

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

UNIVERSITÉ FRÈRES MENTOURI CONSTANTINE
FACULTÉ DES SCIENCES EXACTES
DÉPARTEMENT DE MATHÉMATIQUES

POLYCOPIÉ DE COURS
MASTER 1 SPA

Estimation non paramétrique

Par Sara LEULMI et Fatiha MESSACI

Année 2017/2018

Table des matières

1	Estimation des fonctions de répartition et de densité	3
1.1	Modes de convergence	4
1.1.1	Convergence en loi	4
1.1.2	Convergence en probabilité	4
1.1.3	Convergence en moyenne quadratique	4
1.1.4	Convergence presque sûre	4
1.1.5	Convergence presque complète	5
1.2	Estimation de la fonction de répartition	5
1.3	Estimation de la fonction de densité	6
2	Estimation de la fonction de régression	12
2.0.1	Estimateur de Nadaraya-Watson	12
2.0.2	Consistance de l'estimateur de Nadaraya-Watson	15
3	Prévision non paramétrique dans un cadre fonctionnel	18
3.1	Méthodes d'estimation	19
3.1.1	Estimation de la régression	19
3.1.2	Estimation du quantile conditionnel	19
3.1.3	Estimation du mode conditionnel	20
3.2	Outils préliminaires	21
3.2.1	Noyaux usuels	21
3.2.2	Inégalités exponentielles de type Bernstein	24
3.3	Résultats asymptotiques	25
3.3.1	Convergence presque complète de l'estimateur de la régression	25
3.3.2	Convergence presque complète de l'estimateur de la médiane conditionnelle	28
3.3.3	Convergence presque complète du mode conditionnel . .	32

3.4	Taux de convergence	39
3.4.1	Taux de convergence pour l'estimateur de la régression	39
3.4.2	Taux de convergence pour l'estimateur de la médiane conditionnelle	39
4	Etude sur des données réelles	41

Chapitre 1

Estimation des fonctions de répartition et de densité

La théorie de l'estimation est une préoccupation majeure de la statistique. Cette théorie est divisée en deux volets principaux, à savoir l'estimation paramétrique et l'estimation non-paramétrique. Cette dernière consiste à estimer à partir des observations, une fonction inconnue, élément d'une certaine classe fonctionnelle. Une procédure non-paramétrique est définie indépendamment de la loi de l'échantillon d'observation. En fait, on parle de méthode d'estimation non-paramétrique lorsque celle-ci ne se ramène pas à l'estimation d'un paramètre appartenant à un espace de dimension finie. Plus généralement un modèle de statistique semi paramétrique comporte à la fois une composante paramétrique et une autre non paramétrique (exemple : le modèle de Cox). Un des problèmes centraux en statistique est celui de l'estimation de caractéristiques fonctionnelles associées à la loi des observations, comme par exemple, la fonction de répartition, la densité ou la fonction de régression.

Le but de ce chapitre est d'introduire des estimateurs non paramétriques des fonctions de répartition et de densité, basés sur des données indépendantes et identiquement distribués, et donner quelques unes de leurs propriétés. Ces dernières s'énoncent par un ou plusieurs des modes de convergence, que nous rappelons ci dessous.

Soit (Ω, \mathcal{A}, P) un espace de probabilité, (X_n) une suite de variables aléatoires réelles définies sur Ω et X une variable aléatoires réelle définie sur Ω .

1.1 Modes de convergence

1.1.1 Convergence en loi

(X_n) converge en loi vers X si la suite des fonctions de répartition des variables X_n converge vers la fonction de répartition de X , en tout point de continuité de cette dernière. Ceci est équivalent, grâce au théorème de Paul Lévy, à la convergence des suites des fonctions caractéristiques des variables X_n vers la fonction caractéristique de X , en tout point.

On note $X_n \xrightarrow{\mathcal{L}} X$.

1.1.2 Convergence en probabilité

(X_n) converge en probabilité vers X si

$$\forall \delta > 0, \lim_{n \rightarrow \infty} P(|X_n - X| > \delta) = 0.$$

On note $X_n \xrightarrow{\mathcal{P}} X$. Elle correspond à la convergence en mesure.

1.1.3 Convergence en moyenne quadratique

Si $X_n \in L^2$ et $X \in L^2$, on dit que X_n converge en moyenne quadratique vers X si

$$E |X_n - X|^2 = \int |X_n - X|^2 dP \xrightarrow[n \rightarrow \infty]{} 0.$$

On note $X_n \xrightarrow{m.q.} X$. Elle correspond à la convergence dans L^2 .

1.1.4 Convergence presque sûre

(X_n) converge vers X presque sûrement s'il existe un ensemble P négligeable N tel que

$$\forall \omega \in N^c, |X_n(\omega) - X(\omega)| \xrightarrow[n \rightarrow \infty]{} 0.$$

On note $X_n \xrightarrow{p.s.} X$. Elle correspond à la convergence P presque partout.

1.1.5 Convergence presque complète

La suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ converge presque complètement vers la v.a.r. X si

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

On note $X_n \xrightarrow{p.co.} X$.

Il est connu que ces différents modes de convergence sont liés par les relations suivantes.

$$X_n \xrightarrow{p.co.} X \Rightarrow X_n \xrightarrow{p.s.} X \Rightarrow X_n \xrightarrow{\mathcal{P}} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X \text{ et } X_n \xrightarrow{m.q.} X \Rightarrow X_n \xrightarrow{\mathcal{P}} X.$$

Soulignons le fait que toutes les autres implications ne sont pas vraies et qu'en particulier la convergence presque complète n'implique pas la convergence en moyenne quadratique qui, à son tour, n'entraîne pas la convergence presque complète (voir exercice 1.3).

1.2 Estimation de la fonction de répartition

La fonction de répartition caractérise complètement la loi d'une variable aléatoire X et permet, par exemple, de calculer la probabilité que X appartienne à un intervalle. Son estimation est fondamentale en statistique. Soient donc X_1, \dots, X_n des variables aléatoires réelles indépendantes et identiquement distribuées, de même loi que X , de fonction de répartition inconnue F . Une estimation naturelle de F se fait par la fonction de répartition empirique donnée par :

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

où $1_{\{\cdot\}}$ désigne la fonction indicatrice.

F_n est elle-même une fonction de répartition d'une loi discrète.

D'après la loi forte des grands nombres, on a pour tout $x \in \mathbb{R}$

$$F_n(x) \xrightarrow[n \rightarrow \infty]{p.s.} F(x). \quad (1.1)$$

Donc, F_n est un estimateur fortement consistant de F . De plus, la convergence presque sûre est uniforme en $x \in \mathbb{R}$, d'après le résultat suivant.

Théorème 1 (*Glivenko-Cantelli*)

$$P \left[\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0 \right] = 1.$$

Preuve. En tenant compte de la relation (1.1), elle découle immédiatement de l'exercice 1.2. ■

Le taux de convergence est précisé par la loi du logarithme itéré (LIL) qui est donné dans ? comme suit

$$P \left(\limsup_{n \rightarrow \infty} \frac{\sqrt{n} \sup_x |F_n(x) - F(x)|}{\sqrt{\frac{1}{2} \log \log n}} \leq 1 \right) = 1.$$

Il est facile de voir que, pour tout x , $F_n(x)$ converge en moyenne quadratique vers $F(x)$ et que $F_n(x)$ est asymptotiquement normal par application du théorème central limite.

1.3 Estimation de la fonction de densité

Supposons que la loi de X est absolument continue, par rapport à la mesure de Lebesgue, de densité de probabilité f_X , que nous cherchons à estimer. Ce problème tire son intérêt du fait qu'on peut mieux visualiser la loi (que par utilisation de la fonction de répartition) en déduisant de l'estimateur des éléments concernant la symétrie, la multimodalité ou l'aplatissement de la loi étudiée. De plus cela constitue une aide précieuse au choix d'un modèle de probabilité et permet d'introduire un estimateur non paramétrique de la régression, sur lequel nous revenons au chapitre suivant.

Une solution intuitive a été proposée par Rosenblatt (1956) comme suit. Pour f_X continue et $h > 0$ assez petit, nous avons

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

En remplaçant F par l'estimateur de la fonction de répartition F_n , nous obtenons l'estimateur suivant de f_X :

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h},$$

qui peut aussi s'écrire sous la forme :

$$f_n(x) = \frac{1}{2nh} \sum_{i=1}^n 1_{\{x-h \leq X_i \leq x+h\}} = \frac{1}{nh} \sum_{i=1}^n K_0\left(\frac{X_i - x}{h}\right),$$

où $K_0(u) = \frac{1}{2} 1_{\{|u| \leq 1\}}$.

Rosenblatt (1956) puis Parzen (1962) ont suggéré une généralisation de cet estimateur, en posant :

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

où $K : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction intégrable, telle que $\int K(u) du = 1$, c'est **l'estimateur à noyau** de la densité ou **l'estimateur de Parzen-Rosenblatt**.

La fonction K est le noyau (kernel en anglais) et le paramètre h est la fenêtre (bandwidth en anglais).

Voici quelques exemples de noyaux usuels.

- ♠ $K(u) = \frac{1}{2} 1_{\{|u| \leq 1\}}$: noyau rectangulaire,
- ♠ $K(u) = (1 - |u|) 1_{\{|u| \leq 1\}}$: noyau triangulaire,
- ♠ $K(u) = \frac{3}{4} (1 - u^2) 1_{\{|u| \leq 1\}}$: noyau d'Epanechnikov ou parabolique,
- ♠ $K(u) = \frac{15}{16} (1 - u^2)^2 1_{\{|u| \leq 1\}}$: noyau quadratique,
- ♠ $K(u) = \frac{1}{\sqrt{2\Pi}} \exp(-u^2/2)$: noyau gaussien,
- ♠ $K(u) = \frac{1}{\sqrt{2\Pi}} \exp(-|u|/\sqrt{2}) \sin(|u|/\sqrt{2} + \Pi/4)$: noyau de Silverman.

Quant à la fenêtre, nous la faisons dépendre de n et nous la notons h_n , la suite $(h_n)_{n \geq 1}$ tendant vers zéro lorsque n tend vers l'infini.

Intéressons nous maintenant à la convergence en moyenne quadratique de cet estimateur. Pour cela, nous avons besoin du résultat suivant.

Théorème 2 (Bochner) Soit $K : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ($\mathcal{B}(\mathbb{R})$ étant la tribu borélienne de \mathbb{R}) une application bornée et intégrable telle que :

$$|z|K(z) \xrightarrow{|z| \rightarrow \infty} 0.$$

Par ailleurs, soit $g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une application intégrable.

Posons $g_n(x) = \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{z}{h_n}\right) g(x-z) dz$, où $0 < h_n \xrightarrow{n \rightarrow \infty} 0$.

Si g est continue au point x alors $\lim_{n \rightarrow \infty} g_n(x) = g(x) \int_{-\infty}^{+\infty} K(z) dz$. Si g est uniformément continue alors la convergence de g_n est uniforme.

Preuve. Preuve : Soit $\varepsilon > 0$, on a : $|g_n(x) - g(x) \int_{-\infty}^{+\infty} K(z) dz| =$
 $|\frac{1}{h_n} \int_{-\infty}^{+\infty} K(\frac{z}{h_n}) g(x - z) dz - \int_{-\infty}^{+\infty} g(x) K(z) dz| = \left| \int_{-\infty}^{+\infty} K(y) g(x - h_n y) dy - \int_{-\infty}^{+\infty} g(x) K(y) dy \right|$.
 g étant continue au point x , il vient

$$\forall \varepsilon > 0, \exists \delta > 0 / |x - z| < \delta \Rightarrow |g(x) - g(z)| \leq \varepsilon.$$

Donc :

$$\begin{aligned} & \int |(g(x - h_n y) - g(x)) K(y)| dy \\ & \leq \int_{\{y: |y| < \frac{\delta}{h_n}\}} |g(x - h_n y) - g(x)| |K(y)| dy + \int_{\{y: |y| \geq \frac{\delta}{h_n}\}} |g(x - h_n y) - g(x)| |K(y)| dy \\ & \leq \varepsilon \int_{\{y: |y| < \frac{\delta}{h_n}\}} |K(y)| dy + \int_{\{y: |y| \geq \frac{\delta}{h_n}\}} |g(x - h_n y)| |K(y)| dy + |g(x)| \int_{\{y: |y| \geq \frac{\delta}{h_n}\}} |K(y)| dy \end{aligned}$$

Par l'absolue continuité de l'intégrale et le fait que $h_n \xrightarrow[n \rightarrow \infty]{} 0$:

$$\exists n_0 \in \mathbb{N} / \forall n \geq n_0 : \int_{\{y: |y| \geq \frac{\delta}{h_n}\}} |K(y)| dy < \varepsilon.$$

De plus, sur $\left\{ |y| \geq \frac{\delta}{h_n} \right\}$, on a $|y| \frac{h_n}{\delta} \geq 1$, d'où

$$\begin{aligned} & \int_{\{y: |y| \geq \frac{\delta}{h_n}\}} |g(x - h_n y)| |K(y)| dy \leq \int_{\{y: |y| \geq \frac{\delta}{h_n}\}} |y| \frac{h_n}{\delta} |g(x - h_n y)| |K(y)| dy \\ & \leq \frac{1}{\delta} \sup_{|y| \geq \frac{\delta}{h_n}} |y| |K(y)| \int |g(z)| dz \rightarrow 0 \text{ par hypothèse puisque } h_n \rightarrow 0. \end{aligned}$$

Nous pouvons montrer que si g est uniformément continue (et du fait qu'elle soit intégrable) alors $\lim_{|x| \rightarrow \infty} g(x) = 0$, g est donc bornée et nous obtenons la seconde assertion du théorème. ■

Nous imposons au noyau K d'être une densité de probabilité (ce qui fait de f_n une densité) et de satisfaire les conditions suivantes.

K_1 : K est bornée et paire ;

K_2 : $|x| K(x) \xrightarrow[|x| \rightarrow \infty]{} 0$.

Remarquons que les noyaux gaussien, parabolique et rectangulaire (par exemple) vérifient ces conditions.

Nous sommes, maintenant, en mesure d'énoncer et de montrer le résultat visé

Théorème 3 Si f_X est continue au point x , $h_n \xrightarrow[n \rightarrow \infty]{} 0$, $nh_n \rightarrow \infty$ et si la densité K vérifie les conditions K_1 et K_2 , alors $f_n(x)$ est un estimateur convergeant en moyenne quadratique vers $f_X(x)$.

Preuve. D'abord par equidistribution des X_i , l'application du théorème de transfert et l'invariance par translation de la mesure de Lebesgue, il vient

$$\begin{aligned} Ef_n(x) &= E \left\{ \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{x-X_i}{h_n} \right) \right\} \\ &= \frac{1}{h_n} EK \left(\frac{x-X_1}{h_n} \right) = \frac{1}{h_n} \int K \left(\frac{x-y}{h_n} \right) f_X(y) dy \\ &= \frac{1}{h_n} \int K \left(\frac{z}{h_n} \right) f_X(x-z) dz \rightarrow f_X(x) \int K(z) dz = f_X(x), \end{aligned}$$

en vertu du théorème de Bochner (toutes les conditions sont remplies) et du fait que K est une densité. Ceci montre que $f_n(x)$ est un estimateur asymptotiquement sans biais de $f_X(x)$.

Ensuite, par indépendance et équidistribution des X_i , on peut écrire

$$\begin{aligned} Var f_n(x) &= \frac{1}{nh_n^2} var \left\{ K \left(\frac{x-X_1}{h_n} \right) \right\} \leq \frac{1}{nh_n^2} E \left\{ K^2 \left(\frac{x-X_1}{h_n} \right) \right\} \\ &= \frac{1}{nh_n} \left\{ \frac{1}{h_n} \int K^2 \left(\frac{z}{h_n} \right) f_X(x-z) dz \right\} \\ &\xrightarrow[n \rightarrow \infty]{} \lim \frac{1}{nh_n} \left\{ f_X(x) \int K^2(z) dz \right\} = 0, \end{aligned}$$

par application, encore une fois, du théorème de Bochner et du fait que $nh_n \rightarrow \infty$.

■

Remarque 1 1) On peut montrer que si K est à variation bornée ($K = K_1 - K_2$ où $K_1 \nearrow$ et $K_2 \nearrow$) et si pour tout $\theta > 0$, $\sum_{i=1}^{\infty} \exp(-n\theta h_n^2) < \infty$ alors $\sup |f_n(x) - f(x)| \xrightarrow[n \rightarrow \infty]{p.s.} 0$, ssi f est uniformément continue (se référer à Prakasa (1983) pour la preuve).

2) Cette méthode d'estimation (à noyau) se généralise au cas de \mathbb{R}^p . Ainsi, si X_1, \dots, X_n , sont n vecteurs aléatoires i.i.d. de \mathbb{R}^p , de même densité f_X inconnue, alors on peut l'estimer par $f_n(x) = \frac{1}{nh_n^p} \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right)$, $x \in \mathbb{R}^p$. Le théorème de Bochner reste valable si K et g sont définies sur \mathbb{R}^p et si on

pose $g_n(x) = \frac{1}{h_n} \int K\left(\frac{z}{h_n}\right) g(x-z) dz$, et permet de déduire la consistance de $f_n(x)$, sous des conditions adéquates sur f_X , h_n et K .

3) Il existe bien d'autres méthodes d'estimation de la densité, à savoir : la méthode de projection et celle des delta suites, une des plus récentes et des plus générales (elle englobe les précédentes) étant la méthode des ondelettes (cf Doukhan et Léon (1990), Kerkycharian et Picard (1992) et Walter (1992), mais l'estimation par la méthode des noyaux reste très utilisée.

4) Il est connu que le choix du noyau n'influe pas beaucoup sur l'estimateur à noyau, mais ce dernier est très sensible au choix de la fenêtre. Il existe donc des méthodes qui aident à un choix plus ou moins "optimum", parmi lesquelles la méthode du plug-in ou la méthode du pouce ; une petite initiation à ce problème est donnée à l'exercice 1.4.

EXERCICE 1.1

Soit F la fonction de répartition d'une v.a.r. X et soit F^{inv} son inverse généralisée, définie pour $t \in]0, 1[$ par

$$F^{inv}(t) = \inf\{x \in \mathbb{R} / F(x) \geq t\}.$$

- 1) Montrer qu'on a les propriétés suivantes.
 - a) F^{inv} est une fonction croissante.
 - b) $\forall t \in]0, 1[, F(F^{inv}(t)) \geq t$.
 - c) $\forall x \in \mathbb{R}, F^{inv}(F(x)) \leq x$.
 - d) $F(x) \geq t \Leftrightarrow x \geq F^{inv}(t)$.
 - e) Si F est bijective alors $F^{inv} = F^{-1}$.
- 2) Donner la loi de $F^{inv}(U)$ où U suit la loi uniforme sur $[0, 1]$.
- 3) Si F est continue, quelle est la loi de $F(X)$?

EXERCICE 1.2

Soit F une fonction de répartition et soit $(F_n(x))$ une suite de fonctions croissantes convergeant vers $F(x)$ (en tout point x). Montrer que cette convergence est en fait uniforme, c'est à dire que

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{n \rightarrow \infty} 0.$$

EXERCICE 1.3

- 1) Montrer que la convergence presque complète implique la convergence presque sûre et la convergence en probabilité.
- 2) Montrer que la convergence presque complète n'implique pas la convergence en moyenne quadratique qui, à son tour, n'entraîne pas la convergence presque complète. Indication : Soit l'espace de probabilité $([0, 1], \mathcal{B}([0, 1]), \lambda)$, où $\mathcal{B}([0, 1])$ désigne la tribu borélienne sur $[0, 1]$ et λ est la restriction de la mesure de Lebesgue à cet ensemble, considérer les suites $f_n = n1_{[0, \frac{1}{n^2}]}$ et $g_n = 1_{[0, \frac{1}{n}]}$.

EXERCICE 1.4

Soit X une v. a. r. de fonction de répartition F et de fonction de répartition empirique F_n .

- 1) Calculer $\text{var} F_n(x)$ et $\text{cov}(F_n(x), F_n(y))$ en fonction de F .
- 2) Soit

$$g_n(x) = \frac{F_n(x + h_n) - F_n(x - h_n)}{2h_n} \quad (h_n > 0).$$

Calculer la variance de $g_n(x)$ et $\text{cov}(g_n(x), g_n(y))$ en fonction de F .

- 3) On suppose que X admet une densité f_X , soit $f_n(x)$ l'estimateur à noyau de f_X , où le noyau positif K vérifie les conditions suivantes

$$K1 : \int K(x)dx = 1$$

$$K2 : \sup K(x) \leq M < \infty$$

$$K3 : \int |x^3| K(x)dx < \infty$$

$$K4 : \int x K(x)dx = 0.$$

On suppose que f_X a des dérivées d'ordre inférieur ou égal à 3 sur \mathbb{R} (avec $f_X(x) \neq 0$ et $f_X^{[3]}$ bornée) et que $h_n \xrightarrow{n \rightarrow \infty} 0$. On se propose de trouver la "meilleure" fenêtre, au sens MSE, autrement dit on cherche la fenêtre qui minimise l'erreur quadratique moyenne de $f_n(x)$.

Montrer qu'on a asymptotiquement

i)

$$E f_n(x) = f_X(x) + \frac{h_n^2}{2} \left\{ f_X^{[2]}(x) \int x^2 K(x)dx + o(1) \right\}.$$

ii)

$$\text{Var} f_n(x) = \frac{1}{nh_n} f_X(x) \int K^2(x)dx (1 + o(1)).$$

En déduire la valeur de la fenêtre optimale asymptotiquement.

Chapitre 2

Estimation de la fonction de régression

La méthode connue de la régression linéaire sert à prédire des valeurs pour une variable Y , dite variable expliquée ou dépendante, à partir des valeurs d'une variable X , dite variable explicative ou indépendante. Elle se base sur l'hypothèse d'existence d'une relation linéaire entre X et Y , qui peut être vérifiée graphiquement ou plus rigoureusement par un test. Mais une telle hypothèse n'est pas toujours raisonnable. En statistique non paramétrique, nous introduisons le modèle plus général $Y = r(X) + \varepsilon$ où ε et X sont indépendantes et $E\varepsilon = 0$, ce qui montre que $r(X) = E(Y/X)$. Notre problème se ramène à l'estimation de l'espérance conditionnelle de Y sachant X que nous étudions ci dessous.

2.0.1 Estimateur de Nadaraya-Watson

Soient $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ des couples aléatoires indépendants, à valeurs dans \mathbb{R}^2 , et de même loi que (X, Y) . Le couple de variables aléatoires (X, Y) est supposé admettre une densité jointe sur \mathbb{R}^2 notée $f_{X,Y}$ et nous désignons par $f_X(\cdot)$ la densité marginale de X , donnée par

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy.$$

La densité conditionnelle de Y en $X = x$ est

$$f_{Y/X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{si } f_X(x) \neq 0.$$

L'espérance conditionnelle ou la fonction de régression de Y en $X = x$ s'écrit

$$\begin{aligned} r(x) &= E(Y/X = x) = \int_{-\infty}^{+\infty} y f_{Y/X=x}(y) dy \\ &= \int_{-\infty}^{+\infty} \frac{y f_{X,Y}(x, y)}{f_X(x)} dy. \end{aligned}$$

Soit $J(x, y)$ une fonction de densité sur \mathbb{R}^2 , on pose

$$K(x) = \int_{-\infty}^{+\infty} J(x, y) dy$$

Soit $h_n \rightarrow 0$ quand $n \rightarrow \infty$, on a alors

$$\frac{1}{nh_n^2} \sum_{j=1}^n J \left[\frac{(x - X_j)}{h_n}, \frac{(y - Y_j)}{h_n} \right]$$

est un estimateur de $f_{X,Y}(x, y)$. On a aussi

$$f_n(x) = \frac{1}{nh_n} \sum_{j=1}^n K \left[\frac{(x - X_j)}{h_n} \right].$$

est un estimateur de $f_X(x)$.

Un estimateur naturel de $r(x)$ est alors donné par

$$r_n(x) = \int_{-\infty}^{+\infty} \frac{y \frac{1}{nh_n^2} \sum_{j=1}^n J \left[\frac{(x - X_j)}{h_n}, \frac{(y - Y_j)}{h_n} \right]}{f_n(x)} dy.$$

En posant $z = \frac{y - Y_j}{h_n}$, il vient

$$r_n(x) = \frac{h_n \sum_{j=1}^n p \left(\frac{x - X_j}{h_n} \right)}{\sum_{j=1}^n K \left(\frac{x - X_j}{h_n} \right)} + \frac{\sum_{j=1}^n Y_j K \left(\frac{x - X_j}{h_n} \right)}{\sum_{j=1}^n K \left(\frac{x - X_j}{h_n} \right)},$$

où

$$p(x) = \int_{-\infty}^{+\infty} y J(x, y) dy.$$

En choisissant

$$J(x, y) = K(x)L(y),$$

avec $\int yL(y)dy = 0$, nous obtenons $p(x) = 0$. D'où l'expression de l'estimateur de la fonction de régression :

$$r_n(x) = \begin{cases} \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} & \text{si } \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

C'est l'estimateur de Nadaraya-Watson, introduit simultanément par ces deux auteurs en 1964. Notons que la covariable X peut être à valeurs dans \mathbb{R}^d ($d \in \mathbb{N}^*$), il faut alors aussi prendre un noyau K défini sur \mathbb{R}^d .

Plus généralement, des estimateurs à poids de $r(x)$ peuvent s'écrire

$$\hat{r}_n(x) = \sum_{i=1}^n W_{n,i}(x)Y_i,$$

où $W_{n,i}(x)$ sont des poids dépendant de x et de (X_1, \dots, X_n) . La valeur de $W_{n,i}$ dépend du type d'estimateur considéré, nous citons ci-dessous des exemples connus.

Estimateur des plus proches voisins

Soit k_n un paramètre de l'estimation, on pose

$$W_{n,i}(x) = \begin{cases} \frac{1}{k_n}, & \text{si } X_i \text{ est parmi les } k_n \text{ plus proches voisins de } x \text{ dans } \{X_1, \dots, X_n\}, \\ 0, & \text{sinon.} \end{cases}$$

(Nous renvoyons le lecteur à Devroye et al. (1994)).

Estimateur à partitions

On utilise une partition $\pi_n = \{A_{n,j} : j\}$ de \mathbb{R} et on pose

$$W_{n,i}(x) = \sum_j \frac{1_{A_{n,j}}(X_i)}{\sum_{k=1}^n 1_{A_{n,j}}(X_k)} 1_{A_{n,j}}(x).$$

Pour plus de détails, voir Devroye et Györfi (1983).

Signalons que d'autres types d'estimateurs non paramétriques existent, parmi les quels nous citons l'estimateur des moindres carrés, ou plus généralement

les estimateurs spline de lissage (cf, par exemple Kohler et Krzyżak (2001), Kohler (1997, 1999) et Györfi et al. (1997)), ainsi que des estimateurs à ondelettes (cf Donoho et al. (1995)).

2.0.2 Consistance de l'estimateur de Nadaraya-Watson

Théorème 4 *Soit K une densité vérifiant les hypothèses K_1 et K_2 , si $E(Y^2) < \infty$, $f_X(x)$ est strictement positive, r , f_X et $m(x) = \int y^2 f_{X,Y}(x, y) dy$ sont continues au point x et si $h_n \rightarrow 0$, et $nh_n \rightarrow +\infty$ (quand $n \rightarrow \infty$), alors $r_n(x)$ est un estimateur consistant de $r(x)$.*

Preuve. Puisque $f_n(x)$ est un estimateur consistant de $f_X(x)$, il suffit donc de montrer que $\Phi_n(x) = \frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)$ est un estimateur consistant de

$$\Phi(x) = \int_{-\infty}^{+\infty} y f_{X,Y}(x, y) dy.$$

Nous avons

$$\begin{aligned} E(\Phi_n(x)) &= E\left[\frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} E\left[Y K\left(\frac{x-X}{h_n}\right)\right] \\ &= \frac{1}{h_n} \int_{\mathbb{R}} E(Y/X=t) K\left(\frac{x-t}{h_n}\right) f_X(t) dt \\ &= \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x-t}{h_n}\right) r(t) f_X(t) dt. \\ &= \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x-t}{h_n}\right) \Phi(t) dt \rightarrow \Phi(x) \end{aligned}$$

par le théorème de Bochner. De plus,

$$\begin{aligned}
\text{var}(\Phi_n(x)) &= \frac{1}{nh_n^2} \text{var}(YK[(x-X)/h_n]) \\
&\leq \frac{1}{nh_n^2} E(Y^2 K^2[(x-X)/h_n]) \\
&= \frac{1}{nh_n^2} \int y^2 K^2[(x-u)/h_n] f_{X,Y}(u,y) du dy \\
&= \frac{1}{nh_n} \frac{1}{h_n} \int m(u) K^2[(x-u)/h_n] du \rightarrow 0 \text{ quand } n \rightarrow \infty,
\end{aligned}$$

par le théorème de Bochner et le fait que $nh_n \rightarrow \infty$.

Ceci implique que

$$\Phi_n(x) \xrightarrow{\mathcal{P}} \Phi(x).$$

D'où

$$r_n(x) = \frac{\Phi_n(x)}{f_n(x)} \xrightarrow{\mathcal{P}} \frac{\Phi(x)}{f_X(x)} = r(x).$$

■ Remarquons que l'exercice 2.1 donne des conditions suffisantes pour que cet estimateur soit asymptotiquement sans biais. Bien d'autres résultats concernant cet estimateur existent (comme la normalité asymptotique, la convergence en moyenne quadratique et la convergence presque complète).

EXERCICE 2.1

Soit $r_n(x)$ l'estimateur de Nadaraya-Watson de la fonction de régression $r(x) = E(Y/X = x)$. Soit

$$\Phi_n(x) = \frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{hn}\right)$$

et

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{hn}\right).$$

Supposons que toutes les hypothèses du théorème 4 sont vérifiées et que Y est bornée presque sûrement. Montrer que

- 1) $r_n(x)$ converge en moyenne (c'est à dire dans L^1) vers $r(x)$.
- 2) Montrer que

$$E(r_n(x)) = \frac{E(\Phi_n(x))}{E(f_n(x))} + O\left(\frac{1}{nh_n}\right).$$

(Indication : utiliser l'identité $\frac{1}{u} = \frac{1}{c} - \frac{u-c}{c^2} + \frac{(u-c)^2}{uc^2}$ ($u \neq 0, c \neq 0$)).

- 3) En déduire des conditions suffisantes pour que $r_n(x)$ soit un estimateur asymptotiquement sans biais de $r(x)$.

EXERCICE 2.2

Soit l_x et l_y deux nombres réels et (u_n) une suite de nombres réels tels que $\lim_{n \rightarrow \infty} u_n = 0$, montrer que si

1. Si $\lim_{n \rightarrow \infty} X_n = l_x$ p.co. et $\lim_{n \rightarrow \infty} Y_n = l_y$ p.co., alors

- $\lim_{n \rightarrow \infty} (X_n + Y_n) = l_x + l_y$ p.co.,
- $\lim_{n \rightarrow \infty} (X_n Y_n) = l_x l_y$ p.co.,
- $\lim_{n \rightarrow \infty} \frac{1}{X_n} = \frac{1}{l_x}$ p.co. lorsque $l_x \neq 0$.

2. Si $X_n - l_x = O_{p.co.}(u_n)$ et $Y_n - l_y = O_{p.co.}(u_n)$, alors

- $(X_n + Y_n) - l_x - l_y = O_{p.co.}(u_n)$,
- $(X_n Y_n) - l_x l_y = O_{p.co.}(u_n)$,
- $\frac{1}{X_n} - \frac{1}{l_x} = O_{p.co.}(u_n)$ lorsque $l_x \neq 0$.

3. Si $X_n = O_{p.co.}(u_n)$ et $\lim_{n \rightarrow \infty} Y_n = l_y$ p.co., alors

- $X_n Y_n = O_{p.co.}(u_n)$,
- $\frac{X_n}{Y_n} = O_{p.co.}(u_n)$, lorsque $l_y \neq 0$.

Chapitre 3

Prévision non paramétrique dans un cadre fonctionnel

Une variable est dite fonctionnelle si elle prend ses valeurs dans un espace de dimension infinie, tel est le cas d'une courbe ou d'une surface. Les progrès en informatique ont permis de stocker et de traiter de telles variables que nous pouvons rencontrer et observer dans divers cas dans la pratique.

Supposons donc que nous observons un échantillon de n variables aléatoires indépendantes et de même loi que le couple (X, Y) à valeurs dans $E \times \mathbb{R}$, où E est un espace semi métrique muni de la semi métrique d .

Soit x un élément fixé de E et $h > 0$, nous définissons $B(x, h) = \{x' \in E, d(x', x) \leq h\}$ qui est la boule de centre x et rayon h et nous posons $\phi_X(h) = P(X \in B(x, h))$.

Notre but est de prédire la valeur de la variable réponse réelle Y à partir d'une valeur de X .

Comme exemples tirés de la pratique, nous pouvons citer

- La prédiction de la quantité de matière grasse contenue dans une tranche de viande à partir de la courbe spectrométrique de cette tranche (au lieu d'une analyse biologique fort coûteuse) sur la base de l'échantillon qu'on peut trouver à l'adresse [http : //lob.stat.cmu.edu/datasets/tecator](http://lob.stat.cmu.edu/datasets/tecator).
- La prédiction de la consommation de l'électricité d'un mois d'une année à partir de la consommation de toute l'année précédente (donnée de manière continue) sur la base de l'échantillon qu'on peut trouver à l'adresse : www.economagic.com. Cela peut se faire, entre autres, par l'utilisation d'estimateurs non paramétriques que nous explicitons ci dessous (cf au livre fondateur de Ferraty et Vieu (2006)).

3.1 Méthodes d'estimation

3.1.1 Estimation de la régression

La régression de la variable réponse scalaire Y sur une variable fonctionnelle X est définie par :

$$r(x) = E(Y|X = x).$$

Ferraty et Vieu ont adapté l'estimateur de Nadaraya et Watson au cas où la variable explicative est à valeurs dans un espace semi-métrique en posant

$$\hat{r}_n(x) = \frac{\sum_{i=1}^n Y_i K(h^{-1}d(x, X_i))}{\sum_{i=1}^n K(h^{-1}d(x, X_i))} = \sum_{i=1}^n W_{n,i}(x) Y_i, \quad (3.1)$$

où

$$W_{n,i}(x) = \frac{K(h^{-1}d(x, X_i))}{\sum_{i=1}^n K(h^{-1}d(x, X_i))},$$

($h := h_n$) est une suite de réels positifs et K est une fonction réelle positive .

Cet estimateur est obtenu en moyennant les Y_i suivant une pondération donnant plus de poids aux Y_i pour lesquels les X_i s'approchent de x .

3.1.2 Estimation du quantile conditionnel

La fonction de répartition de la variable aléatoire réelle Y sachant la variable fonctionnelle X , notée F^x , est définie pour tout y dans \mathbb{R} par

$$F^x(y) = P(Y \leq y|X = x).$$

Puisque

$$F^x(y) = E(1_{\{Y \leq y\}}|X = x),$$

nous pouvons estimer $F^x(y)$ par :

$$\frac{\sum_{i=1}^n K(h^{-1}d(x, X_i)) 1_{]-\infty, y]}(Y_i)}{\sum_{i=1}^n K(h^{-1}d(x, X_i))}.$$

En lissant $1_{\{Y \leq y\}}$ dans le but d'obtenir de bonnes propriétés asymptotiques pour l'estimateur introduit, $F^x(y)$ est estimée par

$$\widehat{F}_n^x(y) = \frac{\sum_{i=1}^n K(h^{-1}d(x, X_i))H(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(x, X_i))}, \quad (3.2)$$

où

$$\forall u \in \mathbb{R}, \quad H(u) = \int_{-\infty}^u K_0(v) \, dv,$$

K_0 étant un noyau symétrique et $(g := g_n)$ une suite de nombre réels strictement positifs .

La médiane conditionnelle s'écrit

$$m(x) = \inf\{y \in \mathbb{R}, F^x(y) \geq \frac{1}{2}\}.$$

D'une manière plus générale, les quantiles conditionnels sont définis, pour $\alpha \in]0, 1[$, par

$$t_\alpha(x) = \inf\{y \in \mathbb{R}, F^x(y) \geq \alpha\},$$

A partir de l'estimateur donné en (3.2), nous déduisons les estimateurs respectifs de $m(x)$ et de $t_\alpha(x)$ comme suit

$$\widehat{m}_n(x) = \inf\{y \in \mathbb{R}, \widehat{F}_n^x(y) \geq \frac{1}{2}\}$$

et

$$\widehat{t}_{\alpha,n}(x) = \inf\{y \in \mathbb{R}, \widehat{F}_n^x(y) \geq \alpha\}.$$

3.1.3 Estimation du mode conditionnel

Dans ce paragraphe nous supposons que la loi de probabilité conditionnelle de Y sachant $X = x$ est absolument continue par rapport à la mesure de Lebesgue, désignons par $f^x(y)$ sa densité au point y .

Comme \widehat{F}_n^x est l'estimateur de F^x , il est naturel de proposer, pour estimateur de f^x

$$\forall y \in \mathbb{R}, \quad \widehat{f}_n^x(y) = \frac{\partial}{\partial y} \widehat{F}_n^x(y).$$

Ceci est possible dès que K_0 est continue et nous obtenons alors

$$\hat{f}_n^x(y) = \frac{\sum_{i=1}^n K(h^{-1}d(x, X_i))^{\frac{1}{g}} K_0(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(x, X_i))} \quad (3.3)$$

le mode conditionnel $\theta(x)$, sur un ensemble S , s'écrit :

$$\theta(x) = \arg \sup_{y \in S} f^x(y),$$

(cette définition suppose implicitement que $\theta(x)$ existe).

A partir de l'estimateur donné en 3.3, nous proposons d'estimer le mode conditionnel $\theta(x)$ par

$$\hat{\theta}_n(x) = \arg \sup_{y \in S} \hat{f}_n^x(y).$$

Afin de prouver la convergence presque complète des estimateurs introduits, nous faisons usage de certains outils et résultats que nous introduisons ci dessous.

3.2 Outils préliminaires

3.2.1 Noyaux usuels

Définition 1 Une densité K définie sur \mathbb{R} est appelée

1. noyau de type 0, si son support est $[-1, 1]$ et si pour tout $u \in]0, 1[$, $K(u) > 0$;
2. noyau de type 1, s'il existe deux constantes réelles $0 < C_1 < C_2 < \infty$ telles que :

$$C_1 1_{[0,1]} \leq K \leq 1_{[0,1]} C_2;$$

3. noyau du type 2, si son support est $[0, 1]$ et si sa dérivée K' existe, est continue sur $[0, 1]$ et vérifie

$$C_3 \leq K' \leq C_4,$$

où C_3 et C_4 sont deux constantes réelles telles que $-\infty < C_3 < C_4 < \infty$.

Aux chapitres précédents, le théorème de Bochner a permis de traiter le terme $\frac{EK(\frac{X-x}{h})}{h}$, dans le cadre fonctionnel le terme $\frac{EK(\frac{d(X,x)}{h})}{h}$ qui le remplace est traité grâce au lemme suivant.

Lemme 1

i) Si K est un noyau de type 1, alors il existe deux constantes positives C et C' telles que

$$C\phi_x(h) \leq EK\left(\frac{d(x, X)}{h}\right) \leq C'\phi_x(h).$$

ii) Si K est un noyau de type 2 et si ϕ_x satisfait

$$\exists C'' > 0, \exists \epsilon_0, \forall \epsilon < \epsilon_0, \int_0^\epsilon \phi_x(u) du > C''\epsilon\phi_x(\epsilon), \quad (3.4)$$

alors il existe deux constantes réelles non négatives C et C' , telles qu'on ait pour h suffisamment petit

$$C\phi_x(h) \leq EK\left(\frac{d(x, X)}{h}\right) \leq C'\phi_x(h).$$

Preuve.

i) Par définition d'un noyau de type 1, nous avons

$$C_1 1_{[0,1]} \leq K \leq C_2 1_{[0,1]},$$

ce qui implique que

$$C_1 1_{B(x,h)}(X) \leq K\left(\frac{d(x, X)}{h}\right) \leq C_2 1_{B(x,h)}(X).$$

Le resultat visé s'en déduit par passage à l'espérance en posant $C = C_1$ et $C' = C_2$.

ii) Nous pouvons écrire

$$K(t) = K(0) + \int_0^t K'(u) du,$$

et puisque

$$EK\left(\frac{d(x, X)}{h}\right) = \int_0^1 K(t) dP^{\frac{d(x, X)}{h}}(t),$$

nous obtenons

$$\begin{aligned} EK\left(\frac{d(x, X_i)}{h}\right) &= \int_0^1 K(0) dP^{\frac{d(x, X)}{h}}(t) + \int_0^1 \left(\int_0^t K'(u) du\right) dP^{\frac{d(x, X)}{h}}(t) \\ &= \int_0^1 K(0) dP^{\frac{d(x, X)}{h}}(t) + \int_0^1 \left(\int_0^1 K'(u) 1_{[u, 1]}(t) du\right) dP^{\frac{d(x, X)}{h}}(t). \end{aligned}$$

En appliquant le théorème de Fubini, la relation précédente s'écrit

$$\begin{aligned} EK\left(\frac{d(x, X)}{h}\right) &= K(0) \int_0^1 dP^{\frac{d(x, X)}{h}}(t) + \int_0^1 K'(u) \left(\int_0^1 1_{[u, 1]}(t) dP^{\frac{d(x, X)}{h}}(t)\right) du \\ &= K(0) P(X \in B(x, h)) + \int_0^1 K'(u) P(u \leq \frac{d(x, X)}{h} \leq 1) du \\ &= K(0) \phi_x(h) + \int_0^1 K'(u) [P(\frac{d(x, X)}{h} \leq 1) - P(\frac{d(x, X)}{h} \leq u)] du \\ &= K(0) \phi_x(h) + \int_0^1 K'(u) [\phi_x(h) - \phi_x(hu)] du \\ &= K(0) \phi_x(h) + \phi_x(h) [K(1) - K(0)] - \int_0^1 K'(u) \phi_x(hu) du \\ &= K(1) \phi_x(h) - \int_0^1 K'(u) \phi_x(hu) du. \end{aligned}$$

Si $K(1) \neq 0$, le noyau K est de type 1 (en vertu de l'exercice 3.1) pour lequel l'inégalité recherchée est vérifiée d'après la partie i).

Si $K(1) = 0$, on a :

$$EK\left(\frac{d(x, X)}{h}\right) = - \int_0^1 K'(u) \phi_x(hu) du.$$

Le changement de variable suivant $t = hu$ donne

$$EK\left(\frac{d(x, X)}{h}\right) = -\frac{1}{h} \int_0^h K'\left(\frac{t}{h}\right) \phi_x(t) dt,$$

K étant un noyau de type 2, combiné avec la relation (3.4), nous donne pour $h < \epsilon_0$

$$EK\left(\frac{d(x, X)}{h}\right) \geq -\frac{1}{h} \cdot h C''' C_3 \phi_x(h) \geq C \phi_x(h),$$

où

$$C = -C''C_3.$$

Comme K est borné et à support dans $[0, 1]$, en posant $C' = \sup_{t \in [0, 1]} K(t)$, les mêmes arguments utilisés en i) impliquent que

$$EK \left(\frac{d(x, X)}{h} \right) \leq C' \phi_x(h). \quad (3.5)$$

■

3.2.2 Inégalités exponentielles de type Bernstein

Les inégalités exponentielles suivantes nous serviront à montrer la convergence presque compète des estimateurs que nous étudions, la première ne suppose pas la bornitude de la variable (mais impose une condition plus générale) tandis que la seconde n'est valable sous cette condition.

Proposition 1 (cf corollaire A.8 dans Ferraty et Vieu (2006))

Soit $(Z_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires indépendantes et centrées.

i) Si $\forall m \geq 2, \exists C_m > 0 ; E|Z_1^m| \leq C_m a^{2(m-1)}$, nous avons

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

ii) Si les variables Z_n dépendent de n (i.e. $Z_i := Z_{i,n}$ et si $\forall m \geq 2, \exists C_m > 0 ; E|Z_1^m| \leq C_m a_n^{2(m-1)}$ et si $u_n = n^{-1} a_n^2 \log n$ vérifie $\lim_{n \rightarrow \infty} u_n = 0$, nous avons

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

Proposition 2 (cf corollaire A.9 dans Ferraty et Vieu (2006))

Soit $(Z_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires indépendantes et centrées.

i) Si $\exists M < \infty ; |Z_1| \leq M$ et si $EZ_1^2 \leq \sigma^2$, nous avons

$$\forall \epsilon > 0, P \left(\left| \sum_{i=1}^n Z_i \right| > \epsilon n \right) \leq 2 \exp \left(-\frac{\epsilon^2 n}{2\sigma^2(1 + \epsilon \frac{M}{\sigma^2})} \right).$$

ii) Si les variables Z_n dépendent de n (i.e. $Z_i := Z_{i,n}$) et si $\exists M = M_n < \infty$ telle que $|Z_1| \leq M_n$, si $EZ_1^2 \leq \sigma_n^2$, $u_n = n^{-1}\sigma_n^2 \log n$ telle que $\lim_{n \rightarrow \infty} u_n = 0$ et $\lim_{n \rightarrow \infty} \frac{M\sqrt{u_n}}{\sigma_n^2} = 0$, nous avons

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

3.3 Résultats asymptotiques

Remarquons que dans les preuves données dans la suite de ce document, les constantes (par rapport à la taille n de l'échantillon) utilisées sont génériques.

3.3.1 Convergence presque complète de l'estimateur de la régression

Nous avons besoin des hypothèses suivantes.

- $H_1 : \lim_{d(y,x) \rightarrow 0} r(y) = r(x)$ (hypothèse de continuité).
- $H_2 : \forall h > 0, P(X \in B(x, h)) = \phi_x(h) > 0$.
- H_3 : le paramètre de lissage h , le noyau K et $\Phi_x(h)$ sont tels que :

- $$\left\{ \begin{array}{l} h \text{ est une suite de réels positifs tels que} \\ \lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{\log n}{n\phi_x(h)} = 0, \\ K \text{ est un noyau de type 1} \end{array} \right. \text{ ou } \left\{ \begin{array}{l} K \text{ est un noyau de type 2 et } \phi_x(.) \text{ satisfait la condition (3.4).} \end{array} \right.$$
- $H_4 : \forall m \geq 2$, les moments conditionnels d'ordre m de la variable aléatoire réelle Y vérifie :

$$\mathbb{E}(|Y^m| | X = x) < \sigma_m(x) < \infty \quad (3.6)$$

avec $\sigma_m(.)$ continue en x

Théorème 5 *Sous les hypothèses $H_1 - H_4$, nous avons*

$$\lim_{n \rightarrow \infty} \hat{r}_n(x) = r(x) \quad p.c.o.$$

Preuve.

En posant

$$\Delta_i = \frac{K\left(\frac{d(x, X_i)}{h}\right)}{EK\left(\frac{d(x, X_1)}{h}\right)}, \quad (3.7)$$

la démonstration de ce résultat est basée sur la décomposition suivante

$$\widehat{r}_n(x) - r(x) = \frac{1}{\widehat{r}_1(x)}[(\widehat{r}_2(x) - E(\widehat{r}_2(x)) - (r(x) - E(\widehat{r}_2(x))))] - \frac{r(x)}{\widehat{r}_1(x)}[\widehat{r}_1(x) - 1] \quad (3.8)$$

où

$$\widehat{r}_1(x) = \frac{1}{n} \sum_{i=1}^n \Delta_i \quad \text{et} \quad \widehat{r}_2(x) = \frac{1}{n} \sum_{i=1}^n \Delta_i Y_i,$$

Remarquons que Le lemme 1 et l'hypothèse H_2 assurent que $E(K(\frac{d(x, X)}{h})) > 0$. Les termes de cette décomposition sont traités dans les lemmes 2 et 3 .

Lemme 2 *Sous les hypothèses H_1 et H_3 , nous obtenons*

$$\lim_{n \rightarrow \infty} E\widehat{r}_2(x) = r(x).$$

Preuve.

En utilisant l'équidistribution des données, le fait que $E\Delta_1 = 1$ et que Δ_1 est X_1 mesurable, il vient

$$\begin{aligned} r(x) - E\widehat{r}_2(x) &= r(x) - E\left(\frac{1}{n} \sum_{i=1}^n \Delta_i Y_i\right) \\ &= r(x) - E(\Delta_1 Y_1) \\ &= r(x) - E(E(\Delta_1 Y_1 | X_1)) \\ &= r(x)E(\Delta_1) - E[\Delta_1 r(X_1)]. \end{aligned}$$

Donc

$$r(x) - E\widehat{r}_2(x) = E[(r(x) - r(X_1))\Delta_1]$$

Comme K est à support compact sur $[0, 1]$, nous obtenons

$$|r(x) - r(X_1)|\Delta_1 \leq \sup_{x' \in B(x, h)} |r(x) - r(x')|\Delta_1.$$

Il reste à appliquer l'hypothèse de continuité de r (H_1) pour déduire le résultat visé. ■

Lemme 3

i) Sous les hypothèses $H_1 - H_4$, nous obtenons

$$\widehat{r}_2(x) - E\widehat{r}_2(x) = O_{p.c.o}(\sqrt{\frac{\log n}{n\phi_x(h)}}).$$

ii) Sous les hypothèses $H_1 - H_3$, nous avons

$$\widehat{r}_1(x) - 1 = O_{p.c.o}(\sqrt{\frac{\log n}{n\phi_x(h)}}).$$

Preuve. :

i) On a

$$\widehat{r}_2(x) - E(\widehat{r}_2(x)) = \frac{1}{n} \sum_{i=1}^n Z_i,$$

où

$$Z_i = Y_i \Delta_i - E(Y_i \Delta_i).$$

Posons $K_i = K(h^{-1}d(x, X_i))$, nous pouvons écrire pour $m \geq 2$

$$\begin{aligned} E[|Y_1|^m \Delta_1^m] &= \frac{1}{(EK_1)^m} E[|Y_1|^m K_1^m] \\ &= \frac{1}{(EK_1)^m} E[E(|Y_1|^m | X) K_1^m] \\ &\leq \frac{1}{(EK_1)^m} E[\sigma_m(X) K_1^m] \\ &\leq \frac{1}{(EK_1)^m} E[(\sigma_m(X) - \sigma_m(x) + \sigma_m(x)) K_1^m] \\ &\leq \frac{1}{(EK_1)^m} [E(\sigma_m(X) - \sigma_m(x)) K_1^m + \sigma_m(x) K_1^m]. \end{aligned}$$

D'où

$$E[|Y_1|^m \Delta_1^m] \leq \sup_{x' \in B(x, h)} |\sigma_m(x') - \sigma_m(x)| E\Delta_1^m + \sigma_m(x) E\Delta_1^m.$$

Puisque si K est de type 1 (resp. de type 2), alors $K^m / \int K^m$ est aussi de type 1 (resp. de type 2), en appliquant le lemme 1 à K et $K^m / \int K^m$ simultanément, nous obtenons

$$\frac{C_1 \phi_x(h)}{\phi_x^m(h)} \leq E(\Delta_1^m) = \frac{EK_1^m}{(EK_1)^m} \leq \frac{C_2 \phi_x(h)}{\phi_x^m(h)}. \quad (3.9)$$

Ce qui implique que

$$\forall m \geq 2, \quad E|Y_1|^m \Delta_1^m = O(\phi_x^{-m+1}(h)). \quad (3.10)$$

De plus on a $(Y_1\Delta_1 - E(Y_1\Delta_1))^m = \sum_{k=0}^m C_m^k (-1)^{m-k} (Y_1\Delta_1)^k (EY_1\Delta_1)^{m-k}$.

Ce qui nous donne puisque $\phi_x(h) \leq 1$

$$\begin{aligned} E|Y_1\Delta_1 - EY_1\Delta_1|^m &\leq \sum_{k=0}^m C_m^k E|(Y_1\Delta_1)^k| |(EY_1\Delta_1)^{m-k}| \\ &\leq C \sum_{k=0}^m C_m^k E|(Y_1\Delta_1)^k| \\ &\leq C \max_{k=0,\dots,m} \phi_x^{-k+1}(h), \end{aligned} \quad (3.11)$$

par application de l'inégalité (3.10) pour $k \geq 2$ et pour $k = 0$ ou $k = 1$, il suffit d'utiliser le lemme 2 pour avoir $|EY_1\Delta_1| = E(\widehat{r}_2(x)) = O(1)$.

Nous obtenons donc

$$E|Y_1\Delta_1 - EY_1\Delta_1|^m = O(\phi_x(h)^{-m+1}).$$

En posant $a_n^2 = \phi_x^{-1}(h)$, on a pour $u_n = (a_n^2 \log n)/n$, en vertu de l'hypothèse H_3 ,

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \frac{\log n}{n\phi_x(h)} = 0.$$

Les conditions de la proposition 1 étant satisfaites, nous déduisons

$$\frac{1}{n} \sum_{i=1}^n Z_i = O_{p.c.o}(\sqrt{u_n}).$$

(ii) En posant $Y = 1$ dans (i), nous obtenons (ii). ■

3.3.2 Convergence presque complète de l'estimateur de la médiane conditionnelle

Afin de donner un résultat de consistance forte pour l'estimateur de la médiane conditionnelle, nous avons besoin d'étudier l'estimateur de la fonction de répartition conditionnelle. À cette fin, nous introduisons de nouvelles hypothèses. La première (H_5) impose des conditions de régularité à l'application $(x, y) \rightarrow F^x(y)$ et la seconde (H_6) concerne la fenêtre g et le noyau H

nouvellement introduits pour l'estimateur de F^x .

L'hypothèse H_5 , englobe les conditions suivantes.

$$\begin{cases} H_{5,1} : \text{L'application } y \rightarrow F^x(y) \text{ est une fonction de répartition strictement croissante.} \\ H_{5,2} : \forall y \in \mathbb{R}, \lim_{d(x',x) \rightarrow 0} F^{x'}(y) = F^x(y) \text{ et } \lim_{|y'-y| \rightarrow 0} F^x(y') = F^x(y). \end{cases}$$

L'hypothèse H_6 , quant à elle, comporte les conditions suivantes.

$$\begin{cases} H_{6,1} : g \text{ est une suite de réels positifs tels que } \lim_{n \rightarrow \infty} g = 0, \\ H_{6,2} : K_0 = H' \text{ est de type zéro.} \end{cases}$$

Remarquons que la condition H_5 (resp. $H_{6,2}$) assure que la médiane conditionnelle $m(x)$ (resp. son estimateur $\hat{m}_n(x)$) soit la solution unique de l'équation

$$m(x) = (F^x)^{-1}(1/2), \quad (3.12)$$

$$(\text{resp. de l'équation } \hat{m}_n(x) = (\hat{F}_n^x)^{-1}(1/2)). \quad (3.13)$$

Nous pouvons maintenant énoncer le résultat de convergence presque complète pour l'estimateur de la fonction de répartition conditionnelle comme suit.

Proposition 3 *Sous les hypothèses $H_2 - H_3$ et $H_5 - H_6$, nous avons*

$$\lim_{n \rightarrow \infty} \hat{F}_n^x(y) = F^x(y) \text{ p.co.}$$

Preuve. :

Cette preuve se base sur la décomposition suivante.

$$\hat{F}_n^x(y) - F^x(y) = \frac{1}{\hat{r}_1(x)} [(\hat{r}_3(x, y) - E\hat{r}_3(x, y)) - (F^x(y) - E\hat{r}_3(x, y))] - \frac{F^x(y)}{\hat{r}_1(x)} [\hat{r}_1(x) - 1] \quad (3.14)$$

où

$$\hat{r}_3(x) = \frac{1}{n} \sum_{i=1}^n \Delta_i \Gamma_i,$$

avec

$$\Gamma_i(y) = H(g^{-1}(y - Y_i)).$$

Rappelons que $\hat{r}_1(x) = \frac{1}{n} \sum_{i=1}^n \Delta_i$, où la quantité Δ_i est définie à la relation (3.7).

Puisque le dernier terme et le dénominateur de la décomposition (3.14) ont été traités dans le lemme 3, il suffit de montrer que

$$\lim_{n \rightarrow \infty} E(\widehat{r}_3(x, y)) = F^x(y) \quad (3.15)$$

et

$$\lim_{n \rightarrow \infty} \widehat{r}_3(x, y) - E(\widehat{r}_3(x, y)) = O_{p.c.o} \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right). \quad (3.16)$$

• Preuve de la relation (3.15)

Comme K est à support compact sur $[0, 1]$ et $E\Delta_1 = 1$, nous pouvons écrire

$$\begin{aligned} E\widehat{r}_3(x, y) - F^x(y) &= E(\Delta_1 \Gamma_1(y)) - F^x(y) \\ &= E[\Delta_1(E \Gamma_1(y)|X_1) - F^x(y)] \\ &= E[\Delta_1 1_{B(x, h)}(X_1)(E[\Gamma_1(y)|X_1] - F^x(y))]. \end{aligned} \quad (3.17)$$

En appliquant le théorème de Fubini et en utilisant le fait que $H(u) = \int_{-\infty}^u K_0(t) dt$, nous obtenons

$$\begin{aligned} E[\Gamma_1(y)|X_1] &= \int_{\mathbb{R}} H(g^{-1}(y - u)) dP(u|X_1) \\ &= \int_{\mathbb{R}} \int_{-\infty}^{g^{-1}(y - u)} K_0(v) dv dP(u|X_1) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_0(v) 1_{[v, \infty[}(g^{-1}(y - u)) dv dP(u|X_1) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_0(v) 1_{[v, \infty[}(g^{-1}(y - u)) dP(u|X_1) dv \\ &= \int_{\mathbb{R}} K_0(v) F^{X_1}(y - gv) dv \end{aligned}$$

Comme K_0 est une densité, il vient

$$E[\Gamma_1(y)|X_1] - F^x(y) = \int_{\mathbb{R}} K_0(v)(F^{X_1}(y - gv) - F^x(y)) dv. \quad (3.18)$$

De plus, nous avons

$$|F^{X_1}(y - gv) - F^x(y)| \leq |F^{X_1}(y - gv) - F^x(y - gv)| + |F^x(y - gv) - F^x(y)|. \quad (3.19)$$

Comme K_0 est à support sur $[-1, 1]$ et g et h tendent vers zéro, l'hypothèse H_5 de continuité de $F^x(y)$ entraîne

$$\lim_{n \rightarrow \infty} \sup_{v \in [-1, 1]} 1_{B(x, h)}(X_1) |F^{X_1}(y - gv) - F_Y^x(y - gv)| = 0 \quad (3.20)$$

et

$$\lim_{n \rightarrow \infty} \sup_{v \in [-1, 1]} [F_Y^x(y - gv) - F_Y^x(y)] = 0. \quad (3.21)$$

Il suffit de combiner ces deux derniers résultats avec (3.17), (3.18) et (3.19) pour aboutir à (3.15).

• Pour prouver (3.16), nous utilisons la décomposition suivante

$$\widehat{r}_3(x, y) - E\widehat{r}_3(x, y) = \frac{1}{n} \sum_{i=1}^n (T_i - ET_i),$$

où

$$T_i = \Delta_i \Gamma_i(y).$$

En utilisant le fait que $\Gamma_i(y)$ est bornée, nous prouvons à l'exercice 3.2 que

$$E|T_1 - ET_1|^m = O(\phi_x(h)^{-m+1}). \quad (3.22)$$

Il suffit maintenant d'appliquer la Proposition 1 à la suite $u_n = \frac{\log n}{n\phi_x(h)}$ pour obtenir

$$\widehat{r}_3(x, y) - E\widehat{r}_3(x, y) = O_{a.c.o} \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right),$$

qui est le résultat recherché. ■

Nous sommes maintenant en mesure de montrer la consistance forte de l'estimateur de la médiane conditionnelle.

Théorème 6 *Sous les hypothèses $H_2 - H_3$ et $H_5 - H_6$, nous avons*

$$\lim_{n \rightarrow \infty} \widehat{m}_n(x) = m(x) \quad p.c.o.$$

Preuve. Sous l'hypothèse H_6 , l'estimateur $(\hat{F}_n^x)^{-1}$ existe et est continu, en particulier sa continuité au point $\hat{F}_n^x(m(x))$ donne

$$\forall \epsilon > 0 \exists \delta(\epsilon) > 0, \forall y / |\hat{F}_n^x(y) - \hat{F}_n^x(m(x))| \leq \delta(\epsilon) \Rightarrow |y - m(x)| \leq \epsilon.$$

En posant $y = \hat{m}_n(x)$, la relation précédente s'écrit

$$\forall \epsilon > 0 \exists \delta(\epsilon) > 0 / |\hat{F}_n^x(\hat{m}_n(x)) - \hat{F}_n^x(m(x))| \leq \delta(\epsilon) \Rightarrow |\hat{m}_n(x) - m(x)| \leq \epsilon.$$

Ce qui permet d'écrire

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, P[|\hat{m}_n(x) - m(x)| > \epsilon] \leq P[|\hat{F}_n^x(\hat{m}_n(x)) - \hat{F}_n^x(m(x))| > \delta(\epsilon)].$$

Puisque d'après (3.12) et (3.3.2), nous avons

$$\hat{F}_n^x(\hat{m}_n(x)) = F^x(m(x)) = 1/2,$$

il vient

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, P[|\hat{m}_n(x) - m(x)| > \epsilon] \leq P[|F^x(m(x)) - \hat{F}_n^x(m(x))| > \delta(\epsilon)].$$

Il reste maintenant à utiliser la proposition 3 pour aboutir à

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} P[|\hat{m}_n(x) - m(x)| > \epsilon] < \infty.$$

Autrement dit, le théorème est démontré. ■

3.3.3 Convergence presque complète du mode conditionnel

Afin d'établir la convergence presque complète de l'estimateur du mode conditionnel $\hat{\theta}_n$, nous avons besoin de montrer la convergence uniforme (cette fois ci) en y de l'estimateur de la densité conditionnelle \hat{f}_n^x . Pour cela, nous faisons usage d'hypothèses que nous introduisons.

L'hypothèse H_7 se compose des deux conditions suivantes.

$$\left\{ \begin{array}{l} H_{7,1} : \exists \xi > 0, \exists ! y_0 \in S / y \rightarrow f^x(y) \text{ est strictement croissante sur }]y_0 - \xi, y_0[\\ \text{et strictement décroissante sur }]y_0, y_0 + \xi[. \\ H_{7,2} : \lim_{d(x', x) \rightarrow 0} f^{x'}(y) = f^x(y) \text{ uniformément sur } y \in S \text{ et} \\ \forall y \in \mathbb{R}, \lim_{|y' - y| \rightarrow 0} f^x(y') = f^x(y). \end{array} \right.$$

L'hypothèse H_8 est aussi constituée de deux conditions, à savoir.

$$\begin{cases} H_{8,1} : \exists C < \infty, \forall (x, x') \in \mathbb{R} \times \mathbb{R}, |K_0(x) - K_0(x')| \leq C |x - x'|. \\ H_{8,2} : \lim_{n \rightarrow \infty} \frac{\log n}{n g \phi_x(h)} = 0 \text{ et } \exists \xi \geq 0, \lim_{n \rightarrow \infty} g n^\xi = \infty. \end{cases}$$

Proposition 4 *Sous les hypothèses $H_2 - H_3$ et $H_6 - H_8$, nous avons*

$$\lim_{n \rightarrow \infty} \sup_{y \in S} |\hat{f}_n^x(y) - f^x(y)| = 0 \text{ p.co.}$$

Preuve. Nous pouvons décomposer $\hat{f}_n^x(y) - f^x(y)$ comme suit

$$\hat{f}_n^x(y) - f^x(y) = \frac{(\hat{r}_4(x, y) - E\hat{r}_4(x, y)) - (f_Y^x(y) - E\hat{r}_4(x, y))}{\hat{r}_1(x)} - \frac{f^x(y)}{\hat{r}_1(x)} \{\hat{r}_1(x) - 1\}. \quad (3.23)$$

où (rappelons le)

$$\hat{r}_1(x) = \frac{1}{n} \sum_{i=1}^n \Delta_i, \quad \Delta_i = \frac{K\left(\frac{d(x, X_i)}{h}\right)}{EK\left(\frac{d(x, X_1)}{h}\right)},$$

quant à $\hat{r}_4(x, y)$, il est défini par

$$\hat{r}_4(x, y) = \hat{r}_1(x) \hat{f}_n^x(y) = \frac{1}{n} \sum_{i=1}^n \Delta_i \Omega_i(y),$$

avec

$$\Omega_i(y) = g^{-1} K_0(g^{-1}(y - Y_i)).$$

Puisque nous avons déjà prouvé au lemme 3 que

$$\lim_{n \rightarrow \infty} \frac{\hat{r}_1(x) - 1}{\hat{r}_1(x)} = O_{p.co} \left(\sqrt{\frac{\log n}{n \phi_x(h)}} \right)$$

et puisque $y \rightarrow f^x(y)$ est bornée sur S (par continuité sur un compact), le dernier terme de la décomposition vérifie

$$\lim_{n \rightarrow \infty} \sup_{y \in S} \frac{f^x(y)}{\hat{r}_1(x)} \{\hat{r}_1(x) - 1\} = 0 \text{ p.co.}$$

Il reste à traiter les deux autres termes de la décomposition, autrement dit il suffit de montrer

$$\lim_{n \rightarrow \infty} \frac{1}{\widehat{r}_1(x)} \sup_{y \in S} |E\widehat{r}_4(x, y) - f^x(y)| = 0 \text{ p.co.} \quad (3.24)$$

et

$$\lim_{n \rightarrow \infty} \frac{1}{\widehat{r}_1(x)} \sup_{y \in S} |\widehat{r}_4(x, y) - E\widehat{r}_4(x, y)| = 0 \text{ p.co.} \quad (3.25)$$

• **Preuve de (3.24)**

Nous avons

$$\begin{aligned} E\widehat{r}_4(x, y) - f^x(y) &= E(\Delta_1 \Omega_1(y)) - f^x(y) \\ &= E(E(\Delta_1 \Omega_1(y) | X_1)) - f^x(y). \end{aligned}$$

K_0 étant une densité et $E\Delta_1 = 1$, il vient

$$\begin{aligned} E\widehat{r}_4(x, y) - f^x(y) &= E \left[\Delta_1 \left(\int_{\mathbb{R}} g^{-1} K_0(g^{-1}(y - u)) dP(u | X_1) - f^x(y) \right) \right] \\ &= E \left[\Delta_1 \left(\int_{\mathbb{R}} g^{-1} K_0(g^{-1}(y - u)) f^{X_1}(u) du - f^x(y) \right) \right] \\ &= E \left[\Delta_1 \left(\int_{\mathbb{R}} K_0(v) (f^{X_1}(y - gv) - f^x(y)) dv \right) \right] \\ &= E \left[\Delta_1 1_{B(x, h)}(X_1) \left(\int_{\mathbb{R}} K_0(v) (f^{X_1}(y - vg) - f^x(y)) dv \right) \right]. \end{aligned}$$

De plus, nous avons

$$|f^{X_1}(y - gv) - f^x(y)| \leq |f^{X_1}(y - gv) - f^x(y - gv)| + |f^x(y - gv) - f^x(y)|.$$

En utilisant le fait que h et g tendent vers zéro (hypothèses H_3 et $H_{6,1}$) et l'hypothèse de continuité $H_{7,2}$, nous pouvons déduire que

$$\lim_{n \rightarrow \infty} \sup_{v \in [-1, 1]} \sup_{y \in S} 1_{B(x, h)}(X_1) |f^{X_1}(y - gv) - f^x(y - gv)| = 0$$

et

$$\lim_{n \rightarrow \infty} \sup_{v \in [-1, 1]} \sup_{y \in S} |f^x(y - gv) - f^x(y)| = 0,$$

il en résulte puisque K_0 est à support sur $[-1, 1]$ que

$$\sup_{y \in S} |E\widehat{r}_4(x, y) - f^x(y)| = o(1).$$

Ceci, combiné avec le lemme 3 conduit à l'obtention de (3.24) .

• **Preuve de (3.25)**

S étant un compact de \mathbb{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

$$l_n = Cz_n^{-1} \sim cn^{-2\xi}, \quad (3.26)$$

où ξ est défini à l'hypothèse $H_{8,2}$. Posons

$$t_y = \arg \min_{t \in \{t_1, \dots, t_{z_n}\}} |y - t|.$$

Nous avons

$$\frac{1}{\widehat{r}_1(x)} \sup_{y \in S} |\widehat{r}_4(x, y) - E\widehat{r}_4(x, y)| \leq A_{1,n} + A_{2,n} + A_{3,n}$$

avec

$$\begin{aligned} A_{1,n} &= \frac{1}{\widehat{r}_1(x)} \sup_{y \in S} |\widehat{r}_4(x, y) - \widehat{r}_4(x, t_y)|, \\ A_{2,n} &= \frac{1}{\widehat{r}_1(x)} \sup_{y \in S} |\widehat{r}_4(x, t_y) - E\widehat{r}_4(x, t_y)|, \\ A_{3,n} &= \frac{1}{\widehat{r}_1(x)} \sup_{y \in S} |E\widehat{r}_4(x, t_y) - E\widehat{r}_4(x, y)|. \end{aligned}$$

- Etude du terme $A_{1,n}$

Comme K_0 est Lipschitzien d'ordre 1 (hypothèse $H_{8,1}$), nous avons

$$\begin{aligned}
|\widehat{r}_4(x, y) - \widehat{r}_4(x, t_y)| &= \frac{1}{n} \sum_{i=1}^n \Delta_i |\Omega_i(y) - \Omega_i(t_y)| \\
&= \frac{1}{n} \sum_{i=1}^n \Delta_i |K_0(g^{-1}(y - Y_i)) - K_0(g^{-1}(t_y - Y_i))| \\
&\leq \frac{C}{ng} \sum_{i=1}^n \Delta_i \frac{|y - t_y|}{g} \\
&\leq C \widehat{r}_1(x) l_n g^{-2}.
\end{aligned} \tag{3.27}$$

Il reste à appliquer la condition (3.26) pour arriver à

$$A_{1,n} \leq \frac{C}{(g n^\xi)^2},$$

qui tend vers zéro en raison de l'hypothèse $H_{8,2}$.

- Etude du terme $A_{3,n}$

Puisque

$$|E\widehat{r}_4(x, t_y) - E\widehat{r}_4(x, y)| \leq E|\widehat{r}_4(x, t_y) - \widehat{r}_4(x, y)|,$$

La relation (3.27), combinée avec le fait que $\widehat{r}_1(x) \rightarrow 1$ par le lemme 3, permet alors de conclure que

$$\lim_{n \rightarrow \infty} A_{3,n} = 0 \text{ p.co.}$$

- Etude du terme $A_{2,n}$

$$\begin{aligned}
\forall \varepsilon, \quad P(\sup_{y \in S} |\widehat{r}_4(x, t_y) - E\widehat{r}_4(x, t_y)| > \varepsilon) \\
&= P(\max_{j=1, \dots, z_n} |\widehat{r}_4(x, t_j) - E\widehat{r}_4(x, t_j)| > \varepsilon) \\
&\leq z_n \max_{j=1, \dots, z_n} P(|\widehat{r}_4(x, t_j) - E\widehat{r}_4(x, t_j)| > \varepsilon) \\
&\leq z_n \max_{j=1, \dots, z_n} P(|\frac{1}{n} \sum_i^n (U_i - E(U_i))| > \varepsilon),
\end{aligned}$$

où

$$U_i = \Delta_i \Omega_i(t_j).$$

Calculons le moment d'ordre m ($m \geq 1$) de U_i

$$\begin{aligned} EU_i^m &= E [\Delta_i^m (E(\Omega_i^m(t_j)/X = x))] \\ &= E \left[\Delta_i^m \left(\int_R g^{-m} K_0^m(g^{-1}(t_j - u)) f^x(u) du \right) \right] \\ &= E \left[\Delta_i^m g^{1-m} \left(\int_R K_0^m(z) f^x(t_j - zg) dz \right) \right]. \end{aligned}$$

Puisque K_0 est de type 0 et f^x est bornée dans cette intégrale (car continue), l'application du résultat donné en (3.9), permet d'obtenir l'existence d'une constante finie C tel que

$$E(U_i^m) \leq C(g\phi_x(h))^{1-m}.$$

IL reste alors à appliquer encore une fois la formule du binôme pour aboutir à

$$\forall m \geq, \quad 2E|U_i - EU_i|^m \leq C(g\phi_x(h))^{1-m}.$$

Nous somme donc en mesure d'appliquer la proposition 1 pour arriver à

$$\forall \epsilon > 0, \quad P[\sup_{\{y \in S\}} |\hat{r}_4(x, t_y) - E\hat{r}_4(x, t_y)| \geq \epsilon] \leq z_n \exp(-C n \epsilon^2 g\phi_x(h)),$$

et on en déduit, en vertu de la relation (3.26), que

$$P[\sup_{\{y \in S\}} |\hat{r}_4(x, t_y) - E\hat{r}_4(x, t_y)| \geq \epsilon] \leq C n^{2\xi} \exp(-C n \epsilon^2 g\phi_x(h)).$$

Comme $\frac{\log n}{ng\phi_x(h)}$ tend vers zéro, le terme de droite de l'inégalité précédente est le terme général d'une série convergente en choisissant $\frac{\log n}{ng\phi_x(h)} < C \frac{\epsilon^2}{2\xi+1}$. Nous arrivons finalement à

$$\forall \epsilon > 0, \quad \sum_{y \in S} P[\sup_{y \in S} |\hat{r}_4(x, t_y) - E\hat{r}_4(x, t_y)| > \epsilon] < \infty$$

■

Il nous est maintenant possible d'établir la convergence presque complète de l'estimateur $\hat{\theta}_n(x)$ vers $\theta(x)$.

Théorème 7 *Sous les hypothèses $H_2 - H_3$ et $H_6 - H_8$, nous avons*

$$\lim_{n \rightarrow \infty} \widehat{\theta}_n(x) = \theta(x) \text{ p.co.}$$

Preuve. L'hypothèse H_7 assure

- d'une part que y_0 est le seul mode noté ($\theta(x)$) et le compact S peut être pris égal à $[\theta(x) - \xi, \theta(x) + \xi]$

et

- d'autre part la continuité et la monotonie stricte de $f^x(\cdot)$ sur $(\theta(x) - \xi, \theta(x))$ et sur $(\theta(x), \theta(x) + \xi)$. la fonction $(f^x)^{-1}$ existe donc et est continue au point $f^x(\theta(x))$, ce qui implique que

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, \forall y \in (\theta(x) - \xi, \theta(x) + \xi), |f^x(y) - f^x(\theta(x))| \leq \delta(\varepsilon) \implies |y - \theta(x)| \leq \varepsilon.$$

Comme $\widehat{\theta}_n(x) \in [\theta(x) - \xi, \theta(x) + \xi]$ par définition,

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, |f^x(\widehat{\theta}_n(x)) - f^x(\theta(x))| \leq \delta(\varepsilon) \implies |\widehat{\theta}_n(x) - \theta(x)| \leq \varepsilon.$$

Nous obtenons donc

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, P[|\widehat{\theta}_n(x) - \theta(x)| > \varepsilon] \leq P[|f^x(\widehat{\theta}_n(x)) - f^x(\theta(x))| > \delta(\varepsilon)].$$

D'autre, part il vient directement des définitions de $\widehat{\theta}_n(x)$ et $\theta(x)$ que

$$\begin{aligned} |f^x(\theta(x)) - f^x(\widehat{\theta}_n(x))| &= f^x(\theta(x)) - f^x(\widehat{\theta}_n(x)) \\ &= f^x(\theta(x)) - \widehat{f}_n^x(\theta(x)) + \widehat{f}_n^x(\theta(x)) - f^x(\widehat{\theta}_n(x)) \\ &\leq [f^x(\theta(x)) - \widehat{f}_n^x(\theta(x))] + [\widehat{f}_n^x(\widehat{\theta}_n(x)) - f^x(\widehat{\theta}_n(x))] \\ &\leq 2 \sup_{y \in [\theta(x) - \xi, \theta(x) + \xi]} |f^x(y) - \widehat{f}_n^x(y)|. \end{aligned}$$

Il reste à appliquer la convergence uniforme presque complète de l'estimateur \widehat{f}_n^x sur le compact $[\theta(x) - \xi, \theta(x) + \xi]$, montrée à la proposition 4, pour arriver à

$$\sum_{n=1}^{\infty} P[|\widehat{\theta}_n(x) - \theta(x)| > \varepsilon] < \infty,$$

qui est le résultat visé. ■

3.4 Taux de convergence

Afin d'établir quelques taux de convergence, nous avons besoin de renforcer les hypothèses de régularité imposées aux fonctions à estimer. Les techniques de démonstration utilisées se basent sur la propriété de Lipschitz.

3.4.1 Taux de convergence pour l'estimateur de la régression

Nous imposons à r d'être lipschitzienne, autrement dit

$$H'_1 : \exists \beta > 0, \exists C > 0, / \forall x' \in E \quad |r(x) - r(x')| \leq C d^\beta(x, x').$$

Nous obtenons ainsi le taux de convergence de \hat{r}_n donné comme suit.

Théorème 8 *Sous les hypothèses H'_1 et $H_2 - H_4$, nous avons*

$$\hat{r}_n(x) - r(x) = O(h^\beta) + O_{p.co} \left(\sqrt{\frac{\log n}{n \phi_x(h)}} \right).$$

Preuve. Cette preuve se fait comme celle du théorème 5 sauf qu'il faut remplacer le lemme 2 par le lemme suivant. ■

Lemme 4 *Sous les hypothèses H'_1 et H_3 , nous avons*

$$E(\hat{r}_2(x)) - r(x) = O(h^\beta).$$

Preuve. Nous avons

$$\begin{aligned} r(x) - E(\hat{r}_2(x)) &= E[\Delta_1 1_{B(x,h)}(X_1)(r(x) - r(X_1))] \\ &\leq CE(\Delta_1 1_{B(x,h)}(X_1) d^\beta(x, X_1)) \leq Ch^\beta, \end{aligned}$$

par application de l'hypothèse H'_1 et le fait que $E(\Delta_1) = 1$. ■

3.4.2 Taux de convergence pour l'estimateur de la médiane conditionnelle

En plus d'imposer à la fonction de répartition conditionnelle d'être lipschitzienne, le comportement de l'estimateur $\hat{m}_n(x)$ dépend de la "platitude" de la fonction de répartition conditionnelle F^x autour de la médiane conditionnelle $m(x)$

et ceci peut être contrôlé par le nombre de dérivées de F^x qui s'annulent au point $m(x)$. Nous introduisons donc l'hypothèse H'_5 englobant les deux conditions suivantes.

$$\left\{ \begin{array}{l} H'_{5,1} : \exists \beta > 0 \exists C > 0 / \forall x', x'' \in V_x^2, \forall y', y'' \in V_y^2 \quad |F^{x'}(y') - F^{x''}(y'')| \leq C d^\beta(x', x'') + |y' - y''|^\beta \\ H'_{5,2} : \exists j \geq 2 \quad \forall 1 \leq l \leq j-1, \quad F^{x[l]}(m(x)) = 0 \text{ et } F^{x[j]}(m(x)) > 0 \\ \text{où } F^{x[l]} \text{ désigne la dérivée d'ordre } l \text{ de } F^x. \end{array} \right.$$

Puisque des estimateurs des dérivées $F^{x[l]}$ ($1 \leq l \leq j$) nous sont nécessaires, nous rajoutons l'hypothèse H''_5 englobant les deux conditions suivantes.

$$\left\{ \begin{array}{l} H''_{5,1} : \lim_{ng^{2j-1}\varphi_x(h)} \frac{\log n}{ng^{2j-1}\varphi_x(h)} = 0 \\ H''_{5,2} : H \text{ est } j \text{ fois continûment dérivable} \end{array} \right.$$

Nous pouvons maintenant établir le taux de convergence de \hat{F}_n^x .

Proposition 5 *Si les hypothèses $H_2 - H_3$, $H'_{5,1}$ et H_6 sont satisfaites, nous obtenons*

$$\hat{F}_n^x(y) - F^x(y) = O(h^\beta + g^\beta) + O_{a.c.o} \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right).$$

Chapitre 4

Etude sur des données réelles

La spectroscopie dans l'infrarouge est une technique analytique basée sur le principe d'absorption des rayonnements (infrarouge) par la matière organique en utilisant un spectromètre. Cette méthode d'absorption est utilisée pour estimer la composition chimique des aliments (fourrages, viandes, ...).

Sur la base d'observations antérieures, nous voulons estimer le taux de matière grasse d'un morceau de viande à partir du spectre observé.

On se donne un échantillon de 215 morceaux de viande, on observe pour chaque morceau i , la variable fonctionnelle $X_i(t), t \in [850, 1050]$ qui est la courbe spectrométrique du morceau de viande i , la répartition graphique des 215 spectres montre l'aspect fonctionnel des données (voir Figure 4.1). Nous nous proposons de comparer les estimateurs des fonctions de régression, de médiane conditionnelle et du mode conditionnel étudiés au chapitre précédent. A cette fin, nous utilisons les 161 premières observations pour calculer les estimateurs précédents et le reste des observations servira à les comparer.

Pour calculer ces estimateurs, nous utilisons :

- *Le noyau quadratique.*
- *La semi-métrique d est la semi metric basée sur les dérivées (voir routines "semimetric.deriv" dans le website <http://www.lsp.ups-tlse.fr/staph/npfda> dans Ferraty et Vieu (2006)) cette semi-métrique est plutôt employée pour le traitement de courbes lisses (ou régulières)*

Nous utilisons alors l'algorithme suivant :

- *Etape 1. Nous divisons nos données en deux sous ensembles : $(X_i, Y_i)_{i=1, \dots, 160}$ l'échantillon d'apprentissage utilisé pour construire les estimateurs*

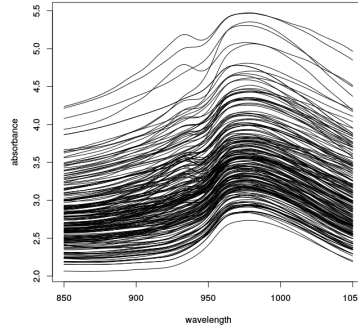


FIGURE 4.1 – Les données spectrometriques.

et $(X_i, Y_i)_{i=161, \dots, 215}$ l'échantillon de test utilisé pour évaluer la qualité des estimateurs.

- Etape 2. Nous calculons le h -optimal (paramètre de lissage) par la méthode de validation croisée.
- Etape 3. Pour chaque X_i dans l'échantillon de test, nous calculons sa valeur réponse Y_i par chaque méthode d'estimation et nous trouvons \hat{Y}_i .
- Etape 4. Pour illustrer la performance des estimateurs, nous représentons les valeurs estimées \hat{Y}_i en fonction des vraies valeurs Y_i de l'échantillon de test (voir Figure 4.2).
- Etape 5. Plus précisément, pour comparer les trois méthodes, nous calculons l'erreur en moyenne quadratique donnée par

$$MSE := \frac{1}{55} \sum_{i=1}^{55} \left(\hat{Y}_i - Y_i \right)^2,$$

où \hat{Y}_i désigne l'estimateur utilisé.

Dans la suite, nous expliquons en quoi consiste la méthode de validation croisée

Validation croisée

Il y a au moins trois techniques de validation croisée : “testset validation” ou “holdout method”, “ k -fold cross-validation” et “leave-one-out cross-

validation" (LOOCV).

- *Testset validation ou holdout méthode (HOCV) :*

1. On divise l'échantillon d'apprentissage en deux parties disjointes : la première est utilisée pour calculer d'estimateur et la deuxième pour le tester.
2. Le test se fait via la commande : $MSE = \text{mean}[(Y_{\text{estim}} - Y)^2]$,

- *k-fold cross-validation (k-FCV) :*

1. On découpe l'échantillon de départ X_1, \dots, X_n en k paquets C_1, \dots, C_k de même taille n/k .
2. Pour $i = 1, \dots, k$:

- * On construit l'estimateur $Y_{\text{estim}}(i)$ à partir de toutes les observations sauf le paquet C_i .

- * On calcule le MSE à partir des observations du paquet C_i qui sont indépendantes des précédentes.

$$MSE(i) = \text{mean}[Y_{\text{estim}}(i) - Y(i)]^2.$$

3. Les tests se font en même temps via la commande :

$$MSE = \text{mean}\{MSE(1) + MSE(2) + \dots + MSE(k)\}.$$

Par exemple pour $k = 2$ On divise l'échantillon d'apprentissage en deux blocs : on calcule d'estimateur en utilisant le premier bloc C_1 (on trouve $Y_{\text{estim}2}$), et on teste sur le deuxième bloc C_2 , ensuite on calcule l'estimateur en utilisant le deuxième bloc (on trouve $Y_{\text{estim}1}$) et on teste sur le premier. Les tests se font en même temps via la commande :

$$MSE = \text{mean}[(Y_{\text{estim}1} - Y_1)^2] + \text{mean}[(Y_{\text{estim}2} - Y_2)^2].$$

- *Leave-p-out cross validation (L-p-OCV) :*

Même principe que k -FCV appliqué à tous les paquets possibles de p variables parmi n , i.e. à chaque étape, on utilise $n - p$ observations pour construire d'estimateur et les p observations restantes permettent d'estimer son risque.
* Comme cas particulier : pour $p = 1$ (Leave-one-out cross validation (LOOCV)).

Cette méthode est aussi un cas particulier de la première méthode où $k = n$,

c'est à dire que l'on apprend sur $(n-1)$ observations puis on valide le modèle sur l'observation restante on répète cette opération n fois.

- *Nous concluons : Dans une validation croisée, on divise toujours l'échantillon soit en deux parties, ou bien en k blocs ou bien en enlevant un seul élément de l'échantillon.*

- *Différence entre ces méthodes :*

1. *Dans la méthode k -FCV, chaque variable appartient à un paquet et un seul. En particulier, chaque variable n'est utilisée qu'une seule fois pour estimer l'erreur.*

2. *Malgré la simplicité de HOCV, elle n'est pas utilisée par les praticiens car elle présente une trop grande variabilité qui vient du choix arbitraire du sous-ensemble. Pour l'améliorer, on peut considérer différentes partitions dans l'espoir d'obtenir une évaluation plus précise de la qualité de chaque méthode. Toute procédure de k -FVC peut se déduire du HOCV dès lors qu'on se donne une collection d'échantillons d'entraînement.*

Programmes

Programmes de Ferraty et Vieu (voir le site <http://www.lsp.ups-tlse.fr/staph/npfda>)

```
source("Routines.r")
spec <- read.table("npfda - spectrometric.txt")
n = 150 ; ntest = 65
X = spec[1 : n, 1 : 100]
Y = spec[1 : n, 101]
Xtest = spec[(n + 1) : (n + ntest), 1 : 100]
Ytest = spec[(n + 1) : (n + ntest), 101]
K = "quadratic"
xlimits <- -range(Ytest)
```

- *Pour la fonction de régression :*

```
Yestimes.reg <- -0
Yestimes.reg <- -funopare.knn.lcv(Y, X, Xtest, 2, nknot = 20, c(0, 1))
mse.reg = round(mean((Yestimes.reg$Predicted.values - Ytest) ** 2), 3)
```

```

print(mse.reg)
ylimits1 < -range(Yestimes.reg$Predicted.values)
plot(Ytest, Yestimes.reg$Predicted.values, xlim = xlimits, ylim = ylimits1, main =
paste("Régression : MSE = ",mse.reg,""), xlab = "Ytest", ylab =
"Yestimes.reg")
abline(0,1)

```

- Pour la médiane conditionnelle :

```

Yestimes.mediane < -0
Yestimes.mediane < -funopare.quantile.lcv(Y, X, Xtest, 2, nknot =
20, c(0,1), alpha = 0.5)
mse.mediane = round(mean((Yestimes.mediane$Predicted.values - Ytest)*
*2), 3)
print(mse.mediane)
ylimits2 < -range(Yestimes.mediane$Predicted.values)
plot(Ytest, Yestimes.mediane$Predicted.values, xlim = xlimits, ylim =
ylimits2, main = paste("Médiane.Cond : MSE = ",mse.mediane,""), xlab =
"Ytest", ylab = "Yestimes.mediane")
abline(0,1)

```

- Pour le mode conditionnel :

```

Yestimes.mode < -0
Yestimes.mode < -funopare.mode.lcv(Y, X, Xtest, 2, nknot = 20, c(0,1))
print(mean((Yestimes.mode$Predicted.values - Ytest) * *2))
ylimits3 < -range(Yestimes.mode$Predicted.values)
plot(Ytest, Yestimes.mode$Predicted.values, xlim = xlimits, ylim =
ylimits3, main = paste("Mode.Cond : MSE = ",mse.mode,""), xlab =
"Ytest", ylab = "Yestimes.mode")

```

- Pour tracer les trois graphes dans la même figure :

```

par(mfrow = c(1,3))
xlimits < -range(Ytest)
ylimits < -range(c(Yestimes.reg$Predicted.values, Yestimes.médiane$Predicted.values,
Yestimes.mode$Predicted.values) )
plot(Ytest, Yestimes.reg$Predicted.values, xlim = xlimits, ylim = ylimits1, main =
paste("Régression : MSE = ",mse.reg,""), xlab = "Ytest", ylab =
"Yestimes.reg")
abline(0,1)

```

```

plot(Ytest, Yestimes.mediane$Predicted.values, xlim = xlimits, ylim =
ylimits2, main = paste("Mediane.Cond : MSE = ", mse.mediane, ""), xlab =
"Ytest", ylab = "Yestimes.mediane")
abline(0,1)
plot(Ytest, Yestimes.mode$Predicted.values, xlim = xlimits, ylim =
ylimits3, main = paste("Mode.Cond : MSE = ", mse.mode, ""), xlab =
"Ytest", ylab = "Yestimes.mode")
abline(0,1)

```

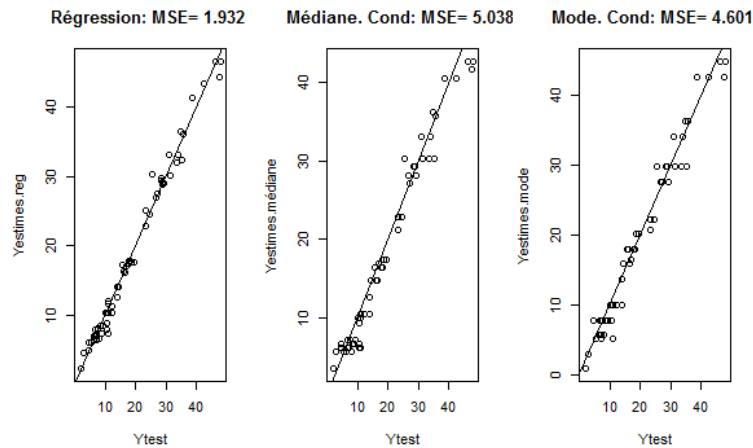


FIGURE 4.2 – Comparaison graphique des trois estimateurs

Les résultats obtenus sont :

- $MSE(reg)=1.932$,
- $MSE(med.cond)=5.038$,
- $MSE(mod.cond)=4.601$.

Sur la base de cet ensemble de données, nous voyons que la prédiction via la régression fournit la plus petite erreur. Ne peut on pas imputer cela au critère de comparaison utilisé (MSE) du moment que nous savons que la fonction de régression minimise l'erreur L^2 ?

Autres Programmes

- Pour la fonction de régression :

```
source("Routines.r")
estimateur.freg <- -function(X, Y, x, K, h, sm) {
  d <- -sm(X, x)
  W <- -K(d/h)
  calcul.estimateur(Y, W) }
calcul.estimateur <- -function(Y, W) {
  sum(W * Y)/sum(W)}
h.optimal <- -function(X, Y, K, semimetric){
  n = length(Y)
  I1 = 1 : (n/2)
  I2 = (n/2 + 1) : n
  ks = seq(ceiling(length(I1)/20), length(I1), ceiling(length(I1)/100))
  critere <- -function(I1, I2){
    sm = semimetric(X[I1,], X[I2,])
    ssm = apply(sm, 2, sort)
    crit = Vectorize(function(kh) {
      Yp = sapply(1 : length(I2), function(i) {
        h <- -ssm[kh, i]
        W <- -K(sm[, i]/h)
        F <- -calcul.estimateur(Y[I1], W) })
      mean((Yp - Y[I2]) * *2) })(ks)
    }
  crit = (critere(I1, I2) + critere(I2, I1))/2
  kopt = ks[which(crit == min(crit, na.rm = TRUE), arr.ind = TRUE)]
  print(hopt)
  hopt }
spec <- -read.table("npfda - spectrometric.dat")
n = 160; ntest = 55
X = spec[1 : n, 1 : 100]
Y = spec[1 : n, 101]
Xtest = spec[(n + 1) : (n + ntest), 1 : 100]
Ytest = spec[(n + 1) : (n + ntest), 101]
kernel <- -function(t) 2 * (1 - t) * (t <= 1 & t >= 0)
sm <- -function(X1, X2) semimetric.deriv(X1, X2, 2, 20, c(0, 1))
hopt = h.optimal(X, Y, kernel, sm)
```



```

print(hopt)
Reg = sapply(1 : length(Ytest), function(i){
  x = Xtest[i,]
  sm1 = sm(X, x)
  h = sort(sm1)[hopt]
  F = estimateur.freg(X, Y, x, kernel, h, sm1)})
print(Reg)
mse.Reg = mean((Reg - Ytest) ** 2)
print(mse.Reg)
xlimits <- -range(Ytest)
ylimits1 <- -range(Reg)
plot(Ytest, Reg, xlim = xlimits, ylim = ylimits1, main = paste("Régression :
MSE = ", mse.Reg, ""), xlab = "Ytest", ylab = "Yestimes.Reg")
abline(0, 1)

```

- Pour la médiane conditionnelle :

```

source("Routines.r")
estimateur.frcond <- function(X, Y, x, K, H, h, kg, sm){
  d <- -sm(X, x)
  W <- -K(d/h)
  calcul.estimateur(Y, W, H, kg) }
calcul.estimateur <- function(Y, W, H, kg){
  F <- function(y){
    g = sort(abs(y - Y))[kg]
    (H(outer(y, Y, '-'))/g)
  }
  Vectorize(F)} mediane <- function(F, m, M){
  if(any(is.na(c(F(m), F(M)))))
  return(NA)
  if(F(m) > 1/2 || F(M) < 1/2)
  return(NA)
  Res <- uniroot(function(x) F(x) - 1/2, c(m, M))$root}
hopt.hoptimal <- function(X, Y, K, H, m, M, semimetric){ n = length(Y)
  I1 = 1 : (n/2)
  I2 = (n/2 + 1) : n
  ks = seq(ceiling(length(I1)/20), length(I1), ceiling(length(I1)/100))
  critere <- function(I1, I2){
  sm = semimetric(X[I1,], X[I2,])

```

```

ssm = apply(sm, 2, sort)
crit = outer(ks, ks, Vectorize(function(kh, kg){
Yp = sapply(1 : length(I2), function(i){
h < -ssm[kh, i]
W < -poids(sm[, i], sm[, i], K, h)
F < -calcul.estimateur(Y[I1], W, H, kg)
mediane(F, m - (M - m), M + (M - m)))})
mean((Yp - Y[I2])**2)}))})
crit = (critere(I1, I2) + critere(I2, I1))/2
hopt = ks[which(crit == min(crit, na.rm=TRUE), arr.ind=TRUE)]
print(hopt)}
specdata$specj-read.table("npfda-spectrometric.dat")
n=160; ntest= 55
X = spec[1 :n,1 :100]
Y = spec[1 :n,101]
Xtest = spec[(n+1) :(n+ntest),1 :100]
Ytest = spec[(n+1) :(n+ntest),101]
kernel j- function(t) 2*(1-t) * (t j= 1 & t j= 0)
sm j- function(X1, X2) semimetric.deriv(X1, X2, 2, 20, c(0, 1))
m = min(Y); M = max(Y)
H j- function(t) (3/4*(t-t^3/3) + 1/2) * (t > -1 & t < 1) + (t >= 1)
sm < -function(X1, X2)semimetric.deriv(X1, X2, 2, 20, c(0, 1))
hopt = hopt.optimal(X, Y, kernel, H, m, M, sm)
kh = kopt[1]
kg = kopt[2]
Res = sapply(1 : length(Ytest), function(i){
x = Xtest[i,]
sm1 = sm(X, x)
h = sort(sm1)[kh]
F = estimateur.frcond(X, Y, x, kernel, H, h, kg, sm1)})
mediane(F, m - (M - m), M + (M - m)))
mse.Res = mean((Res - Ytest) **2)
print(mse.Res]
plot(Ytest, Res, xlim = xlimits, ylim = ylimits1, main = paste("Mediane.Cond :
MSE = ", mse.Res, ""), xlab = "Ytest", ylab = "Yestimes.Mediane.Cond")
abline(0, 1)

```

Les résultats obtenus sont :

- $MSE(reg)=1.886$,
- $MSE(med.cond)=4.04$.

Sur la base de cet ensemble de données, nous voyons que la prédiction basée sur la régression fournit encore l'erreur la plus petite.

EXERCICE 3.1

- 1) Donner des exemples de noyaux de type 0, de type 1 et de type 2.
- 2) Montrer que si K est un noyau de type 2 et si $K(1) \neq 0$ alors K est de type 1.
- 3) Montrer que si K est un noyau de type 1 (resp. type 2) alors pour tout $m \in \mathbb{N}^*$ $\frac{K^m}{\int K^m}$ est aussi un noyau de type 1 (resp. type 2) après avoir montré que $0 < \int K^m < \infty$.

EXERCICE 3.2

Prouver la validité de la relation (3.22).

EXERCICE 3.3

Soit $\alpha \in]0, 1[$, montrer que sous les hypothèses $H_2 - H_3$ et $H_5 - H_6$, nous avons

$$\lim_{n \rightarrow \infty} \widehat{t}_{\alpha, n}(x) = t_{\alpha}(x) \quad \text{p.c.o.}$$

EXERCICE 3.4

Soit H un espace de Hilbert séparable muni du produit scalaire $\langle \cdot, \cdot \rangle$ et soit $\{e_j, j = 1, \dots, \infty\}$ une base orthonormale. Soient $k \in \mathbb{N}^*$ un entier fixé et $x = \sum_{j=1}^{\infty} x^j e_j$ un élément fixé de H .

- 1) Montrer que l'application donnée par

$$\forall (x', x'') \in H \times H, \quad d_k(x', x'') = \sqrt{\sum_{j=1}^k \langle x' - x'', e_j \rangle^2}$$

est une semi métrique dans l'espace H , est elle une métrique ?

- 2) Soit $X = \sum_{j=1}^{\infty} X^j e_j$ une variable aléatoire (à valeurs dans H) de carré intégrable. Supposons que le vecteur aléatoire (X^1, \dots, X^k) admet une densité (par rapport à la mesure de Lebesgue sur \mathbb{R}^k) f continue au point (x^1, \dots, x^k) et telle que $f(x^1, \dots, x^k) > 0$. Montrer que X est un processus fractal d'ordre k par rapport à la semi métrique d_k (voir la définition ci dessous), en déduire que la fonction de concentration $\phi_x(\varepsilon)$ associée au processus X satisfait la condition 3.4.

Définition

Un processus X est dit fractal d'ordre τ , par rapport à une semi métrique d , s'il existe une constante finie $C > 0$ telle que la fonction de concentration

$\phi_x(\varepsilon)$ associée à X s'écrit

$$\phi_x(\varepsilon) \sim C\varepsilon^\tau \text{ quand } \varepsilon \rightarrow 0.$$

Bibliographie

- Devroye L. and Györfi L. (1983). *Distribution-free exponential bounds on the L_1 error of partitioning estimates of a regression function.* In : Konecny, F., Mogyordi, J., Wertz, W. (Eds.), *Proceedings of the Fourth Pannonian Symposium on Mathematical Statistics.* Akadémian Kiadó, Budapest, Hungary, pp. 67–76.
- Devroye L. and Krzyżak A. (1989). *An equivalence theorem for L_1 convergence of the kernel regression estimates.* *J.Statist. Plann. Inference.* **23**, 71–82.
- Devroye L., Györfi L., Krzyżak A. and Lugos, G. (1994). *On the strong universal consistency of nearest neighbor regression function estimates.* *Ann. Statist..* **22**, 1371–1385.
- Donoho D. L., Johnstone, I. M., Kerkyacharian G. and Picard D. (1995). *Wavelet shrinkage : asymptopia (with discussion) ?* *J. Roy. Statist. Soc., Ser. B* **57** (2), 301–370.
- Doukhan P. and Léon J. (1990). *Déviations quadratiques d’estimateurs de densité par projection orthogonale.* *Compt. Rend. Acad. Sci. Paris A* **310**, 424–430.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice.* Springer Series in Statistics. New York.
- Kerkyacharian G. and Picard D. (1992). *Density estimation in Besov Spaces.* *Statistics and Probability Letters* **13**, 15–24.
- Kohler M., Krzyżak A. (2001). *Nonparametric regression estimation using penalized least squares,* *IEEE Trans. Inform. Theory.* **47**, 3054–3058.

- Nadaraya, E. A. (1964). *On estimating regression. Theor. Probab. Appl.* , **9**, 157–159.
- Parzen E. (1962). *On estimating of a probability density and mode. Annals of Mathematicla Statistics*, **33**, 1065–1076.
- Prakasa Rao B. L. S. (1983). *Nonparametric Functional Estimation. Academic Press, New York.*
- Rosenblatt M. (1956). *Remarks on some nonparametric estimates of a density function (1956). Annals of Mathematical Statistics*, **27**, 832–837.
- Walter, G. G. (1992). *Approximation of the Delta Function by Wavelets. J. Approx. Theory.* **71**, 329–343.
- Watson, G. S. (1964). *Smooth Regression Analysis. Sankhya A* , **26**, 359–372.