



**UNIVERSITÉ KASDI MERBAH
OUARGLA**

**Faculté des mathématiques et sciences de la
matière**

N° d'ordre :
N° de série :

DÉPARTEMENT DE MATHÉMATIQUE

MASTER

Spécialité : Mathématiques

Option : Probabilité et Statistique

Par : Rahmani Asmaa

Thème

**Estimation non paramétrique de la régression par la
méthode du noyau**

Soutenu publiquement le : 01/07/2019

Devant le jury composé de :

Mrs.r.khorsi	M.A.A université de KASDI Merbah - Ouargla	Examineur
Mr.A.hocine	M.A.A université de KASDI Merbah - Ouargla	président
Mr.R.Agoune	M.A.A université de KASDI Merbah - Ouargla	Rapporteur

Dédication

Je dédie ce travail à :

A mes parents :M.laid,R.Djemila

A mes frères :Hamed,Marwan,Abd Elhai

et mes sœurs :Khawla,Djomana ;Hadil,Anfel

et tout ma famille

A mes chers amies Fida,Naima ,hadjer et tout amies

Remerciement

En premier lieu, je remercie Dieu, le miséricordieux, sans lui rien de tout cela n'aurait pu être.

je remercie particulièrement mon encadreur :**Mr r.agoune** pour son aide précieuse, sa patience et ses encouragements. Je voudrais également remercier : **Mr :A.hocine** pour nous avoir fait l'honneur de présider le jury de notre mémoire, je remercie vont également aux : **Mrs.R.KHORSI** pour examiner le mémoire . s'adressent également à tous ce qui m'ont aidé et permis de faire aboutir ce travail.

Table des matières

Dédication	i
Remerciement	ii
Notations et Préliminaires	v
1 Les outils probabilistes	2
1.1 Convergence des variable aléatoire	2
1.1.1 Convergence en loi	2
1.1.2 Convergence en probabilité :	3
1.1.3 Condition suffisante de convergence en probabilité	3
1.1.4 Convergence presque sûre :	4
1.1.5 Convergence presque complète :	4
1.2 Espérance Conditionnelle	5
1.2.1 Espérance	5
1.2.2 loi conditionnelle :	6
1.2.3 Espérance conditionnelle	6
1.3 Principe d'estimation non paramétrique	7
1.4 propriété de l'estimateur	9
1.5 Estimation par la méthode du noyau :	10
1.5.1 Estimation de Parzen-Rosenblatt :	10
1.5.2 Noyaux usuels :	12
1.6 Propriétés d'un estimateur à noyau :	13
1.6.1 Critères de convergence :	14

1.7	Choix du paramètre de lissage	16
1.7.1	introduction	16
1.8	Méthodes plug in (ré-injection) :	17
1.8.1	Estimateur optimal :	17
1.8.2	Règle de Thumb	18
1.8.3	La validation croisée :	18
2	Étude asymptotique	20
2.1	Hypothèses	20
2.2	Convergence presque complète ponctuelle	20
2.3	Convergence presque complète uniforme	22
3	Application et exploration des observations sur des données simulées	25
3.1	Simulée les données	25
3.1.1	Algorithme 1	25
3.1.2	Programme sous Matlab	26
3.2	Résultats de simulation	26

Notations et Préliminaires

- X : Variable aléatoire.
- $E(X)$: Espérance de la variable aléatoire X .
- $V(X)$: La variance de la variable aléatoire X .
- $p.co$: Convergence presque complète.
- $i.i.d$: Indépendantes et identiquement distribuées.
- MSE : Erreur quadratique moyenne (Mean square Error).
- $MISE$: Erreur quadratique moyenne intégrée (Mean Integrated square Error).
- CV : Cross Validation (validation croisée)
- $O_{p.co}$: vitesse de convergence presque complète
- h : le paramètre de lissage

- K : une fonction asymétrique du noyau
- r :opérateur de régression non linéaire

Introduction

La théorie de l'estimation est une des branches importantes de la statistique. Cette théorie est habituellement divisée en deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non paramétrique. L'estimation non paramétrique consiste, dans la majeure partie des cas, à estimer, à partir des observations, une fonction inconnue élémentaire 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 associées à la loi de l'échantillon

dans ce mémoire on a trois chapitres :

le premier chapitre. nous avons brièvement défini la méthode d'estimation non paramétrique. Nous nous sommes intéressés à la méthode du noyau, on a défini l'estimateur et donné ses différentes propriétés statistiques et asymptotiques. C'est l'objet du premier chapitre

Le deuxième chapitre est le plus théorique on parle de la convergence presque complète ponctuelle et la convergence presque complète uniforme on utilise les lemmes et l'hypothèse on utilise dans ce chapitre pour expliquer la théorie de l'estimation

Le troisième chapitre est le dernier on utilise l'information du premier chapitre et le deuxième pour extraire le programme dans Matlab pour trouver la simulation de la régression dans le cas d'indépendance des variables aléatoires.

Chapitre 1

Les outils probabilistes

1.1 Convergence des variable aléatoire

1.1.1 Convergence en loi

Définition 1.1 Soit (X_n) suite v.a.r. de fonction de répartition (F_n) et X une v.a.r de fonction de répartition. La suite (X_n) converge en loi vers X si, et seulement si :

$$\lim F_n(x) = F(x)$$

en tout point x où F est continue .

Théorème 1.2 (Théorème de Slutsky)

soit (X_n) et X des vecteurs aléatoires dans R^p , tel que (X_n) converge en loi vers X . Si g est une application continue de

R^p vers R^q , alors on a :

$$g(X_n) \longrightarrow g(X) = \frac{\sum(Y_i)K(\frac{X_i-x}{h})}{\sum(X_i-x)}$$

Théorème 1.3 (Théorème de Paul Lévy) :

Soit (X_n) est une suite de variables aléatoires dans R^p . Si la suite $\varphi(X_n)$ de ses fonctions caractéristiques converge simplement vers une fonction φ continue en 0, alors φ est la fonction caractéristique d'une variable aléatoire X et (X_n) converge en loi vers X , i.e :

$$\varphi(X_n(x)) \longrightarrow \varphi(x), \forall x \in R^p \Rightarrow X_n \rightarrow X$$

1.1.2 Convergence en probabilité :

Définition 1.4 Soit $(X_n)_n$ une suite de variables aléatoires définies sur un espace de probabilité (Ω, A, P) . On dit que (X_n) converge vers X en probabilité si :

$$\forall \varepsilon > 0, \lim(\mathbb{P}(|X_n - X| \geq \varepsilon)) = 0$$

on note : $X_n \xrightarrow{P} X$

1.1.3 Condition suffisante de convergence en probabilité

Proposition 1.5 Soit (X_n) une suite de v. a réel Si :

$$\lim EX_n = a, \text{ et, } \lim \text{Var} X_n = 0$$

alors :

$$X_n \rightarrow a$$

Preuve. par l'inégalité de Bienaymé Tchebychev on a pour tout $\xi > 0$:

$$P(|X_n - a| > \xi) \leq \frac{E(X_n - a)^2}{\xi^2}$$

alors :

$$E(X_n - a)^2 = \text{var}(X_n) + (EX_n - a)^2$$

donc :

$$\forall \xi > 0, P(|X_n - a| > \xi) \leq \frac{\text{Var} X_n + E(X_n - a)^2}{\xi^2}$$

$$\forall \xi > 0 \lim_{n \rightarrow \infty} P(|X_n - a| > \xi) = 0$$

donc $X_n \rightarrow a$

■

propriété :

soient (X_n) et (Y_n) deux suites de variables aléatoires . Si on a :

$$X_n \rightarrow X, \text{ et } Y_n \rightarrow Y \Rightarrow X_n + Y_n$$

1.1.4 Convergence presque sûre :

Définition 1.6 1 - On dit que (X_n) converge presque sûrement vers X si :

$$\mathbb{P}(\lim(X_n = X)) = 1$$

Ou de manière équivalente, s'il existe un sous-ensemble \mathbb{P} -négligeable N dans Ω tel que :

$$\forall \omega \in \Omega/N : (X_n)_\omega \xrightarrow{p.s} X_\omega$$

et on écrit :

$$X_n \xrightarrow{p.s} X$$

Théorème 1.7 La suite de v.a.r. (X_n) converge presque sûrement vers X si la suite de v.a.r. Y_m définie par :

$$Y_m = \sup_{n \geq m} |X_n - X|$$

1.1.5 Convergence presque complète :

Dans la première définition, nous nous référons à une définition convergence presque complète qui a été découvert par Hsu et Robbins (1947) , Convergence presque complète a été utilisé dans la calcul des sommes de variables aléatoires Et dans le domaine des statistiques , surtout en statistique non-paramétrique . Et dans la deuxième définition, nous nous référons la définition de vitesse convergence presque complète qui a été découvert par Ferraty et Vieu (2006). Elle a l'avantage théorique d'impliquer les deux vitesses de convergence classiques en probabilité et presque sûre, et l'avantage pratique d'être souvent plus facile à démontrer. ce mode de convergence a été très utilisé dans des travaux concernant la statistique non-paramétrique .

Définition 1.8 1 - On dit que la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ converge presque complètement vers une variable aléatoire X lorsque $n \rightarrow \infty$ si :

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

2 - on dit que la vitesse de convergence presque complète de la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ vers X est d'ordre (u_n) si :

$$\exists \varepsilon_0 > 0, \sum_{n \in \mathbb{N}} \mathbb{P}[|X_n - X| > \varepsilon_0 u_n] < \infty$$

1.2 Espérance Conditionnelle

1.2.1 Espérance

définition :

Soit X une variable aléatoire discrète et soit f_X sa fonction de masse. On dit que X admet une espérance si :

$$\sum_{x \in X(\Omega)} |x| f_X(x) < \infty$$

Dans ce cas on définit l'espérance de X par :

$$E(X) = \sum_{x \in X(\Omega)} x f_X(x) < \infty$$

Soit X une variable aléatoire avec densité f_X . On dit que X admet une espérance si :

$$\int_{\mathbb{R}} |x| f_X(x) dx < \infty$$

Dans ce cas on définit l'espérance de X par :

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx < \infty$$

Théorème 1.9 soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire et $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction mesurable. Alors on a :

1 - pour un vecteur aléatoire discret :

$$E(\varphi(X)) = \sum \varphi(x) f_X(X)$$

2 - pour un vecteur aléatoire à densité :

$$E(\varphi(X)) = \int_{\mathbb{R}^n} \varphi(x) f_X(X) dx_1 \dots dx_n$$

1.2.2 loi conditionnelle :

Définition 1.10 Soient X , Y deux variables aléatoires discrète La fonction de masse conditionnelle de Y sachant que $X = x$ est la fonction : $f_{Y/x}(Y/x) : R \rightarrow [0, 1]$ ddéfini par :

$$f_{Y/x}(Y/x) = P(Y = y/X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

pour tout x tel que $f_X(x) > 0$. La loi correspondante s'appelle la loi conditionnelle de Y sachant que $X = x$

Définition 1.11 Soient X , Y deux variables aléatoires conjointe $f_{(X,Y)}$ La densité conditionnelle de Y sachant que $X = x$ est la fonction : $f_{Y/x}(Y/x) : R \rightarrow [0, 1]$ ddéfini par :

$$f_{Y/x}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

1.2.3 Espérance conditionnelle

Définition 1.12 - Soient X , Y deux variables aléatoires On appelle espérance conditionnelle de Y sachant X la variable aléatoire si :

$$E(Y/X = x) = \sum_{y \in Y, (\Omega)} y f_{Y/X}(y/x)$$

- Soient X , Y deux variables aléatoires conjointe On appelle espérance conditionnelle de Y sachant la variable aléatoire X si :

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

Proposition 1.13 L'espérance conditionnelle $E(Y/X)$ satisfait

$$E(E(Y/X = x)) = E(Y)$$

généralement, pour toute fonction mesurable φ telle que les espérances existent :

$$E(E(Y/X)\varphi(X)) = E(Y\varphi(X))$$

Preuve.

$$\begin{aligned} E(E(Y/X)\varphi(X)) &= \sum_{x,y} y f_{Y/X}(y/x)\varphi(x)f_X(x) \\ &= \sum_{x,y} f_{X,Y}(x,y) = E(Y\varphi(X)) \end{aligned}$$

Dans le cas a densité, la preuve est formellement identique :

$$\begin{aligned} E(E(Y/X)\varphi(X)) &= \int_R \int_R y f_{Y/X}(y/x) dy \varphi(x) f_X(x) dx \\ &= \int_R \int_R y \varphi(x) f_{X,Y}(x,y) dx dy \end{aligned}$$

■

1.3 Principe d'estimation non paramétrique

Lorsque nous souhaitons décrire l'influence d'une variable quantitative sur un événement en faisant le moins d'hypothèses possible sur la forme de la relation, nous distinguons deux approches : L'approche de la régression paramétrique, L'approche de la régression non paramétrique. Le but d'un modèle de régression consiste a déterminer la façon dont l'espérance d'une variable dépendante Y dépend d'un ensemble de variables explicatives X. Supposons que $x \in \mathbb{R}$ le problème consiste donc a déterminer pour chaque réalisation x de la variable X; la valeur de la fonction $m(x)$, dite fonction de lien ou fonction de régression

Définition 1.14 *On appelle fonction de régression, la fonction $m(x)$ qui pour toute réalisation x de la variable explicative X associe la quantité :*

$$\mathbb{E}(Y | X = x) = m(x)$$

1-Modèle de régression paramétrique :

Pour caractériser cette fonction de régression, la première approche consiste a utiliser un modèle de régression paramétrique. Nous supposons que cette fonction peut s'écrire comme une explicite des valeurs de X. Cette fonction peut être linéaire, logarithmique, non-linéaire etc. Par exemple, dans le cas linéaire on suppose que :

$$\mathbb{E}(Y | X = x) = \alpha + \beta x$$

Nous cherchons alors à déterminer les meilleures valeurs de α et β

compte tenu d'un critère, par exemple celui de la MSE (Mean Square Error).

Remarque 1.15 Dans un modèle de régression paramétrique, la fonction est :

-De forme explicite.

-Peut s'écrire en fonction d'un nombre réduit de paramètres.

exemple :

$$\mathbb{E}(Y | X = x) = m(x, \theta) | Y = m(x, \theta) + \epsilon$$

ou $m(\cdot)$ est connue avec $\theta \in \mathbb{R}^k$

L'exemple typique est celui d'un modèle linéaire, ou l'on suppose que :

$$\mathbb{E}(Y | X = x) = \alpha x + \beta = m(\alpha, \beta, x)$$

1-Modèle de régression non paramétrique :

Nous pouvons retenir une approche non paramétrique dans laquelle on va estimer la relation entre le niveau moyen de Y et toutes les valeurs réalisées de X. Nous ne supposons aucune forme spécifique sur la fonction de régression.

Définition 1.16 Dans un modèle non paramétrique, la fonction de régression

1-N'a pas de forme explicite.

2-Ne peut pas s'écrire en fonction d'un nombre réduit de paramètres

$$\mathbb{E}(Y | X = x) = m(x) | y = m(x) + \epsilon$$

Remarque 1.17 Avec une approche non paramétrique, on aboutit à :

-une représentation graphique de la relation entre X et Y .

-Il n'existe pas de forme analytique de la fonction de lien $m(x)$.

Tout le problème consiste alors à estimer cette fonction de régression $m(x)$.

Le modèle qui va nous intéresser est un modèle non paramétrique, en ce sens que la seule condition que nous ferons sur la fonction m est une condition de régularité

$$\mathbb{E}(Y | X = x) = m(x) \tag{1.1}$$

k étant un entier positif ou nul (le cas $k = 0$ correspondant évidemment à l'hypothèse de continuité de m).

Nous allons donner quelques définitions théoriques des notions utilisées

1.4 propriété de l'estimateur

On dit qu'un estimateur \hat{m} de m est sans biais si :

$$\mathbb{E}(\hat{m}) = m$$

Définition 1.18 On dit qu'un estimateur \hat{m} de m est asymptotiquement sans biais si :

$$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{m}) = m$$

Définition 1.19 Un estimateur \hat{m} de m est dit asymptotiquement uniformément sans biais si :

$$\lim_{n \rightarrow \infty} \sup |[\mathbb{E}(\hat{m}(x) - m(x))]| = 0$$

Définition 1.20 L'erreur moyenne quadratique MSE :

$$\begin{aligned} MSE(m(x), \hat{m}(x)) &= \mathbb{E}(m(x), \hat{m}(x))^2 \\ &= \mathbb{E}(m(x))^2 + \mathbb{E}(\hat{m}(x))^2 - 2\mathbb{E}[m(x)\hat{m}(x)] + [\mathbb{E}\hat{m}(x)]^2 - [\mathbb{E}\hat{m}(x)]^2 \end{aligned}$$

Définition 1.21 L'erreur moyenne quadratique intégrée MISE :

$$\begin{aligned} MISE(\hat{m}, m) &= \int MSE(m(x), \hat{m}(x)) dx \\ &= \int \text{Biais}(\hat{m})^2 dx + \int \text{Var}(\hat{m}(x)) dx \end{aligned}$$

Définition 1.22 On dit qu'un estimateur \hat{m} de m est ponctuellement consistant en moyenne quadratique si :

$$\lim_{n \rightarrow \infty} MSE(m(x), \hat{m}(x)) = 0$$

Définition 1.23 On dit qu'un estimateur \hat{m} de m est uniformément consistant en moyenne quadratique intégrée si :

$$\lim_{n \rightarrow \infty} MISE(m(x), \hat{m}(x)) = 0$$

Définition 1.24 *On dit qu'un estimateur \hat{m} de m est asymptotiquement normal si*

$$\hat{m} \rightarrow \mathcal{N}(\mathbb{E}(\hat{m}), \text{Var}(\hat{m})), \text{ en loi}$$

1.5 Estimation par la méthode du noyau :

Le premier document publié qui a traité explicitement l'estimation de densité de probabilité est dû à Rosenblatt en 1956 qui a proposé l'estimateur naïf, suivi de Parzen en 1962 qui définit une classe de fonctions K , appelées noyau pour désigner la fonction de densité que l'on utilise dans les méthodes non paramétriques, remplaçons du même coup le terme 'fonction- poids' (weight function) qui était généralement utilisé par Rosenblatt.

L'estimateur précédemment construit peut encore être amélioré. En effet, maintenant que la classe est centrée en x , on peut tout de même remarquer que pour l'estimateur histogramme mobile, toutes les observations de cette classe ont le même rôle dans le calcul de f_h .

Il serait plus judicieux de penser que plus une observation est proche de x , plus elle doit intervenir dans le calcul de $f_h(x)$. L'idée la plus naturelle alors est de pondérer les observations en mettant d'autant plus de poids qu'on se trouve proche de x , et d'autant moins qu'on s'en trouve éloigné.

On choisira donc des fonctions de poids (noyaux) dans une classe plus large de densités, comprenant notamment des densités à support non borné, et ayant un seul mode à l'origine comme par exemple la loi normale centrée réduite.

1.5.1 Estimation de Parzen-Rosenblatt :

Soit (X_n) une suite de variables aléatoires réelles indépendantes et de même loi, de densité de probabilité f . On veut estimer f à partir d'un échantillon (X_1, X_2, \dots, X_n) . Soit h le

paramètre de lissage tel que :

$$h(n) \rightarrow 0 \text{ et } nh(n) \rightarrow \infty \text{ quand } n \rightarrow \infty$$

On peut estimer f par l'estimateur de Parzen-Rosenblatt

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (1.2)$$

Généralement la fonction K appelée noyau est une fonction positive et bornée tel que :

K est symétrique ,i.e, $K(u) = K(-u)$

$$\int_{\mathbb{R}} K(u) du = 1$$

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

$$\int_{\mathbb{R}} u^2K(u) du = \mu_2 < \infty$$

Cet estimateur a été largement étudié par de nombreux auteurs, citons par exemple Roussas (2000, 2001), Bosq et al.(1999) . Une première justification concernant la forme de l'estimateur de Parzen Rosenblatt, a été donnée précédemment.

Définition 1.25 *Estimateur à noyau :*

Un estimateur à noyau de la densité f est une fonction défini par :

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

où h est un paramètre appelé paramètre de lissage il dépend de n et il vérifie

$$h(n) \rightarrow 0 \text{ lorsque } n \rightarrow \infty$$

et K est une densité de probabilité appelée noyau.

Noyau

Un noyau K est une fonction intégrable défini de $\mathbb{R} \rightarrow \mathbb{R}$ tel que $\int_{\mathbb{R}} K(x) dx = 1$

.

Remarque 1.26 - *Le noyau K détermine la forme de voisinage autour du point x et h contrôle la taille de ce voisinage*

- f_h possède les mêmes propriétés de continuité et de différentiabilité que le noyau K

1.5.2 Noyaux usuels :

Noyau Uniforme (Rosenblatt) :

$$K(x) = \begin{cases} \frac{1}{2}, & \text{Si } |x| \leq 1; \\ 0, & \text{sinon} \end{cases}$$

Noyau triangulaire

Ce noyau s'écrit sous la forme :

$$K(x) = \begin{cases} 1 - |x|, & \text{Si } |x| \leq 1; \\ 0, & \text{sinon} \end{cases} \quad (1.3)$$

Noyau gaussien :

Ce noyau s'écrit sous la forme :

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), x \in \mathbb{R}$$

Noyau quadratique (Biweight) :

Le noyau quadratique est très intéressant car il donne un estimateur dérivable . En fait, il s'agit du noyau le plus simple parmi les noyaux de forme polynômial dérivable. Ainsi, il assure le lissage locale de la fonction f_h . Ce noyau est d'une forme très proche du noyau gaussien, il est donc préférable de l'utiliser. il s'écrit sous la forme :

$$K(x) = \begin{cases} \frac{15}{16} (1 - x^2), & \text{Si } |x| \leq 1; \\ 0, & \text{sinon} \end{cases}$$

Noyau d'Epanechnikov ou parabolique :

En 1969, Epanechnikov , a donné la forme du noyau K_e défini par :

$$K(x) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{x^2}{5}\right), & \text{Si } |x| \leq \sqrt{5}; \\ 0, & \text{sinon} \end{cases}$$

Noyau sinus :

$$K(x) = \frac{1}{2\pi} \left(\frac{\sin\left(\frac{x}{2}\right)}{\frac{x}{2}} \right)^2, \text{ si } x \neq 0$$

Noyau cosinus :

$$K(x) = \begin{cases} \frac{\pi}{4} \cos\left(\frac{\pi x}{2}\right), & \text{si } |x| \leq 1 \\ 0, & \text{sinon} \end{cases}$$

Noyau de Silverman :

$$K(x) = \frac{1}{2} \exp\left(\frac{-|x|}{\sqrt{2}}\right) \sin\left(\frac{|x|}{\sqrt{2}} + \frac{\pi}{4}\right), x \in \mathbb{R}$$

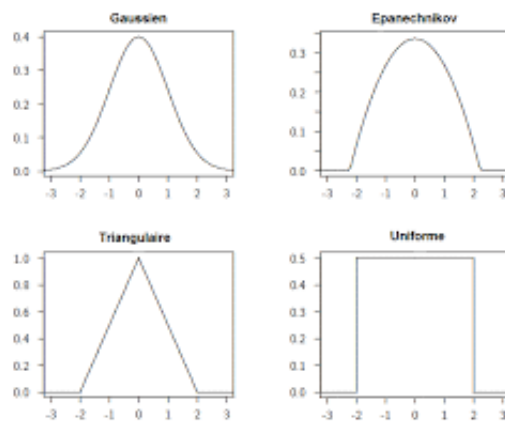


figure1 : quelques Formes des noyaux symétrique

1.6 Propriétés d'un estimateur à noyau :

Nous présentons dans cette partie les propriétés statistiques de l'estimateur de la densité .

Lemme :

Si le noyau K est positif et $\int K(u) du = 1$, alors l'estimateur de f_h est une densité de probabilité. De plus, f_h est continue si K est continue.

Démonstration :

L'estimateur à noyau est positif et continu car la somme des fonctions positives et continues est elle-même une fonction positive et continue. Il faut donc vérifier que l'intégrale de f_h vaut 1. En effet

$$\begin{aligned} \int_{-\infty}^{+\infty} f_h(x) &= \int_{-\infty}^{+\infty} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{+\infty} K\left(\frac{x - X_i}{h}\right) \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u) h du \text{ (changement de variable } u = \frac{x - X_i}{h} \text{)} \\ &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u) du \\ &= 1 \end{aligned}$$

1.6.1 Critères de convergence :

Parmi toutes les qualités que peut avoir un estimateur, on s'intéresse souvent à sa consistance, c'est à dire, au fait qu'un estimateur f_h converge ou non vers f . La convergence d'un estimateur peut être faible (en probabilité) ou forte (presque sûrement ou en moyenne quadratique).

On donne quelques résultats de convergence des estimateurs à noyaux de la littérature.

Théorème : (Parzen)

Soit f une densité continue et f_h son estimateur. Si le noyau K vérifie :

$$-\int K(u) du = 1 \text{ et } \int |K(u)| du < \infty$$

$$-\sup |K(u)| < \infty \text{ et } \lim_{u \rightarrow \infty} |uK(u)| = 0$$

Si le paramètre h satisfait :

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} nh = \infty$$

Alors f_h est un estimateur convergent en moyenne quadratique c'est à dire :

$$MSE(f_h(x)) \xrightarrow[n \rightarrow \infty]{} 0.$$

Convergence en moyenne quadratique intégrée :

Théorème :(Parzen)

Soit f une densité de puissance p^{eme} -intégrable et f_h son estimateur. Si le noyau K vérifie :

$$\int K(u) du = 1 \text{ et } \int |K(u)| du < \infty$$

$$\sup K(u) < \infty \text{ et } \lim_{u \rightarrow \infty} |uK(u)| = 0$$

Si le paramètre h satisfait :

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} nh = \infty$$

Alors f_h est un estimateur convergent en moyenne quadratique intégrée c'est à dire :

$$MISE(f_h(x)) \xrightarrow[n \rightarrow \infty]{} 0$$

Convergence uniforme en probabilité

Théorème(Parzen)

. Soit f la densité à estimer et f_h son estimateur, si les conditions suivantes sont vérifiées :

$$-\int K(u) du = 1 \text{ et } \int |K(u) du| < \infty$$

$$-\sup K(u) < \infty \text{ et } \lim_{u \rightarrow \infty} |uK(u)| = 0$$

$$-\lim_{u \rightarrow \infty} nh^2 = \infty$$

Convergence uniforme presque complète :

Théorème :(Nadaraya)

Soit f une densité uniformément continue et son estimateur à noyau K positif et à bornées.

si $\lim_{u \rightarrow \infty} h = 0$ et $\sum_{i=1}^{\infty} \exp(-\varepsilon nh^2) < \infty, \forall \varepsilon > 0$ alors f_h convergent uniformément avec la probabilité 1 c'est à dire :

$$\sup_{x \in \mathbb{R}} |f_h - f(x)| \rightarrow 0$$

Théorème :(Silverman)

Soit f une densité uniformément continue et f_h son estimateur à noyau K positif et à bornées.

$$\text{si } \lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{\log n}{nh} = 0$$

$$\sup_{x \in \mathbb{R}} |f_h(x) - f(x)| \xrightarrow{p.s} 0$$

Convergence en loi

Ce dernier résultat est tiré des travaux de Parzen. Il montre que l'estimateur à noyau est asymptotiquement normale.

Théorème :(Parzen)

Soit f une densité continue et f_h son estimateur si le noyau K vérifie

$$-\int K(u) du = 1 \text{ et } \int |K(u) du| < \infty$$

$$-\sup K(u) < \infty \text{ et } \lim_{u \rightarrow \infty} |uK(u)| = 0$$

$$-\lim_{h \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} nh = \infty$$

alors :

$$\frac{f_h(x) - E\{f_h(x)\}}{[V\{f_h(x)\}]^2} \xrightarrow{L} N(0, 1)$$

où $\xrightarrow{L} N$ désigne la convergence en loi.

1.7 Choix du paramètre de lissage

1.7.1 introduction

L'estimation de la fonction de probabilité par la méthode des noyaux est principalement conditionnée par le paramètre de lissage h , il est un facteur important dans l'estimation par la méthode des noyaux. Il représente en quelque sorte une fenêtre qui permet de déterminer le degré de lissage de l'estimation d'une fonction de densité. Un faible paramètre de lissage implique un faible degré de lissage et en une fonction de densité irrégulière. À l'opposé, une large valeur de h conduit à une estimation lisse.

Il ne faut pas oublier que ce choix dépend du but pour lequel l'estimation de la densité est utilisée. Plusieurs méthodes pour choisir ce paramètre ont été développées dans la littérature et quelques études comparatives ont été effectuées pour ces méthodes. Deux études comparatives intéressantes ont été publiées. La première est celle de Berlinet et Devroye (1994) et la deuxième est celle de Cao et al. (1994) . Elles comparent plusieurs méthodes pour choisir le paramètre de lissage pour plusieurs distributions différentes. Toutes

ces méthodes nous donnent un paramètre de lissage qui est optimale pour la distribution à estimer. Celles-ci diffèrent au niveau du choix du critère à optimiser. Dans ce chapitre nous allons parler de quelques méthodes permettant le calcul de ce paramètre.

1.8 Méthodes plug in (ré-injection) :

1.8.1 Estimateur optimal :

Le décision d'un choix optimal pour le paramètre de lissage suppose la spécification d'un critère d'erreur qui puisse être optimisé. Bien sûr, l'optimalité n'est pas un concept absolu : elle est liée aux choix du critère qui peut faire intervenir à la fois la densité inconnue f et l'estimateur f_h (donc h et le noyau K). Dans ce cas, on cherche à minimiser l'Erreur Quadratique Intégrée Moyenne (MISE)

$$MISE(f_h) = \int \mathbb{E} [f_h(x) - f(x)]^2 dx$$

Le MISE est défini par :

$$MISE = \frac{h^2}{4} \mu_2^2(K) R(f''(x)) + \frac{R(K)}{nh} + O\left(h^4 + \frac{1}{nh}\right) \quad (1.4)$$

et l'Erreur Quadratique Intégrée Moyenne Asymptotique $AMISE = MISE - O\left(h^4 + \frac{1}{nh}\right)$ est :

$$AMISE = \frac{h^2}{4} \mu_2^2(K) R(f''(x)) + \frac{R(K)}{nh} \quad (1.5)$$

avec $R(g) = \int g^2(x) dx$ pour toute fonction g

On peut remarquer que le premier terme du membre de droite du développement (1.4) est un terme de biais, alors que le second est un terme de variance. On constate que dans le MISE, le terme de biais est une fonction croissante en h alors que le terme de la variance est une fonction décroissante en h c'est à dire les deux termes varient en sens inverse par rapport à h , une largeur de fenêtre h trop importante entraînera une augmentation du biais et une diminution de la variance (phénomène de sur lissage), alors qu'une largeur de fenêtre trop petite provoquera une augmentation de la variance et une diminution du biais (phénomène de sous-lissage).

De l'expression (1.5), on peut déterminer le paramètre de lissage h^* qui minimise l'Erreur

Quadratique Intégrée Moyenne Asymptotique :

$$h^* = \left[\frac{R(K)}{\mu_2^2(K) R f''} \right]^{\frac{1}{5}} n^{-\frac{1}{5}} \quad (1.6)$$

h^* peut s'écrire :

$$h^* = \psi(K) \varphi(f) n^{-\frac{1}{5}} \quad (1.7)$$

ou $\psi(K) = \left[\frac{R(K)}{\mu_2^2(K)} \right]^{\frac{1}{5}}$ et $\psi(f) = \left[\frac{1}{R(f'')} \right]^{\frac{1}{5}}$ avec $R(f'') \neq 0$

Notons que h^* est une quantité déterministe qui dépend du nombre d'observations n . La valeur du AMISE optimale est donnée par :

$$AMISE^* = \frac{5}{4} [\mu_2^2(K) R^4(K) R(f'')]^{\frac{1}{5}} n^{-\frac{4}{5}} \quad (1.8)$$

le paramètre h^* optimal au sens du critère de l'erreur quadratique intégrée moyenne asymptotique devra réaliser un compromis entre les valeurs de la variance et celle du biais.

aussi sa nature asymptotique, la largeur de fenêtre optimale h^* dépend de la densité inconnue f à travers $R(f'')$. Cette largeur de fenêtre "idéale" (relativement au critère d'erreur retenu) n'est donc pas directement calculable. Une façon classique de remédier à ce dernier problème consiste à remplacer la quantité $R(f'')$ par un estimateur approprié. Il propose d'utiliser un paramètre de lissage h_1 pour calculer $f_{h_1}(x)$ et estimer $R(f'')$ pour calculer h^* .

1.8.2 Règle de Thumb

On a montré que si on choisit le paramètre de lissage de telle sorte que le MISE soit minimum alors le paramètre de lissage optimal h^* est donnée par la formule (1.6). Dans cette formule il suffit d'assigner une valeur au terme $R(f'')$ (pour obtenir une estimation de h^*). Si on choisit f comme étant la distribution normale de moyenne 0 et de variance σ^2 on a alors

$$R(f'') = \int (f''(x))^2 dx = \frac{3}{8} \pi^{-\frac{1}{2}} \sigma^{-5} \quad (1.9)$$

1.8.3 La validation croisée :

Définition 1.27 La fonction validation croisée est définie par la quantité :

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{m}^{-i}(x_{i,h}))^2$$

La seule différence avec le critère précédent réside dans l'utilisation de l'indice m^{-i} . Cet indice signifie que pour chaque $i = 1, \dots, n$; la valeur de $m(x_i)$ est obtenue en enlevant de l'observation x_i . Le modèle est estimé sur toutes les autres observations $x_j, j \neq i$ puis on estime la valeur de $m(\cdot)$ au point x_i à partir de cette régression. C'est la valeur estimée qui figure dans la formule CV (h) sous la notation $m^{-i}(x, h)$

$$\hat{m}^{-i}(x) = \frac{\sum_{j \neq i} y_j K\left(\frac{x - X_j}{h}\right)}{\sum_{j \neq i} K\left(\frac{x - X_j}{h}\right)}$$

La procédure de validation croisée peut s'interpréter comme étant le meilleur choix de h qui fait de $\hat{m}^{-i}(x_i)$ un estimateur efficace de Y_i

Proposition 1.28 (Hurlin) *Soit h_{CV} la valeur de h telle que :*

$$h_{CV} = \arg \min_{\{h \in \mathbb{R}^{*+}\}} CV(h)$$

Alors

$$MISE(h_{CV}) \rightarrow MISE(h_{opt}) \text{ en probabilit quand } n \rightarrow \infty$$

L'utilisation de la fonction CV permet ainsi d'obtenir un estimateur du paramètre optimal h_{opt}

Ferraty et Vieu [22] ont donné une autre forme pour la fonction validation croisée .

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{m}^{-i}(x_i, h))^2 w(x_i)$$

Théorème 1.29 Härdle et Marron

si H ne contient que des largeurs de fenêtres vérifiant

$$h = Cn^{-\frac{1}{5}}, 0 < C < \infty$$

alors on a la propriété suivante

$$\frac{\inf_{h \in H} MISE(h)}{MISE(h_{CV})} \rightarrow 1, \text{ presque sûrement :}$$

Chapitre 2

Étude asymptotique

dans ce paragraphe et dans un premier temps quelques notations et hypothèses qui paraissent importantes pour la suite de ce travail .

2.1 Hypothèses

h_1 : Les fonctions de répartition des variables Y , R et L sont continues,

$I_Y \leq I_L \leq I_R$ et $T_R < T_Y$

$h_{2,1}$: r et f sont l fois continument dérivables dans un voisinage de x .

$h_{2,2}$: $f(x) > 0$,

$h_{2,3}$: $\lim_{n \rightarrow \infty} h_n = 0$, $\frac{nh_n}{\log n} = +\infty$

$h_{2,4}$: K est bornée, intégrable, à support compact et $\int K(t) dt = 1$

$h_{2,5}$: $\int t^j K(t) dt = 0, \forall j = 1, \dots, l-1$ et $0 < \int t^l K(t) dt < \infty$

Soit S un sous-ensemble compact de \mathbb{R}

$h'_{2,1}$: r et f sont l fois continûment dérivables autour de S

$h'_{2,2}$: $\exists \theta > 0, \inf_{x \in S} f(x) > 0$

$h'_{2,6}$: $\exists \beta > 0, C < \infty, \forall x \in S, \forall y \in S : |K(x) - K(y)| \leq C |x - y|^\beta$

$H_{1,2}$: (R, L) et (X, Y) sont indépendants.

$H_{1,3}$: $\exists T < T_R$ ET $I > I_L$ tel que $\forall n \in \mathbb{N}, \forall i (1 \leq i \leq n), A_i = 0 \Rightarrow I \leq Z_i \leq T_{P.S.}$

2.2 Convergence presque complète ponctuelle

Le théorème suivant donne vitesse de convergence presque complète ponctuelle de r_n

Théorème 2.1 *Sous les hypothèses $h_1, h_{2,1} - h_{2,5}, H_{1,2}$ et $H_{1,3}$, nous avons*

$$r_n(x) - r(x) = O(h_n^\ell) + O_{a.co} \left(\sqrt{\frac{\log n}{nh_n}} \right)$$

Démonstration. Posons

est l'estimateur à noyau de f , puis effectuons la décomposition suivante

$$r_n(x) - r(x) = r_n(x) - \hat{r}_n + \frac{\hat{r}_n, N - g(x)}{f_n(x)} + \frac{f(x) - f_n(x)}{f_n(x)} r(x) \quad (2.1)$$

où $g = rf$.

La démonstration du théorème est alors une conséquence directe des lemmes suivants.

Lemme 1 *Sous les hypothèses $h_1, h_{2,3}$, et $H_{1,3}$, nous avons*

$$r_n(x) - \hat{r}_n(x) = O_{p.co} \left(\sqrt{\frac{\log n}{nh_n}} \right)$$

Démonstration : Les définitions de $r_n(x)$ de $\hat{r}_n(x)$ et l'hypothèse $H_{1,3}$ nous permettent d'écrire

$$r_n(x) - \hat{r}_n(x) = O_{p.co} \left(\sqrt{\frac{\log n}{n}} \right)$$

De plus, puisque $\lim_{n \rightarrow \infty} = 0$ hypothèse $h_{2,3}$, nous déduisons que

$$r_n(x) - \hat{r}_n(x) = O_{p.co} \left(\sqrt{\frac{\log n}{nh_n}} \right)$$

Lemme 2 *Sous les hypothèses $h_{2,1}, h_{2,4}, h_{2,5}$ et $H_{1,2}$, nous obtenons*

$$E\hat{r}_{n,N}(x) - g(x) = O(h_n^\ell)$$

sous les condition de régularité

$$\lim_{n \rightarrow \infty} \hat{r}(x) = r(x)$$

Quelque indication pour la preuve qui été détaillé en ferraty et view

Posant pour $i = 1, \dots, n$

$$\Delta_i = K(d \frac{(X_i, x)}{h}) / EK(d \frac{X_i - x}{h})$$

on $EK(d\frac{X_i-x}{h}) > 0$ soit $\hat{r}_1 = 1/4 \sum_{i=1}^n \delta_i$ et $\hat{r}_2(x) = 1/4 \sum Y_i \delta_i$ on a $\hat{r}(x) = \frac{\hat{r}_2(x)}{\hat{r}_1(x)}$

En utilisant la décomposition

$$\hat{r}(x) - r(x) = \frac{1}{\hat{r}_1(x)} [\hat{r}_2(x) - E\hat{r}_2(x) - (r(x) - E\hat{r}_2(x))] - \frac{r(x)}{\hat{r}_1(x)} [\hat{r}_1(x) - 1] \quad (2.2)$$

Comme $\lim_{n \rightarrow \infty} E\hat{r}_2(x) = (r(x))$,

$$\hat{r}_2(x) - E\hat{r}_2(x) = O\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) p.co$$

$$\text{et } \hat{r}_1(x) - 1 = O\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) p.co$$

sous les conditions de régularité

pour plus de détails on se réfère a[ferraty et view] la version de pondération et le lieu entre les petits boules $\Phi_x(x)$ et la pondération locales.

Lemme 3 *Sous les hypothèses $h_{2,1}, -h_{2,5}$ nous avons*

$$Ef_n(x) - f(x) = O(h_n^\ell), Ef_n(x) - \hat{f}_n(x) = O_{p.co}\left(\sqrt{\frac{\log n}{nh_n}}\right)$$

$$\text{et } \exists \eta > 0, \sum_{n=1}^{\infty} P\left(\hat{f}_n(x) < \eta\right) < \infty$$

La preuve des deux premières relations se fait de la même manière que celles prévisions lemmes en utilisant la constante 1 ,Il reste à prouver la dernière relation. Remarquons que $\hat{f}_n(x)$ converge presque complètement vers f(x),donc il suffit de poser $\eta = \frac{f(x)}{2}$ et remarquer que

$$P\left(\left|\hat{f}_n(x)\right| < f(x)/2\right) \leq P\left(\left|\hat{f}_n(x) - f(x)\right| > \frac{f(x)}{2}\right)$$

2.3 Convergence presque complète uniforme

Théorème 2.2 *Sous les hypothèses $h_1, h_{2,3} - h_{2,5}, H_{1,2}H_{1,3}, h'_{2,1}h'_{2,2}eth'_{2,6}$ nous avons*

$$\sup_{x \in S} |r_n(x) - r(x)| = O(h_n^l) + \left(\sqrt{\frac{\log n}{nh_n}}\right)$$

Démonstration. Nous suivons la même démarche que pour la démonstration du théorème (2.1) et (??) L'utilisation de les relation Précédemment et les quatre lemmes suivants permettent de déduire directement le résultat du théorème.

Lemme 4 *Sous les hypothèses , nous avons $h_1, h_{2,3}H_{1,3}$*

$$\sup_{x \in S} |r_n(x) - \hat{r}(x)| = O_{p.co} + \left(\sqrt{\frac{\log n}{nh_n}} \right)$$

Démonstration : est la même que celle du lemme 1

Lemme 5 *Sous les hypothèses $h_{2,4}, h_{2,5}, h'_{2,1}$ nous avons*

Lemme 6 *Sous les hypothèses $h_{2,4}, h_{2,3}H_{1,3}, h'_{2,1}, h'_{2,6}$*

$$\sup_{x \in S} |E\hat{r}_{n,N}(X) - \hat{r}_{n,N}(x)| = O_{p.co} \left(\sqrt{\frac{\log n}{nh_n}} \right)$$

Démonstration. Dans cette preuve C' désigne une constante générique. La compacité de S nous permet d'écrire $S \subset U_{k=1}^{z_n} t_k - l_n, t_k + l_n$ où l_n et z_n vérifient

$$l_n = C' z_n^{-1} e t z_n \sim C' n^{\frac{\beta+1}{\beta}} \quad (2.3)$$

on posant $t_x = \arg \min_{t \in \{t_1, t_2, \dots, t_{z_n}\}} |x - t|$ nous obtenons

$$\begin{aligned} \sup_{x \in S} |E\hat{r}_{n,N}(x) - \hat{r}_{n,N}(x)| &\leq \sup_{x \in S} |\hat{r}_{n,N}(x) - \hat{r}_{n,N}(t_x)| \\ &+ \sup_{x \in S} |E\hat{r}_{n,N}(x) - E\hat{r}_{n,N}(t_x)| + \sup_{x \in S} |E\hat{r}_{n,N}(x) - \hat{r}_{n,N}(t_x)| \end{aligned}$$

En utilisant les hypothèses $H_{1,3}, h'_{2,6}$ nous obtenons

$$\sup_{x \in S} |\hat{r}_{n,N}(x) - \hat{r}_{n,N}(t_x)| \leq C' \frac{l_n^\beta}{h_n^{1+\beta}}$$

et nous pouvons déduire que

$$\sup_{x \in S} |E\hat{r}_{n,N}(x) - \hat{r}_{n,N}(t_x)| \leq C' \frac{l_n^\beta}{h_n^{1+\beta}}$$

Maintenant, en vertu de (2,3)et de l'hypothèse $H_{1,3}$ nous pouvons voir que $\frac{l_n^\beta}{h_n^{1+\beta}} \sqrt{nh_n} \rightarrow 0$ nous en déduisons que pour tout $\varepsilon > 0$ et pour n suffisamment grand

$$P \left(\frac{l_n^\beta}{h_n^{1+\beta}} \sqrt{\frac{nh_n}{\log n}} > \varepsilon \right) = 0$$

Finalement, l'utilisation de la relation (??) et (??) permet d'obtenir

$$\begin{aligned} & P \left[\sup_{x \in S} |E\hat{r}_{n,N}(t_x) - \hat{r}_{n,N}(t_x)| > \varepsilon_0 \sqrt{\frac{nh_n}{\log n}} \right] \\ & P \left[\max_{k \in 1, \dots, z_n} |E\hat{r}_{n,N}(t_k) - \hat{r}_{n,N}(t_k)| > \varepsilon_0 \sqrt{\frac{nh_n}{\log n}} \right] \\ & \leq 2z_n e^{-\frac{n\varepsilon_0^2 h_n}{4C}} \leq C' n^{\frac{\beta+1}{\beta} - \frac{\varepsilon_0^2}{4C}} \end{aligned}$$

Le choix d'un ε_0 suffisamment grand permet d'avoir

$$\sum_1^n P \left[\sup_{x \in S} |E\hat{r}_{n,N}(t_x) - \hat{r}_{n,N}(t_x)| > \varepsilon_0 \sqrt{\frac{nh_n}{\log n}} \right] < \infty$$

Lemme 7 *Sous les hypothèses $h_{2,3}, -h_{2,5}, h'_{2,1} h'_{2,2} \text{eth}'_{2,6}$ nous avons*

$$\begin{aligned} \sup |Ef_n(x) - f(x)| &= O(h_n^\ell), \sup_{x \in S} |Ef_n(x) - f(x)| = O_{p.co} \left(\sqrt{\frac{\log n}{nh_n}} \right) \\ \text{et } \exists \eta > 0, \sum_{n=1}^{\infty} P \left[\inf_{x \in S} |f_n| \leq \eta \right] &< \infty \end{aligned}$$

Démonstration. La preuve des deux premières relations se fait de la même manière que celles des lemme ??,?? en utilisant la constante 1 au lieu de la variable $\frac{1_{\{A_i=0\}} Z_i}{S_R(Z_i) F_L(Z_i)}$. Il reste à prouver la dernière relation. Remarquons que $\hat{f}_n(x)$ converge presque complètement vers $f(x)$ donc il suffit de poser $\eta = \frac{f(x)}{2}$ et de remarquer que

$$P \left(\left| \hat{f}_n(x) < \frac{f(x)}{2} \right| \right) \leq P \left(\left| \hat{f}_n(x) - f(x) \right| > \frac{f(x)}{2} \right)$$

Chapitre 3

Application et exploration des observations sur des données simulées

cette section est spécifique à la simulation de la régression dans le cas d'indépendance des variables aléatoires.

Pour l'effecteur, On considère un échantillon de taille =1000 la fonction théorique choisi est :

$$y = \sin(x/10) + (x/50)^2$$

son estimateur :

3.1 Simulée les données

on a la fonction de régression du noyau gaussien représenté par

$$y = \sin(x/10) + (x/50)^2$$

3.1.1 Algorithme 1

- Nous choisirons la fonction pour représenter la régression
 $y = \sin(x/10) + (x/50)^2$
- Nous prendrons les valeurs de variables du réponse.
- Nous utilisons le résultat extrait

3.1.2 Programme sous Matlab

premièrement gaussien kern reg

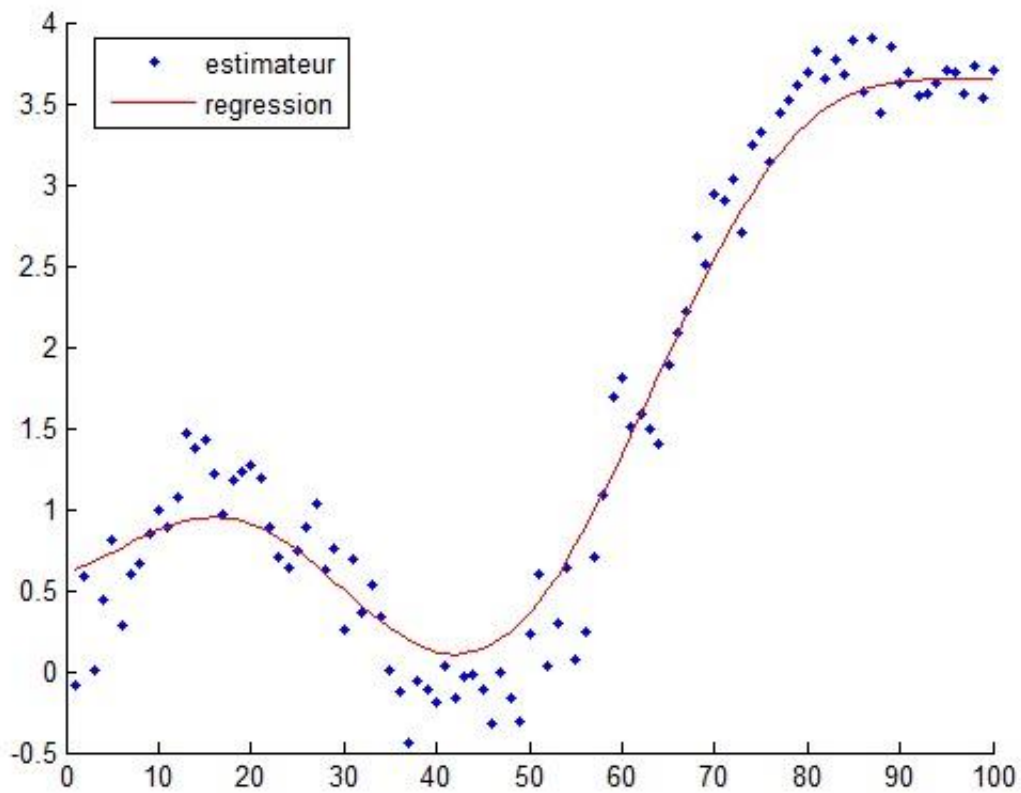
```
function gaussien kern reg(xs(i), x, y, h)
kerf = @(z)exp(-z.*z/2)/sqrt(2*pi);
z = kerf((xs - x)/h);
ys = sum(z.*y)/sum(z);
```

deuxièmement main function

```
x = 1 : 100;
y = sin(x/10) + (x/50).^2 + 0.2(randon(1,100));
randon(e) = e + i h = 5; for i = 1 : 100
xs(i) = i;
ys = (i) = gaussienkernreg(xs(i), x, y, h);
end
figure;hold on;
plot(x,y,'.');
plot(xs,ys,'r');
title(regression,'-',estimateur,'...')
```

3.2 Résultats de simulation

Avec le programme du logiciel Matlab , on obtient la courbe dans la figure qui représente la fonction théorique comme régression et son estimateur.



la valeur de la largeur fenêtre h a été choisi par balayage selon le critère MCE comme l'indique le tableau suivant :

la valeur retenu selon le critère MCE $MCE = \frac{1}{100} \sum_{i=1}^{100} (y_i - \hat{y}_i)^2$ le choix de l'initialisation de l'algorithme est $x_0 = x_{100}$

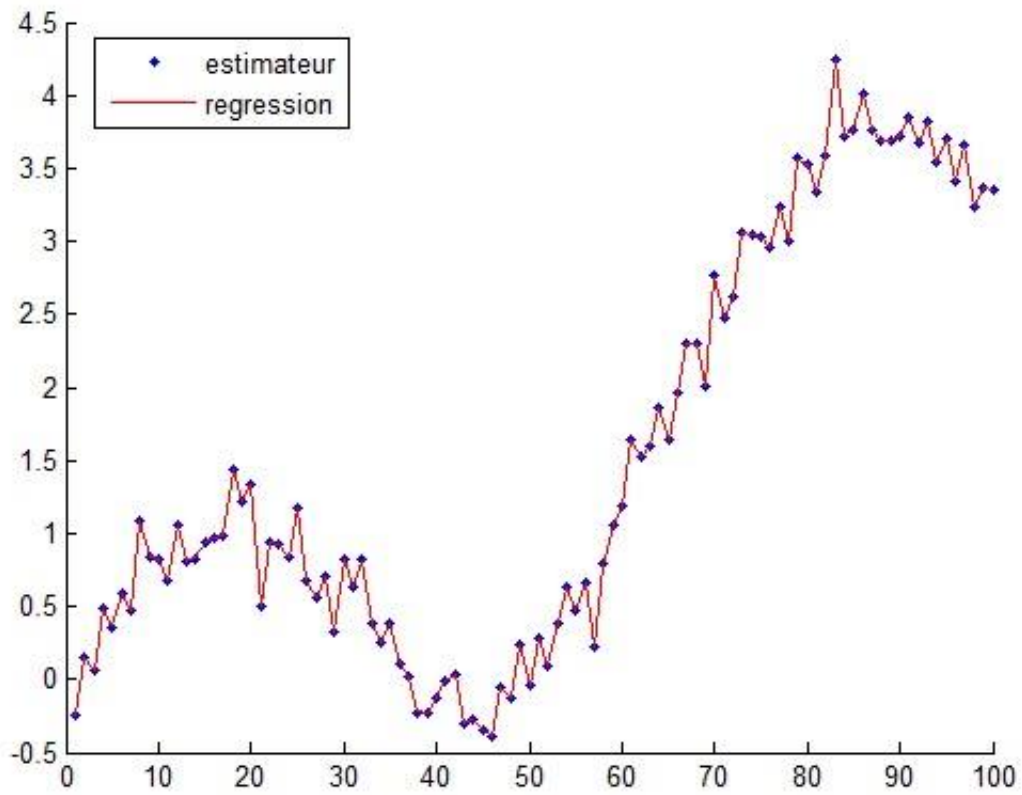
CHAPITRE 3. APPLICATION ET EXPLORATION DES OBSERVATIONS SUR DES DONNÉES SIMULÉES

h	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
MCE	6.3910	5.4666	4.4705	4.8462	5.9370	8.2111	14.3888	30.1714	63.7808

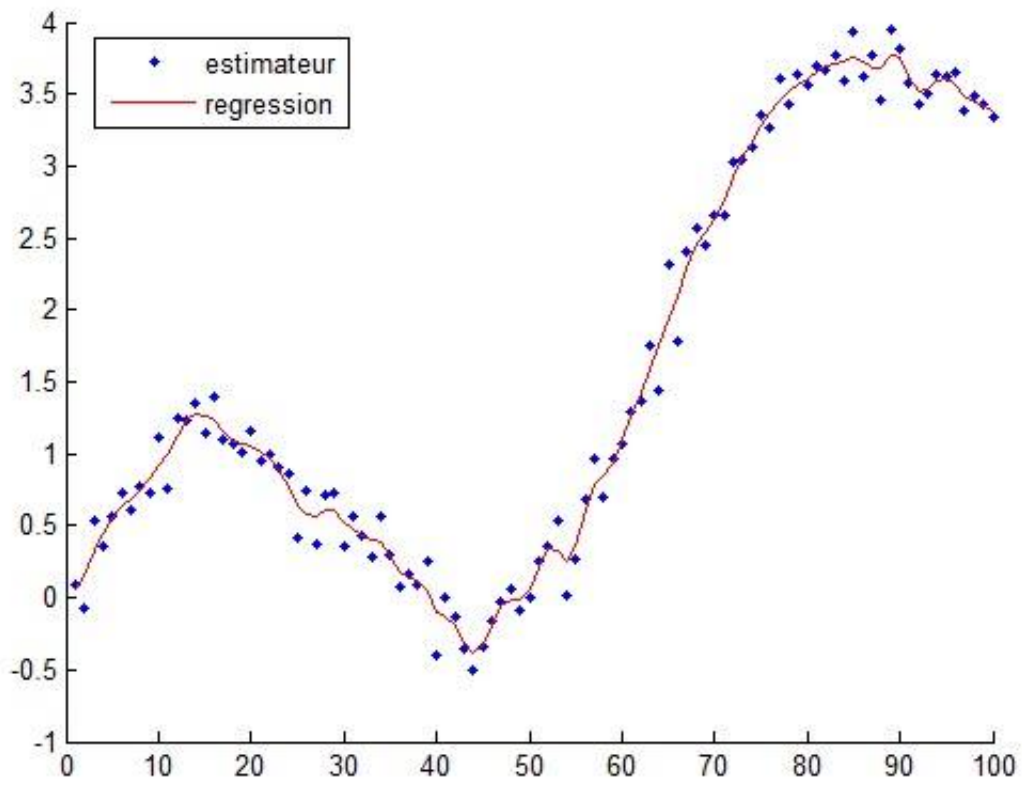
Commentaires

On remarque que les courbes obtenu sont proches dans les approximations sont bonnes.
remarque maintenant,on changé dans la paramètre de lissage h . On utilise les variables suivant de h pour $h=0.1$ on la courbe suivant :

CHAPITRE 3. APPLICATION ET EXPLORATION DES OBSERVATIONS SUR DES DONNÉES SIMULÉES



pour $h=0.3$ on la courbe suivant :



CONCLUSION

L'estimation de la régression joue un très important dans la prédiction .A ce effet il fascine les chercheurs de toutes disciplines.

Dans ce mémoire,nous avons propose d'avoir une synthèse sur la convergence presque complète des estimateurs .Notre objectifs est base sur l'estimateur de la courbe de régression.

Avant de réaliser l'étude en question,nous avons exposé,les principales notion que nous voyons intéressantes pour l'étude du problème de l'estimation ,qui sont suivi de quelques outils probabilistes et statistique qui nous permettant étudier l'approximation entre une telle courbe et son est estimateur .parmi les modèle existant nous avons utilisé le modèle de régression non paramétrique .Comme nous avons mis l'accent sur la méthode du noyau qui s'utilise dans dans ce genre de situation .En effet ,nous avons introduit le principale de base de l'estimation a noyau de parzan-rosenblant, ses propriété et les méthodes de sélection du noyau et du paramétrique de lissage .

En fin nous avons fait une simulation sur un exemple avec le logiciel Matlab qui nous a permis de valider notre étude théorique.

Bibliographie

- [1] Bowman, A. W. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* 71, 2 (1984), 353–360.
- [2] Coudret, R., Durrieu, G., and Saracco, J. Estimateurs a noyau bimodaux d'une densité bimodale et comparaison avec d'autres estimateurs non paramétriques. *Proc de la société Française de Statistique* (2012).
- [3] Deheuvels, P., and Hominal, P. Estimation non paramétrique de la densité compte tenu d'informations sur le support. *Revue de Statistique Appliquée*, 27, 47,68 (1979).
- [4] Hall, P. Cross validation in density estimation. *Biometrika* ,69 (1982), 383–390.
- [5] Nadaraya, E. On non parametric estimation density function and regression. *Theory Probab P.P.L* (1965).
- [6] Cybakov, A. B. Introduction à l'estimation non paramétrique, vol. 41. Springer Science Business Media, 2003.
- [7] Delsol, L. (2007). Régression non paramétrique fonctionnelle : expression asymptotique des moments, *Ann. I.S.U.P. Vol LI*, 3, 43-67.
- [8] BITOUZÉ, B. LAURENT et P. MASSART : A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 35(6) :735–763, 1999. ISSN 0246-0203.
- [9] Devroye, L. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canad. J. Statist.* 6, No.2, 179-191.

- [10] G. Collomb. Estimation de la régression pour la méthode du noyau : Quelques propriétés de convergence uniforme. C.R. Acad. Sc. Paris, 1977b.
- [11] J. Fox. Non parametric regression. 2003.
- [12] K. KEBABI et F. MESSACI : Rate of the almost complete convergence of a kernel regression estimate with twice censored data. Statistics Probability Letters, 82(11) :1908–1913, 2012.
- [13] KHARDANI, M. LEMDANI et E. OULD SAÏD : Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series. Journal of Statistical Planning and Inference, 141(11) : 3426–3436, 2011.
- [14] E. A. NADARAYA : Nonparametric estimation of probability densities and regression curves. Springer, 1989.

Abstract

Estimation regression has great role in prediction.

In this work we have 3 chapter, the first one about parametric and non parametric ,kernel method and its definition and their characteristics and smoothing parameter .

Second chapter talks about convergence almost filled for estimator,the aim for this work is estimation of the regression curve.

The last chapter we use matlab for simulation regression in case independance and we find that estimator and regression are close.

Key words : non parametric ,kernel method, smothing parameter

Résumé

L'estimation de la régression joue un rôle très important dans la prédiction.

Dans ce mémoire on a 3 chapitre, premier chapitre on parle on méthode non paramétrique est méthode du noyau sa définition et définition de model paramétrique et non paramétrique avec leurs propriétés statistique et asymétrique et on utilisant le paramètre de lissage

Deuxième chapitre nous avons proposé d'avoir une synthèse sur la convergence presque complète des estimateurs .notre objectifs est base sur l'estimation de la courbe de régression

Finalement, on utilise le logiciel matlab pour simulation de la régression dans cas indépendance et on trouve que l'estimation et la régression sont proches.

Mots clé : non paramétrique, méthode du noyau, paramètre de lissage

المخلص

التقدير الانحدار يلعب دورا مهما في التنبؤ

في هاته الرسالة لدينا 3 فصول في الفصل الأول نتحدث عن التقدير المعلمي والامعلمي وطريقة النواة مع ذكر الصيغ والمعادلات لكل منها من اجل استخدامها في تحقيق التقارب كما نستخدم قيمة المعلمة او النافذة

في الفصل الثاني نقدم نظريات حول التقارب الشبه مكتمل للمقدر ونبرهن عليه هدفا يستند على تقدير منحني الانحدار

أخير الفصل 3 نستخدم برنامج الماتلاب من اجل محاكاة المقاربة في حاله مستقلة لكي نجد ان الانحدار والتقدير يتقاربان وهو الهدف من البحث.

الكلمات المفتاحية الالامعلمي طريقه النواه معلمة النافذة