Nous essayons maintenant d'établir une vitesse convergence presque complète uniforme. Il suffit d'une part de considérer un compact S de R tel que l'hypothèse 7 soit remplacée par le modèle suivant.

8. r et f sont k continûment dérivables sur S .

Et de supposer d'autre part l'existence de $\theta > 0$, tel que :

$$\inf_{x \in S} f(x) > \theta \tag{2.18}$$

Nous gardons toutes les conditions citées précédemment, auxquelles nous rajoutons la condition Lipschitzienne (2. 9) sur le noyau k et l'hypothèse (2. 10) sur le paramètre de lissage h .

Théorème 2.26.

Soient les modèles 8, (2. 19) avec $k > 0$ et les hypothèses (K. 7), (2. 16), 2, 3, (2. 9) et (2. 10), on a

$$\sup_{x \in S} (|\hat{r}_{h_n}(x) - r(x)|) = O(h_n^k) + O \left(\sqrt{\frac{\log n}{n}} \right), p. co.$$

Une décomposition similaire au cas ponctuel, nous permet d'écrire :

$$\begin{aligned} \sup_{x \in S} (|\hat{r}_{h_n}(x) - r(x)|) &\leq \frac{\sup_{x \in S} (|\hat{\phi}_{h_n}(x) - \phi(x)|)}{\inf_{x \in S} |\hat{f}_{h_n}(x)|} + \sup_{x \in S} (|f(x) - \hat{f}_{h_n}(x)|) \frac{\sup_{x \in S} (|\hat{r}_{h_n}(x)|)}{\inf_{x \in S} |\hat{f}_{h_n}(x)|} \\ &\leq \frac{\sup_{x \in S} |\hat{\phi}_{h_n}(x) - E(\hat{\phi}_{h_n}(x))|}{\inf_{x \in S} |\hat{f}_{h_n}(x)|} + \frac{\sup_{x \in S} |E(\hat{\phi}_{h_n}(x)) - \hat{\phi}_{h_n}(x)|}{\inf_{x \in S} |\hat{f}_{h_n}(x)|} \\ &\quad + \left\{ \sup_{x \in S} |f(x) - E(\hat{f}_{h_n}(x))| + \sup_{x \in S} (|E(\hat{f}_{h_n}(x)) - \hat{f}_{h_n}(x)|) \right\} \frac{\sup_{x \in S} |\hat{r}_{h_n}(x)|}{\inf_{x \in S} |\hat{f}_{h_n}(x)|} \end{aligned}$$

Les approximations en $O(h_n^k)$ traitées précédemment peuvent se généraliser via (K. 7) et 7 comme suit :

$$\sup_{x \in S} E(\hat{f}_{h_n}(x) - f(x)) = O(h_n^k)$$

et

$$\sup_{x \in S} E(\hat{g}_{h_n}(x) - g(x)) = O(h_n^k).$$

Comme r est borné, la preuve de ce théorème s'achèvera à partir des lemmes suivants :

Lemme 2.27.

Sous les hypothèses 2, 8, (2. 9) et (2. 10) on a

$$\sup_{x \in S} |E[\hat{\phi}_{h_n}(x)] - \hat{\phi}_{h_n}(x)| = O\left(\sqrt{\frac{\log(n)}{n}}\right), \text{ p.co.}$$

Démonstration.

S est un compact de R , il existe un recouvrement fini de S tel que :

$$S \subset \cup_{k=1}^{z_n} S_k$$

où

$$S_k =]t_k - l_n, t_k + l_n[\text{ et } l_n = n^{-\beta}$$

Posons

$$t_x = \operatorname{argmin}_{t \in t_1, t_2, \dots, t_{z_n}} |x - t|$$

avec

$$l_n = n^{-2\varepsilon}, l_n = Cz_n^{-1}$$

On a

$$\sup_{x \in S} |E[\hat{\phi}_{h_n}(x)] - \hat{\phi}_{h_n}(x)| \leq A_1 + A_2 + A_3.$$

où

$$\begin{aligned} A_1 &= \text{Sup}_{y \in S} |[\hat{\phi}_{h_n}(x)] - \hat{\phi}_{h_n}(t_x)|, \\ A_2 &= \text{Sup}_{y \in S} |[\hat{\phi}_{h_n}(t_x)] - E[\hat{\phi}_{h_n}(t_x)]|, \\ A_3 &= \text{Sup}_{y \in S} |[E[\hat{\phi}_{h_n}(t_x)] - \hat{\phi}_{h_n}(x)]|. \end{aligned}$$

Concernant le terme A_1 , comme le noyau k est lipschitzien et la variable Y est bornée, on a

$$\begin{aligned} |\hat{\phi}_{h_n}(t_x) - \hat{\phi}_{h_n}(x)| &= \frac{1}{nh_n} \sum_{i=1}^n |Y_i| \left| \left[k\left(\frac{t_x - X_i}{h_n}\right) - k\left(\frac{x - X_i}{h_n}\right) \right] \right| \\ &\leq \frac{C}{h_n} \frac{|t_x - x|}{h_n^2} \\ &= \frac{Cl_n}{h_n^2}. \end{aligned}$$

L'hypothèse (2. 10) implique

$$A_1 = o\left(\frac{\log n}{nh_n}\right).$$

Une manière de démonstration analogue à la précédente, nous permet d'écrire

$$A_3 = o\left(\frac{\log n}{nh_n}\right).$$

On ce qui concerne le terme A_2 , on a : $\forall \varepsilon > 0$.

$$\begin{aligned} P \left[\text{Sup}_{y \in S} |[\hat{\phi}_{h_n}(t_x)] - E[\hat{\phi}_{h_n}(t_x)]| > \varepsilon \right] &= P \left[\max_{j=1, \dots, z_n} |\hat{\phi}_{h_n}(t_j) - E[\hat{\phi}_{h_n}(t_j)]| > \varepsilon \right] \\ &\leq z_n P[|\hat{\phi}_{h_n}(t_j) - E[\hat{\phi}_{h_n}(t_j)]| > \varepsilon] \\ &\leq z_n P \left[\frac{1}{nh_n} \sum_{i=1}^n |u_i - E[u_i]| > \varepsilon \right]. \end{aligned}$$

Où

$$u_i = Y_i k\left(\frac{X_i - t_k}{h_n}\right).$$

Il suffit de trouver des majorants pour u_i et $E[u_i^2]$, pour pouvoir appliquer le corollaire (2. 13). D'après la démonstration du lemme (2. 16), on a

$$u_i \leq \frac{C}{h_n} \text{ et } E[u_i^2] \leq \frac{C}{h_n}.$$

Maintenant nous sommes en mesure d'appliquer le corollaire (2. 13) :

$$P \left[\text{Sup}_{y \in S} |\hat{\phi}_{h_n} - E(\hat{\phi}_{h_n})(x)| > \varepsilon \right] \leq n^{2\varepsilon} \exp(-Cn\varepsilon^2 h_n).$$

comme :

$$\lim_{n \rightarrow \infty} \sqrt{\frac{\log(n)}{nh_n}} = 0.$$

En posant $\varepsilon = \varepsilon_0 \sqrt{\frac{\log(n)}{nh_n}}$, on obtient alors :

$$\forall \varepsilon > 0, \sum P \left[\text{Sup}_{y \in S} |\hat{\phi}_{h_n} - E(\hat{\phi}_{h_n}(x))| > \varepsilon \right] \leq \infty$$

pour ε_0 choisi suffisamment grand



Conclusion

Le problème posé dans ce mémoire est la convergence presque complète des estimateurs non paramétrique à noyaux de la fonction de densité et de régression. Deux méthodes pour estimer la fonction de répartition ont été utilisées : la fonction de répartition empirique, et l'estimateur à noyau, quatre méthodes pour estimer la fonction de densité : histogramme, estimateur simple, méthode du noyau et système trigonométrique. Concernant l'estimation de la fonction de régression, trois méthodes ont été proposées : méthode à noyau, méthode des k-plus proches voisins et la méthode des Polynômes locaux. Notez que les estimateurs calculés par les méthodes citées précédemment présentent de bonnes propriétés du biais, variance et erreur moyenne quadratique.

Bibliographie

- [1] **A. Plansky**, *Introduction to statistical limit theory*, 2011.
- [2] **E. Cinlar**, *Probability and stochastics*, Springer, 2010.
- [3] **F. Ferraty, F. vieu,** *Non Paraetric Functional Data Analysis*, Springer, 2006.
- [4] **N. Saadi, S. Adjabi**, *On the estimation of the probability density by trigonometric series*, Communications in Statistics-Theory and Methods, 38(3583-3595), 2009.
- [5] **M.Reiss**, *Nonparametric estimation of the smoth distribution function*, Scandinavian Journal of Statistics, P .116-119, 1981.
- [6] **M.Rosenblatt**, *Remarks on some nonparametric estimates of a density function*, Annals of Mathematical Statistics, 27 :8332-837, 1956.
- [7] **D.Bosq.**, *Estimation adaptatif de la fonction de densité par projection tronquée*, C. R. Acad. Sci. Paris, Ser. I, 334 : 591-595, 2002.
- [8] **W. Härdle**, *Applied Nonparametric Regression*, Humboldt-Universität zu Berlin, 1994.
- [9] **M. Lejeune**, *Statistique, la théorie et ses applications*, Deuxième édition, Springer 2010.
- [10] **Z. Lai, Z. Bai**, *Probability Inequaliteies*, Springer 2010.
- [11] **L. Wasserman**, *All of Non parametric statistics*, Springer 2006.
- [12] **P. Hsu, H. Robbins**, *Complete convergence and the law of large numbers*, proc. Math. Acad. Sci. USA33 : 25-31, 1947.
- [13] **E. Parzen**, *On estimation of a probability density function and mode. Ann. Math. Statist*, 33, 1065-1076, (1962).
- [14] **A. Nadaraya**, *Nonparametric Estimation of Probability Densities and Regression Curves*, Kluwer, Dordrecht, (1989).

-
- [15] **J. Tukey**, *Curves as parameters, and touch estimation*. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob*, Vol. I, page 681-694. Univ. California Press, Berkeley, Curves.(1962).
- [16] **G. Watson** *Smooth regression analysis*. *Sankhyà Ser. A*, 26, 359-372, (1964).
- [17] **H. Sabry** *Sur l'estimation non paramétrique des fonctions de régression*. *C. R. Acad. Sci. Paris. Sér. A-B*, 286(20), A941-A944, (1978).
- [18] **J. Lecoutre**, *Contribution à l'estimation non paramétrique de la régression*. *PhD thesis, Université de Pierre et Marie Curie-ParisVI-France*, (1982).
- [19] **L. Lai**, *Large sample properties of k-nearest neighbor procedures*, Ph. d. dissertation, Dept. Mathematics, UCLA, Los Angeles, 1977.
- [20] **P. Mack**, *Local properties of k-nn regression estimates*, *SIAM, J. Alg. Disc. Meth.* 2, 311-23, 1981.
- [21] **C. Loader**, *bandwidth selection : Classical or plug-in*. *Ann. of Statist*, 27, 1999.
- [22] **D. O. Loftsgaarden and G. P. Quesenberry**, *A nonparametric estimate of a multivariate density function*. *Annals of Mathematical Statistics*, 36, 1049-51, 2005.
- [23] **T. M. Cover and P. E. Hart**, *Nearest neighbor pattern classification*. *IEEE Transactions on Information Theory*, 13, 21-7, 1967.
- [24] **J. S. Simonoff.**, *Smoothing Methods in Statistics*, Springer-Verlag, 1996.

Résumé

Dans ce travail, on est intéressé à l'estimation non paramétrique des fonctions de densités, de répartition et de régression par certaines méthodes dont les estimateurs sont consistants, et on a étudié la convergence presque complète des estimateurs à noyaux de chacune des fonctions citées précédemment.

Mots-clés : Estimation non paramétrique, estimation de la densité de probabilité, estimation de la fonction de répartition, estimation de la fonction de régression, estimation à noyau, convergence complète, propriétés asymptotique.

Abstract

In this work, we are interested in the non-parametric estimation of density, distribution and regression functions by some methods whose estimators are consistent, and we have studied the almost complete convergence of the kernel estimators of each of the functions mentioned above.

Keywords : Nonparametric estimation, probability density estimation, distribution function estimation, regression function estimation, kernel estimation, complete convergence, asymptotic properties.