

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ KASDI MERBAH OUARGLA
FACULTÉ DES MATHÉMATIQUES ET SCIENCES DE LA MATIÈRE



DÉPARTEMENT DE MATHÉMATIQUES
MÉMOIRE PRÉSENTÉ EN VUE DE L'OTENTION DU DIPLÔME :

MASTER

Spécialité : **Mathématiques**

Option : **Probabilités Et Statistique**

Par : Bencheikh Samia

Thème

Sur L'estimation non paramétrique de la densité et la fonction de régression

Membres du Comité d'Examen :

Dr. Koudri Mohammed	UKM	Président
Dr. Chahad Abdelkader	UAD Adrar	Encadreur
Dr. Mansoul Brahim	UKM	Examineur

Juin 2021

DÉDICACE

*Je dédie ce modeste travail : Dieu soit loué, qui m'a béni pour terminer ce
mémoire et m'a approuvé.*

Je dédie le fruit de mes efforts à celui qui est resté avec moi contre vents et marées de
ma chère mère, Zainuna Al-Saeedi,

et à ceux qui m'ont soutenu dans la vie, mon cher père bien-aimé, Muhammad,
et à mes sœurs qui étaient mon soutien, Rashid et sa petite famille, Abdul Ghani et sa
petite famille,

Muhammad Al-Eid et sa petite famille, Fateh et sa petite famille et Abdul Haq
. Merci pour votre soutien et à mes sœurs et sœurs chères qui ont soutenu moi, Wassila
et sa petite famille,

Karima et Asma et Shaima, je les remercie pour leur soutien,
à mes amis qui m'ont soutenu, à tous mes proches et à tous ceux qui m'ont soutenu de
près ou de loin.

REMERCIEMENTS

Je tiens premièrement à me prosterner, remerciant «**Allah**» le tout puissant de m'avoir donné le force et la volonté pour terminer ce travail. J'exprime également mes sincères remerciements, ma gratitude et mon appréciation à tous ceux qui m'ont aidé à accomplir cette humble recherche, et je voudrais mentionner l'honorable professeur :Chahad Abdelkader , que Dieu le protège, qui ne m'a pas épargné ses connaissances, et des orientations précieuses pour mener à bien ce travail. Je lui souhaite encore plus de succès et de réussite. Dans sa vie pratique et scientifique, j'exprime également mes remerciements et ma gratitude à tous mes honorables professeurs qui m'ont supervisé.

J'ai été formé tout au long de mon parcours universitaire, et je ne manque pas d'exprimer mes remerciements et ma reconnaissance aux professeurs du comité de discussion : le professeur Kouidri Mohammed et le professeur Mansoul Brahim. Et je leur souhaite ainsi qu'à tous les professeurs Faculté de la médecine, des mathématiques, des travailleurs de l'administration et de la réussite de tous les étudiants.

Table des matières

Remerciements	ii
Table des matières	iii
Introduction	1
1 Rappels et Définitions	3
1.1 Espérance et variance	3
1.2 Les types des convergences des variable aléatoires	4
1.2.1 Convergence en loi	4
1.2.2 Convergence presque sûre	5
1.2.3 La convergence en probabilité	5
1.2.4 Convergence en moyenne quadratique	5
2 Estimation non paramétrique de la fonction du densité	7
2.1 Estimation de la densité par histogramme	7
2.1.1 Propriétés statistiques de l'estimateur par histogramme	8
2.2 Estimation de la densité par noyau	12
2.2.1 Les propriétés de l'estimateur à noyau	17
2.2.2 L'erreur quadratique moyenne	20
2.3 Choix de noyau optimal	20
2.4 Choix paramètre de lissage	22

3 Régression non paramétrique	28
3.1 Estimation de la fonction de régression par noyau	28
3.1.1 La méthode du noyau	28
3.1.2 Définition	30
3.1.3 Propriétés asymptotiques de Nadraya Watson	32
3.1.4 Statistiques de l'estimateur Nadaraya-Watson	35
Conclusion	43
Bibliographie	44
Annexe B : Abréviations et Notations	45

Introduction générale

En statistique, il existe deux types d'estimation, l'estimation paramétrique et l'estimation non paramétrique. L'estimation paramétrique est utilisée lorsque nous connaissons la distribution des données, mais que certains paramètres de la distribution sont inconnus. Nous utilisons alors les techniques d'estimation paramétrique pour estimer les paramètres inconnus. Les méthodes paramétriques les plus utilisées sont la méthode du maximum de vraisemblance, la méthode des moments, la méthode de moindres carrés et par intervalle de confiance).

Par contre l'estimation non paramétrique ne nécessite aucune hypothèse ou information sur les caractéristiques de la distribution de base de la population et les données de l'échantillon. c'est pour ça on trouve que leur champ d'application est plus large que celui des statistiques paramétriques.

Donc dans le cas où l'on n'a pas d'a priori sur la forme de la loi inconnue, on doit alors estimer des fonctions, (par exemples : une fonction de densité f ou une fonction de régression r) et non plus des paramètres. C'est l'objet de la statistique non-paramétrique, qui nécessite moins de connaissances préalables de la loi. En contrepartie, il faut plus de données pour obtenir une précision d'estimation équivalente à celle du cadre paramétrique.

En statistique non paramétrique, il existe différents types d'estimation de la fonction de densité : l'estimation de densité de noyau univarié, l'estimation de la densité conjointe et l'estimation de densité conditionnelle.

Dans ce mémoire, nous allons concentrer notre étude sur :

— l'estimateur de densité (univarié) par noyau.

Il s'agit d'une manière non paramétrique d'estimer la fonction de densité $f(x)$ de probabilité d'une variable aléatoire X , nous allons introduire quelques définitions

et théorèmes de base . Ensuite, nous étudierons les propriétés statistiques et les propriétés asymptotiques de cet estimateur. Nous étudierons également le MSE , le A-MISE et le choix de paramètre associée à cet estimateur.

- Pour l'estimateur de la fonction de régression il y a plusieurs méthodes d'estimation non paramétriques : Estimation par noyau , estimation par des polynômes, estimation sur des bases de splines, et par projection sur des bases orthonormées. Dans ce cadre de l'estimation d'une fonction de régression, que nous noterons r . La variable aléatoire à expliquer sera notée Y , tandis que la variable aléatoire explicative sera notée X . Toutes les variables considérées seront supposées être définies sur le même espace probabilisé (Ω, \mathcal{A}, P) . Notre modèle de régression peut être écrits sous la forme :

$$Y = r(X) + \epsilon$$

Où ϵ est une variable aléatoire centrée et indépendante de X .

Dans ce cas-là, l'objectif consiste à déterminer la fonction $r(X)$. L'idée répandue actuellement est d'estimer $r(X)$ par un lissage local. Le lissage signifie : « laisser les données nous montrer la forme de la fonction de régression » (Latraverse,M) . Nous allons introduire quelques définitions et théorèmes de base . Ensuite, nous étudierons les propriétés statistiques et les propriétés asymptotiques de cet estimateur.

Notre travail se décompose en trois chapitres :

Dans le premier chapitre,nous rappelons quelques préliminaires sur les espérance et variance, Les types des convergences des variable aléatoires.

Dans le deuxième chapitre,nous présentons estimation non paramétrique de la fonction du densité en utilisant la méthode estimation de la densité par histogramme et par noyau .

Et dans la dernière chapitre,nous parlerons par régression non paramétrique par la méthode du noyau dans lequel il y a l'estimation de Nadraya Watson et ses propriétés.

Chapitre 1

Rappels et Définitions

1.1 Espérance et variance

Définition 1.1.1 Soit X une variable aléatoire ayant pour densité f . On dit que X admet une espérance si l'intégrale $\int_{-\infty}^{+\infty} |x| f(x) dx$ converge. Dans ce cas, on appelle espérance de X le réel $E(X)$ défini par :

$$E(x) = \int_{-\infty}^{+\infty} x f(x) dx \quad (1.1)$$

Théorème 1.1.1 Soit X une variable aléatoire admettant f pour densité, et soit $\phi : \Omega \rightarrow \mathbb{R}$ une fonction continue. On suppose en outre que :

$\int_{-\infty}^{+\infty} |\phi(x)| f(x) dx$ converge. Alors la variable aléatoire $\phi(X)$ admet une espérance donnée par :

$$E(\phi(x)) = \int_{-\infty}^{+\infty} \phi(x) f(x) dx$$

Définition 1.1.2 La variance d'une variable aléatoire X est définie par :

$$\begin{aligned} \text{Var}(X) &= E[(X - E(X))^2] = E(X^2) - [E(X)]^2 = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx \\ &=; \int_{-\infty}^{+\infty} (x)^2 f(x) dx - \left(\int_{-\infty}^{+\infty} x f(x) dx \right)^2 \end{aligned}$$

Définition 1.1.3 *On dit que l'estimateur $\hat{\theta}_n(x)$ de $\theta(x)$ est sans biais si :*

$$E(\hat{\theta}_n(x)) = \theta(x) \tag{1.2}$$

Définition 1.1.4 — *Le Biais de l'estimateur $\hat{\theta}_n(x)$ est :*

$$E(\hat{\theta}_n(x)) - \theta(x)$$

— *L'estimateur $\hat{\theta}_n(x)$ est asymptotiquement sans biais si :*

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n(x)) = \theta(x)$$

1.2 Les types des convergences des variable aléatoires

1.2.1 Convergence en loi

Définition 2.1.1

On dit que X_n converge en loi vers la (v.a) X si l'on a, en tout x où sa fonction de répartition F_X est continue

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

et on note

$$X_n \xrightarrow[n \rightarrow \infty]{L} X$$

On dira aussi que la loi de X est la loi limite ou asymptotique de la suite X_n . En pratique la loi limite sera utile pour donner une approximation pour le calcul de la probabilité d'un événement sur X_n quand n sera assez grand.

1.2.2 Convergence presque sûre

On dit que X_n converge presque sûrement (ou converge avec probabilité 1, ou converge fortement) vers la (v.a) X si les (v.a) au point $x \in \mathfrak{R}$ si, quel que soit $\varepsilon > 0$ donné,

$$\lim_{n \rightarrow \infty} P(\sup |X_m - X| < \varepsilon) = 1$$

et on note

$$X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$$

il est clair que la convergence presque sûre entraîne la convergence en probabilité (d'où les qualificatifs de convergence forte et convergence faible)

1.2.3 La convergence en probabilité

On dit que X_n converge en probabilité vers la (v.a) X si, quel que soit $\varepsilon > 0$ donné

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$$

et note

$$X_n \xrightarrow[n \rightarrow \infty]{p} X$$

1.2.4 Convergence en moyenne quadratique

On dit que X_n converge en moyenne quadratique vers la (v.a) X si les (v.a) X_1, X_2, \dots ont un moment d'ordre 2 et si

$$\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0$$

et on note

$$X_n \xrightarrow[n \rightarrow \infty]{m.q.} X$$

On admettra la relation d'implications suivantes entre les différents types de convergence :

$$X_n \xrightarrow[n \rightarrow \infty]{p} X \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{L} X$$

$$X_n \xrightarrow[n \rightarrow \infty]{m.q} X \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{p} X$$

$$X_n \xrightarrow[n \rightarrow \infty]{p.s} X \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{p} X$$

On lui préfère souvent $S_n^2 = \left(\frac{1}{n-1}\right) \sum_{i=1}^n (X_i - \bar{X})^2$ qui est sans biais, appelée variance empirique corrigé Médiane empirique $\hat{m}_n = F_n^{-1}(1/2)$ ou plus généralement le quantile empirique $\hat{q}_\alpha = F_n^{-1}(\alpha)$

Chapitre 2

Estimation non paramétrique de la fonction de densité

Dans ce chapitre on s'intéresse de l'étude de l'estimation de la fonction de densité f d'une variable aléatoire et ses propriétés.

2.1 Estimation de la densité par histogramme

L'histogramme est l'estimateur de la densité non paramétrique le plus ancien et le plus simple, il est considéré comme un estimateur de la densité de probabilité sous-jacente à un ensemble fini. Cette méthode consiste à estimer f en un point x par la proportion des variables aléatoires $(X_1; X_2; \dots; X_n)$, qui se trouvent dans un intervalle de longueur h et qui contient x . Elle est donc basée sur le choix d'un point d'origine a_0 et d'une partition $A_k = ([ak; a_k + 1[) k = 1; \dots; p$ en p classes du support de X . Si nous notons n_k le nombre de variables dans la classe A_k et $h = a_{k+1} - a_k$, l'estimateur de f sur A_k du type histogramme est définie par :

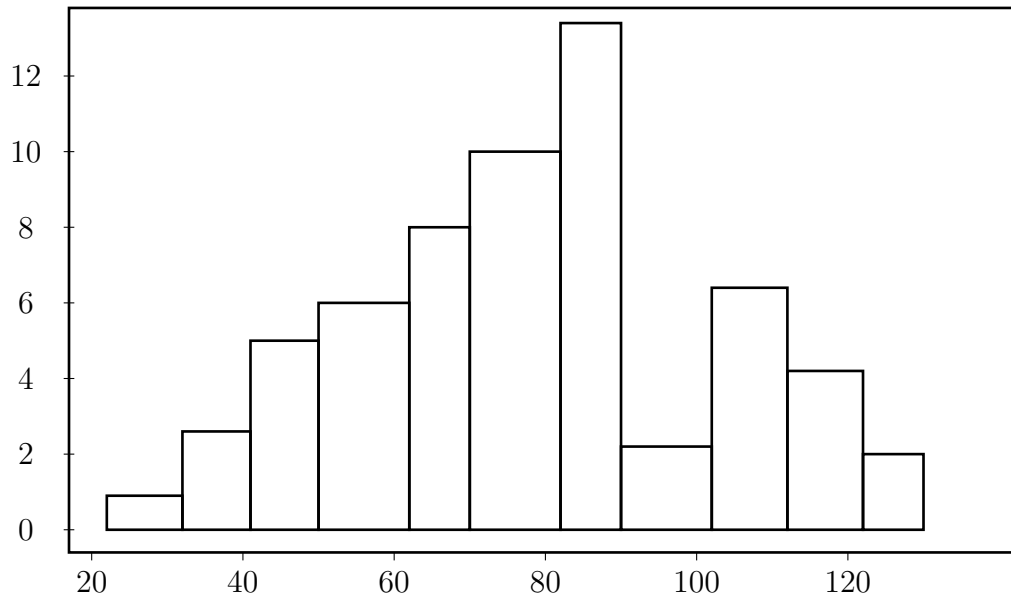
$$\hat{f}_h^n(x) = \frac{n_k}{nh} = \frac{1}{nh} \sum_1^n \mathbf{1}_{A_k}(x_i) \quad x \in A_k \quad (2.1)$$

Dans la suite, nous émettons l'hypothèse que les classes $A_k, k \in \{1, \dots, p\}$ forment une partition de Ω (i.e $\Omega = \cup_{k=1}^p A_k$ et $\forall i, j \in \mathbf{N} : i \neq j, A_i \cap A_j = \Phi$) et définissons pour chaque classe A_k , son centre a_k telles que : $\forall k \in \{1, \dots, p\}, A_k = [a_k - \frac{h}{2}, a_k + \frac{h}{2}]$ et $\forall k \in \{1, \dots, p-1\}, a_{k+1} = a_k + h$

Et par suite on peut écrire :

$$\hat{f}_h^n(x) = \frac{n_k}{nh} = \frac{1}{nh} \sum_1^n \mathbf{1}_{[a_k - \frac{h}{2}, a_k + \frac{h}{2}]}(x_i) \quad x \in [a_k - \frac{h}{2}, a_k + \frac{h}{2}]$$

Exemple de histogramme 2.1.1



Chiffre 1.1 : Histogramme de fréquence de l'ensemble de données sur les chutes de neige à Buffalo

2.1.1 Propriétés statistiques de l'estimateur par histogramme

Nous présentons, dans ce titre, les propriétés statistiques de l'estimateur par histogramme \hat{f}_h^n défini par l'expression 2.1 (pour une étude plus détaillée voir, par exemple, les livres de Bosq et Lecoutre [5], Simonoff [10], Härdle [7]). En statistiques, il est nécessaire de mesurer la qualité d'un estimateur. Pour cela, on évalue, d'une part, l'écart entre la moyenne de l'estimateur et la densité à estimer, ce critère d'évaluation est appelé biais, et d'autre part, la variance de l'estimateur (due au caractère aléatoire d'observations) qui caractérise la dispersion des valeurs de l'estimateur dans l'ensemble d'observations. On essaye généralement de réduire au mieux ces deux quantités.

Biais de l'estimateur

Le biais de l'estimation \hat{f}_h^n définie en 2.1 est donnée, pour tout $x \in A_k, \forall k \in \{1, \dots, p\}$, par :

$$\begin{aligned} \text{Biais } \hat{f}_h^n(x) &= \mathbf{E}(\hat{f}_h^n(x)) - f(x) \\ &= \frac{1}{2}f'(h - 2(x - a_k)) + O(h^2) \end{aligned} \quad (2.2)$$

Où f' la dérivée de f . qui doit être une fonction de $L(\Omega)$ absolument continue et carrée intégrable ($\int_{-\infty}^{+\infty} f'(x) < \infty$) Nous observons que le biais de l'histogramme diminue dans l'ordre $O(h)^2$ lorsque h diminue. Pour étudier la stabilité de l'estimation on calculera sa variance

Variance de l'estimateur

La Variance de l'estimateur \hat{f}_h^n définie en 2.1 est donné pour tout $x \in \Omega$ par :

$$\begin{aligned} \text{Var}(\hat{f}_h^n(x)) &= \mathbf{E}((\hat{f}_h^n(x))^2) - (\mathbf{E}(\hat{f}_h^n(x)))^2 \\ &= \frac{f(x)}{nh} + O\left(\frac{1}{n}\right) \end{aligned} \quad (2.3)$$

Cette variance tend vers zéro quand nh tend vers l'infini et le nombre d'observation n tend vers l'infini.

Afin d'apprécier la qualité de l'estimateur, il faut évaluer la distance entre l'estimateur et la densité à estimer. La distance la plus couramment utilisée est celle définie par la moyenne du carré de leur différence. Elle porte le nom d'erreur quadratique moyenne en anglais mean squared error (MSE) .

$$MSE((\hat{f}_h^n(x)) = \mathbf{E}((\hat{f}_h^n(x) - f(x))^2) \quad (2.4)$$

Proposition 2.1.1

$$\begin{aligned}
 MSE(\hat{f}_h^n)(x) &= \mathbf{E}((\hat{f}_h^n(x) - f(x))^2) \\
 &= \mathbf{E}(\hat{f}_h^n(x) - \mathbf{E}\hat{f}_h^n(x) + \mathbf{E}\hat{f}_h^n(x) - f(x))^2 \\
 &= \mathbf{E}(\hat{f}_h^n(x) - \mathbf{E}\hat{f}_h^n(x))^2 + \underbrace{2\mathbf{E}(\hat{f}_h^n(x) - \mathbf{E}\hat{f}_h^n(x))(\mathbf{E}\hat{f}_h^n(x) - f(x))}_{=0} \\
 &\quad + (\mathbf{E}\hat{f}_h^n(x) - f(x))^2 \\
 &= \mathbf{E}(\hat{f}_h^n(x) - \mathbf{E}\hat{f}_h^n(x))^2 + (\mathbf{E}\hat{f}_h^n(x) - f(x))^2 \\
 &= (\text{Biais}(\hat{f}_h^n(x)))^2 + \text{Var}(\hat{f}_h^n(x))
 \end{aligned}$$

La convergence en moyenne quadratique de l'estimateur \hat{f}_h^n 2.1 a été établie par Lecoutre [9] pour tout $x \in A_k$.

$$\begin{aligned}
 MSE(\hat{f}_h^n)(x) &= (\text{Biais}(\hat{f}_h^n(x)))^2 + \text{Var}(\hat{f}_h^n(x)) \\
 &= \left(\frac{1}{2}f'(h - 2(x - a_k))\right)^2 + \frac{f(x)}{nh} + O(n^{-1}) + O(h^3)
 \end{aligned}$$

Cette erreur quadratique moyenne tend vers zéro quand h tend vers zéro et nh tend vers l'infini quand n tend vers l'infini. Ce critère d'erreur quadratique moyenne est un critère local. Pour généraliser ce critère sur tout le domaine Ω on intègre sur tout le domaine Ω . Ce critère porte le nom d'erreur quadratique moyenne intégrée en anglais mean integrated squared error (MISE)

$$MISE(\hat{f}_h^n) = \int_{\Omega} (\hat{f}_h^n - f)^2(x) dx = \int_{\Omega} MSE(\hat{f}_h^n)(x) dx \quad (2.5)$$

La convergence en moyenne quadratique intégrée de l'estimateur \hat{f}_h^n défini par 2.1 a été établie par Lecoutre [9]. Nous avons, d'après [10] :

$$MISE((\hat{f}_h^n)) = \frac{1}{nh} + \frac{h^2 \int_{\Omega} (f')^2(x) dx}{12} + O(n^{-1}) + O(h^3) \quad (2.6)$$

Et on a : $\lim MISE = O$ quand $h \rightarrow 0$ et $nh \rightarrow \infty$. Notons que $\int_{\Omega} (f')^2$, décrit le degré de régularité de la fonction de densité f . Asymptotiquement, le terme $O(n^{-1}) + O(h^3)$ peut être ignoré, comme $h \rightarrow 0$ et $nh \rightarrow \infty$. Le terme dominant de MISE est défini comme MISE asymptotique (A-MISE).

$$A - MISE((\hat{f}_h^n)) = \frac{1}{nh} + \frac{h^2 \int_{\Omega} (f')^2(x) dx}{12} \quad (2.7)$$

L'approche usuelle pour minimiser MISE est de minimiser A-MISE en fonction de h . Appelez le paramètre de lissage optimal h_o (h_{opt}). Nous pouvons trouver h_o en différenciant par rapport à h

$$\frac{\partial(A - MISE)}{\partial h} = \frac{-1}{nh^2} + \frac{1}{6}h \int_{\Omega} (f')^2(x) dx$$

Et par suite :

$$\begin{aligned} \frac{\partial(A - MISE)}{\partial h} = 0 &\Leftrightarrow \frac{-1}{nh^2} + \frac{1}{6}h \int_{\Omega} (f')^2(x) dx = 0 \\ &\Leftrightarrow h_o = \left(\frac{6}{\int_{\Omega} (f')^2(x) dx} \right)^{\frac{1}{3}} n^{-\frac{1}{3}} \end{aligned}$$

Pour $h = h_o$ on obtient la valeur optimale de (A-MISE) :

$$A - MISE^* = \left(\frac{9 \int_{\Omega} (f')^2(x) dx}{16} \right)^{\frac{1}{3}} n^{-\frac{2}{3}}$$

L'approche la plus simple pour choisir h est de choisir une densité particulière f et de simplement la substituer dans pour obtenir une valeur de h . Le choix typique est une densité normale (gaussienne). On peut montrer que le minimum d'A-MISE est :

$$h_o = 3,49\sigma n^{-\frac{1}{3}}$$

Nous constatons que l'histogramme a de bonnes propriétés statistiques, par contre il n'est robuste ni pour le choix du paramètre de lissage h , ni pour celui de l'origine a_0 . Le deuxième inconvénient est sa discontinuité qui ne peut pas s'adapter au cas où la densité f à estimer, vérifie certaines hypothèses de régularité.

Pour résoudre ce problème, l'estimateur de Parzen Rosenblatt (estimation par noyau) a été introduit, il généralise intuitivement la méthode d'estimation par histogramme, et il est très utilisé en estimation non paramétrique.

2.2 Estimation de la densité par noyau

L'idée de la méthode des noyaux est due à Parzen - Rosenblatt (1956). Celui-ci a proposé une sorte d'histogramme mobile où la fenêtre de comptage des observations se déplace avec la valeur de x . La densité en x est estimée par la fréquence relative des observations dans l'intervalle $[x - h, x + h]$, donc centré sur x , divisée naturellement par la largeur de l'intervalle $2h$. On appelle h la largeur de fenêtre (bien que cette largeur soit en fait égale à $2h$).

Construction de l'estimateur Soit $X_1; \dots; X_n$ n variables aléatoires indépendantes et identiquement distribués, et soit f la fonction de densité de probabilité par rapport à la mesure de Lebesgue de \mathbb{R} dans $[0; +\infty[$. Notre objectif est la construction d'un estimateur de \hat{f}_h^n de f , mesurable par rapport à la tribu engendrée par (X_1, \dots, X_n) . Notons $F(x) = P(X_1 \leq x)$ la fonction de répartition de la loi de X_1 et soit \hat{F}_n la fonction de répartition empirique.

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

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

$$\lim \hat{F}_n(x) = F(x) \tag{2.8}$$

Donc \hat{F}_n peut être considéré comme un estimateur de F .

Mais comment faire pour estimer la densité f

Une des premières idées intuitives est de considérer pour $h > 0$ petit

$$\begin{aligned}
 \hat{f}_n(x) &= \frac{F(x+h) - F(x-h)}{2h} \\
 &= \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}_{x-h \leq X_i \leq x+h} \\
 &= \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}_{-h \leq X_i - x \leq h} \\
 &= \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}_{-1 \leq \frac{X_i - x}{h} \leq 1} \\
 \hat{f}_n(x) &= \frac{1}{nh} \sum_{i=1}^n K_0\left(\frac{x - X_i}{h}\right) \tag{2.9}
 \end{aligned}$$

Avec

$$\begin{aligned}
 K_0(s) &= \frac{1}{2} \mathbf{1}_{\{-1 \leq s \leq 1\}} \\
 &= \begin{cases} \frac{1}{2} & \text{si } x \in [-1, 1] \\ 0 & \text{sinon} \end{cases}
 \end{aligned}$$

Le biais et la variance de cet estimateur a été calculée par Rosenblatt [11] ,

$$\begin{aligned}
 \text{Biais}(\hat{f}_h^n(x)) &= \mathbb{E}(\hat{f}_h^n(x)) - f(x) \\
 &= \frac{1}{2h} \mathbb{E}(F_n(x+h) - F_n(x-h)) - f(x) \\
 &= \frac{1}{2h} \mathbb{E}(F(x+h) - F(x-h)) - f(x) \tag{2.10}
 \end{aligned}$$

$$\text{Var}(\hat{f}_h^n(x)) = \frac{F(x+h) - F(x-h)}{4nh^2} - \frac{(F(x+h) - F(x-h))^2}{4nh^2} \tag{2.11}$$

Cette variance tend vers zéro lorsque nh tend vers l'infini quand n tend vers l'infini. et par suite dans ce cas $\hat{f}_h^n(x)$ converge en moyenne quadratique vers $f(x)$. et par suite :

$$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{f}_h^n(x)) = f(x) \tag{2.12}$$

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{f}_h^n(x)) = 0 \tag{2.13}$$

Donc f_h^n défini par l'expression 2.9 est un estimateur consistant. Nous remarquons qu'il n'a pas le problème du choix d'origine a_0 comme le cas de l'histogramme mais il présente l'inconvénient d'être discontinu aux points $\{(x_i - h, x_i + h)\}_i \in \{1, \dots, n\}$. Ainsi une généralisation de cet estimateur a été introduite par (Parzen-Rozenblatt) est une méthode non paramétrique d'estimation de la densité d'une variable aléatoire. Cette méthode permet d'obtenir une densité continue et constitue en ce sens une généralisation de la méthode de l'histogramme.

En effet, la fonction indicatrice utilisée pour l'histogramme est ici remplacée par une fonction continue appelée (le noyau).

Définition 2.2.1 Soit $K : \mathbb{R} \rightarrow \mathbb{R}$ une fonction intégrable, positive et telle que :

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

K est appelée noyau. Soit $h := h_n > 0$ un paramètre de lissage (fenêtre) qui dépend de la taille de l'échantillon n . Pour toute $n \in \mathbf{N}$; l'estimateur à noyau de la densité f de la v.a X au point x_i , noté $\hat{f}_h^n(x)$ est donné par :

$$\hat{f}_h^K(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (2.14)$$

Généralement le noyau K est une fonction positive et bornée et vérifie les conditions suivantes :

- K est symétrique $\forall x : K(x) = K(-x)$
- $\int_{-\infty}^{+\infty} K(x) dx = 1$
- $\int_{-\infty}^{+\infty} xK(x) dx = 0$
- $\int_{-\infty}^{+\infty} x^2K(x) dx = \mu_2 < \infty$

Cet estimateur a été largement étudié par de nombreux auteurs, citons par exemple Wolverton et Wagner (1969), Roussas (2000, 2001) [6], Bosq et al. (1999) et Lu (2001). Exemple de noyaux K les plus utilisés dans l'estimation de la densité

- le noyau rectangulaire (Rosenblatt) :

$$K(u) = \frac{1}{2} \mathbf{1}_{[-1,1]}(u)$$

- le noyau triangulaire :

$$K(u) = (1 - |u|)\mathbf{1}_{[-1,1]}(u)$$

- le noyau d'Epanechnikov (parabolique) :

$$K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}_{[-1,1]}(u)$$

- le noyau gaussien :

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-u^2/2}$$

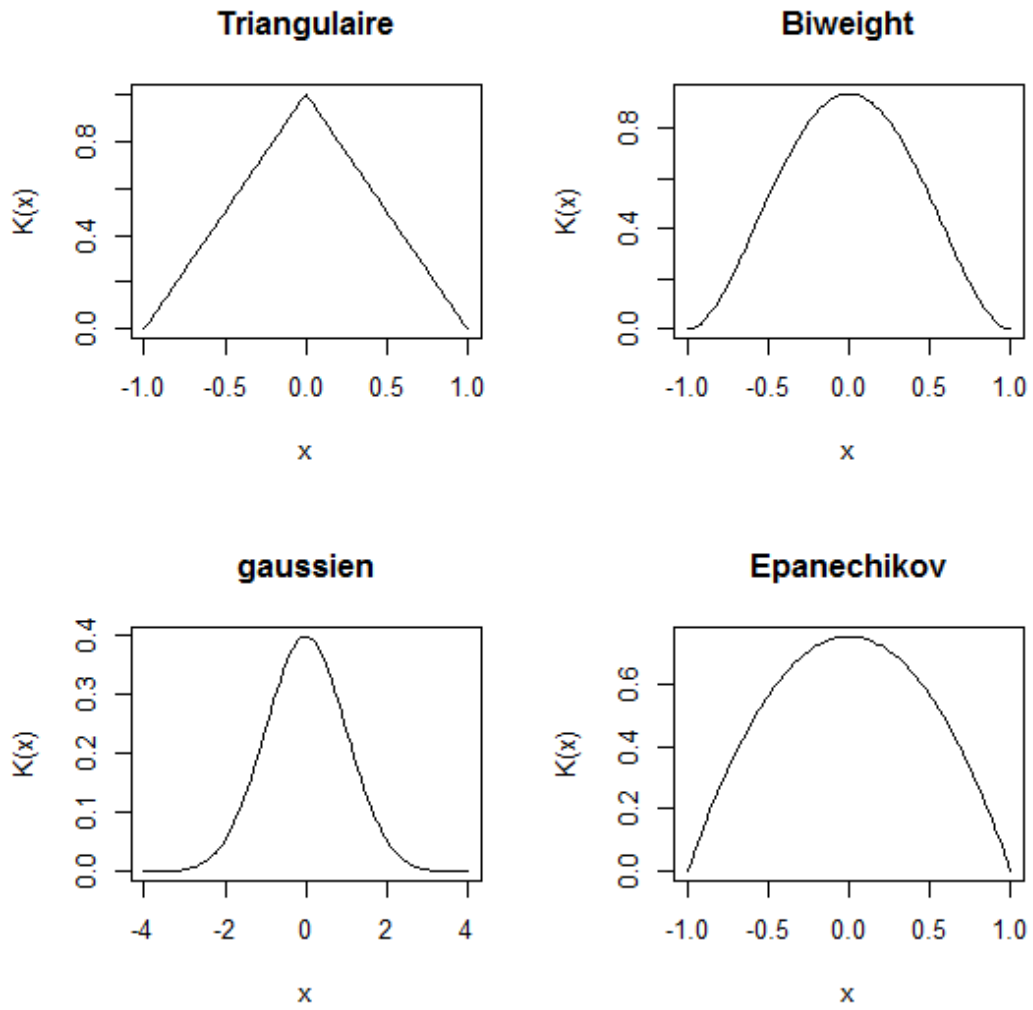


Figure 2.1 : Courbes sur les noyaux communs

Lemme 2.2.1 Si le noyau K est positive et $\int_{-\infty}^{+\infty} K(x)dx = 1$, alors $\hat{f}_h^K(x)$ est une densité de probabilité. De plus, \hat{f}_h^K est continue si K est continue.

Preuve. Si K est positif et continue alors $\hat{f}_h^K(x)$ doit être positif et continue, car c'est une somme des fonctions positifs est continues et plus de ça on a :

$$\begin{aligned} \int_{-\infty}^{+\infty} \hat{f}_h^K(x)dx &= \int_{-\infty}^{+\infty} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)dx \\ &= \frac{1}{nh} \int_{-\infty}^{+\infty} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (\text{posons } t = \frac{x - X_i}{h} \quad \text{alors } hdt = dx) \\ &= \frac{1}{n} \int_{-\infty}^{+\infty} \sum_{i=1}^n K(t)dt \\ &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(t)dt \\ &= \frac{n}{n} = 1 \end{aligned}$$

■ On voit donc que, tout comme l'estimateur par histogramme, l'estimateur à noyau est une densité de probabilité. Il a de plus l'avantage d'être continu à condition que K le soit, ce qui n'était pas le cas pour les histogrammes. Par conséquent, lorsqu'on estime une densité continue, il est naturel de s'attendre que l'estimateur à noyau soit meilleur que l'estimateur par histogramme.

2.2.1 Les propriétés de l'estimateur à noyau

Biais de l'estimateur

Le biais de l'estimateur $\hat{f}_h^K(x)$ 2.14 est définie par :

$$\text{biais}(\hat{f}_h^K(x)) = \frac{h^2 \mu_2(K) f''(x)}{2} + O(h^2) \quad (2.15)$$

Remarquons que Le terme de droite de l'équation 2.15 est différent de zéro, ceci signifie que $\hat{f}_h^K(x)$ est biaisé. Cependant le biais $\text{biais}(\hat{f}_h^K(x))$ converge vers zéro quand la (la fenêtre) h tend vers zéro.

Le biais de $\hat{f}_h^K(x)$ ne dépend pas directement du nombre d'observations n mais seulement de la fonction de noyau K et donc l'augmentation du nombre d'observations n ne suffit pas pour diminuer le *biais*($\hat{f}_h^K(x)$) .

Etude de biais

Tout d'abord, commençons par vérifier le non-alignement asymptotique, où l'estimation de la densité de noyau dépend de deux paramètres.

$$\begin{aligned} E[\hat{f}_h(x)] &= \frac{1}{n} \sum_{i=1}^n E[K_h(x - X_i)] \\ &= E[K_h(x - X)] \\ &= \int K_h(x - u) f(u) du \\ &= \int K(s) f(x + sh) ds. \end{aligned}$$

Si on laisse $h \rightarrow 0$, on voit que

$$E[\hat{f}_h(x)] \rightarrow f(x) \int K(s) ds = f(x), h \rightarrow 0$$

Nous allons calculer le biais

$$\begin{aligned} \text{Biais}[\hat{f}_h(x)] &= \int K(s) f(x + sh) ds - f(x) \\ &= \int K(s) \left[f(x) + sh f'(x) + \frac{h^2 s^2}{2} f''(x) + o(h^2) \right] ds - f(x) \\ &= f(x) + \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2) - f(x) \end{aligned}$$

donc le biais de estimation de noyau c' écrit sa forme :

$$\text{Biais}[\hat{f}_h(x)] = f(x) + \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2) - f(x) \quad h \rightarrow 0 \quad (2.16)$$

Variance de l'estimateur

La variance de l'estimateur $\hat{f}_h^K(x)$ 2.14 est définie par :

$$\text{Var}(\hat{f}_h^K(x)) = \frac{f(x) \int_{-\infty}^{+\infty} (K(u))^2 du}{nh} + O(nh)^{-1} \quad (2.17)$$

Cette variance tend vers zéro quand nh tend vers l'infini quand n tend vers l'infini.

La convergence en moyenne quadratique de l'estimateur $\hat{f}_h^K(x)$ a été établie par Bosq et Lecoutre [5].

Etude de variance

$$\begin{aligned} \text{Var}(\hat{f}_h(x)) &= n^{-2} \text{Var} \left(\sum_{i=1}^n K_h(x - X_i) \right) \\ &= n^{-2} \sum_{i=1}^n \text{Var}[K_h(x - X_i)] \\ &= n^{-1} \text{Var}[K_h(x - X)] \\ &= n^{-1} E[K_h^2(x - X)] - (E[K_h(x - X)])^2 \\ &= n^{-1} \left\{ h^{-2} \int K^2 \left(\frac{x - u}{h} \right) f(u) du - (f(x) + o(h))^2 \right\} \\ &= n^{-1} \left\{ h^{-1} \int K^2(s) f(x + sh) ds - (f(x) + o(h))^2 \right\} \end{aligned}$$

2.2.2 L'erreur quadratique moyenne

Calculons maintenant l'erreur quadratique moyenne de $\hat{f}_h^K(x)$. Nous avons, pour tout $x \in \mathbb{R}$:

$$MSE(\hat{f}_h^K(x)) = \frac{f(x) \int_{-\infty}^{+\infty} (K(u))^2 du}{nh} + \frac{h^4}{4} (\mu_2(K) f''(x))^2 + O(nh)^{-1} + O(h^4) \quad (2.18)$$

Calculons l'erreur quadratique moyenne intégrée $MISE(\hat{f}_h^K(x))$ par l'intégration de MSE sur \mathbb{R} . 2.18 on obtient :

$$MISE(\hat{f}_h^K) = \frac{R(K(u))}{nh} + \frac{h^4}{4} (\mu_2(K))^2 R(f'') + O(nh)^{-1} + O(h^4) \quad (2.19)$$

avec $R(w) = \int_{\mathbb{R}} (w(x))^2 dx$. Donc

$$A - MISE(\hat{f}_h^K) = \frac{R(K(u))}{nh} + \frac{h^4}{4} (\mu_2(K))^2 R(f'')$$

Pour calculer h_o optimale on résout l'équation $\frac{\partial(A-MISE)}{\partial h} = 0$:

$$\begin{aligned} \frac{\partial(A - MISE)}{\partial h} = 0 &= -\frac{R(K(u))}{nh^2} + h^3 (\mu_2(K))^2 R(f'') = 0 \\ \Leftrightarrow h^5 &= \left(\frac{R(k(u))}{(\mu_2(K))^2 R(f'')} \right) n^{-1} \\ \Leftrightarrow h_o &= \left(\frac{R(k(u))}{(\mu_2(K))^2 R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}} \end{aligned}$$

Pour $h = h_o$ on obtient la valeur optimale de (A-MISE) :

$$A - MISE^* = \frac{5}{4} (\mu_2(K) R(K)^2)^{\frac{2}{5}} (Rf'')^{\frac{1}{5}} n^{-\frac{4}{5}}$$

2.3 Choix de noyau optimal

Notre objectif est de choisir le noyau optimal qui minimise $A - MISE$.

Puisque la fonction f'' est inconnue et le paramètre de lissage a été déjà optimiser, alors pour minimiser $A - MISE(\hat{f}_h^K)$, il faut choisir le noyau K qui minimise la valeur de $(\mu_2(K))^{\frac{2}{5}} (R(K))^{\frac{4}{5}}$. Ce qui est équivalent à minimiser $R(K) = \int_{\mathbb{R}} (K(x))^2 dx$ avec les condi-

tions :

- $\int_{-\infty}^{+\infty} K(x)dx = 1$
- $\int_{-\infty}^{+\infty} x^2 K(x)dx = \mu_2(K) < \infty$

Hodges et Lehmann (1956) [?, ?] montraient que ce problème de minimisation est résolu en choisissant le noyau K défini par :

$$K_{ep}(u) = \frac{3}{4}(1 - u^2)\mathbf{1}_{[-1,1]}(u)$$

C'est le noyau d'Epanechnikov (parabolique) : Nous pouvons donc considérer l'efficacité d'un noyau K (notée $eff(K)$) quelconque symétrique en le comparant avec K_{ep} puisque ce dernier minimise le A-MISE si h est choisi de façon optimale ($h = h_{opt}$) .

On pose

$$eff(K) = \left[\frac{C(K_e)}{C(K)} \right]^{\frac{5}{4}}$$

$$eff(K) = \frac{3}{5\sqrt{5}} \frac{1}{\sqrt{\mu_2(K)}R(K)}$$

le tableau suivant présente les valeurs d'efficacité de quelques noyaux continus symétriques usuels.

Noyau	$K(u)$	eff(K)
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}_{[-1,1]}(u)$	1
Rectangulaire	$K(u) = \frac{1}{2}\mathbf{1}_{[-1,1]}(u)$	0,9295
Triangulaire	$K(u) = (1 - u)\mathbf{1}_{[-1,1]}(u)$	0,9859
Gaussien	$K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-u^2/2}$	0,9512

Remarque 2.3.1 *D'après le tableau on remarque que les valeurs d'efficacité des noyaux symétriques sont très proches de 1 et qu'il ya très peu de différence entre les différents noyaux par rapport à l'erreur quadratique moyenne intégrée.*

2.4 Choix paramètre de lissage

L'application scientifique de l'estimateur de densité de probabilité à noyau nécessite la spécification du paramètre de lissage. Pour choisir la densité la plus douce, la sélection du paramètre de lissage par l'œil prend beaucoup de temps. Il existe de nombreuses estimations. La densité nécessaire pour un problème particulier où l'utilisateur n'a pas de connaissance préalable de la structure des données et n'a pas le sentiment que le paramètre de lissage donne une estimation plus proche de la densité réelle dans le cas de l'utilisation d'estimateurs à noyau comme composants des procédures statistiques est grande, il faut donc choisir le paramètre de lissage, ce qui est généralement nécessaire. La sélection du paramètre de lissage est divisée en deux parties :

1- Il est composé de formules simples qui peuvent être calculées facilement, mais sans garantir que le calcul est proche d'un paramètre. Nous appellerons les limiteurs de paramètre de lissage rapide et simple et cette méthode qui utilise des données X_1, \dots, X_n pour produire la paramètre de lissage est appelée paramétrisation touchante.

2- Il est classé comme high-tech car les procédures de sélection dépendent d'arguments mathématiques plus complexes et nécessitent beaucoup d'efforts de calcul, mais la réponse est correcte et précise. Elle est motivée par visant à minimiser $MISE(f_h(x))$.

2.3.1 Méthode de Rule of thumb

Une approche très facile et naturel consiste à utiliser une famille \hat{A} au niveau de la distribution d'attribuer une valeur à $\int (f''(x))^2 dx$ terme de f'' dans l'expression h_{AMIS} pour la fenêtre idéale par exemple Si x_1, x_2, \dots, x_n . Est un échantillon i.i.d. d'une variable aléatoire, alors l'estimateur non-paramétrique par la méthode du noyau de la densité est :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (2.20)$$

qui régit le degré de lissage de l'estimation. Bien souvent, K est choisi comme la densité d'un fonction gaussienne standard (espérance nulle et variance unitaire) :

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}x^2} \quad (2.21)$$

Si on choisit f comme étant la distribution de loi normale de moyenne 0 et de variance $\sigma^2(N(0; \sigma^2))$ On aura alors :

$$\int (f''(x))^2 dx = \sigma^{-5} \int (\phi''(x))^2 dx = \frac{3}{8} \pi^{-\frac{1}{2}} \sigma^{-5} \simeq 0.212 \sigma^{-5} \quad (2.22)$$

De plus, si K est un noyau gaussien, alors la valeur pour le h_{opt} notée dans ce cas par h_{rot} est obtenue en substituant ce noyau et la valeur $R(f'')$ obtenue

$$\begin{aligned} h_{opt} &= (4\pi)^{-\frac{1}{4}} \left[\frac{3}{8} \pi^{-\frac{1}{2}} \sigma \right] n^{-\frac{1}{5}} \\ &= \left(\frac{4}{3} \right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} \\ &= 1.06 \sigma n^{-\frac{1}{5}}. \end{aligned} \quad (2.23)$$

Remarques

- Évidemment, h_{rot} fonction bien si la densité réelle est gaussienne.
- Forcément bien si la densité réelle n'est pas gaussienne, il ne fonctionne pas.

2.3.2 Cross validation

2.1 Validation croisée non biaisée

Cette méthode appelée Validation Croisée non Biaisée a été proposée par Rudemo [51] en 1982 et Bowman [8] en 1984. Le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de

$$\begin{aligned} UCV(h) &= \int_{\mathfrak{R}} |f_h(x) - f(x)|^2 dx - \int_{\mathfrak{R}} f^2(x) dx \\ &= \int_{\mathfrak{R}} f_h^2(x) dx - 2 \int_{\mathfrak{R}} f_h(x) f(x) dx. \end{aligned} \quad (2.24)$$

Puisque $\int_{\mathfrak{R}} f^2(x)dx$ ne dépend pas du paramètre de lissage h . On peut choisir le paramètre de lissage de façon à ce qu'il minimise un estimateur de :

$$\int_{\mathfrak{R}} f_h^2(x)dx - 2 \int_{\mathfrak{R}} f_h(x)f(x)dx \quad (2.25)$$

On veut premièrement trouver un estimateur de $\int_{\mathfrak{R}} f_h(x)f(x)dx$. Remarquons que

$$\int_{\mathfrak{R}} f_h(x)f(x)dx = E(f_h(x))$$

L'estimateur empirique de $\int_{\mathfrak{R}} f_h(x)f(x)dx$, est alors :

$$\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i).$$

Le critère à optimiser est alors :

$$UCV(h) = \int_{\mathfrak{R}} f_h^2(x)dx - \frac{2}{n} \sum_{i=1}^n f_{h,i}(x_i)$$

Où $f_{h,i}(x_i) = \frac{1}{(n-1)h} \sum_{j=1, j \neq i}^n K\left(\frac{x_i - x_j}{h}\right)$ est l'estimation de la densité construit

à partir de l'ensemble de points sauf le point x_i

Montrons maintenant que $UCV(h)$ est un estimateur sans biais de $MISE(h) - R(f)$.

on a :

$$MISE(h) - R(f) = E \left[\int_{\mathfrak{R}} f_h^2(x)dx - 2 \int_{\mathfrak{R}} f_h(x)f(x)dx \right] \quad (2.26)$$

Il suffit de montrer que $\int_{\mathfrak{R}} f_h^2(x)dx$ et $\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i)$ sont des estimateurs

sans biais de $E[\int_{\mathfrak{R}} f_h^2(x)dx]$ et $E[\int_{\mathfrak{R}} f_h(x)f(x)dx]$ respectivement. Or $E[\int_{\mathfrak{R}} f_h^2(x)dx]$

admet l'estimateur sans biais trivial $\int f_h^2 dx$. Il reste donc à montrer que

$\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i)$ est un estimateur sans biais de $E[\int f_h(x)f(x)dx]$

On a d'une part,

$$\begin{aligned} E \left[\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i) \right] &= E \left[\frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \int K\left(\frac{x_i - x_j}{h}\right) \right] \\ &= E \left[\frac{1}{n(n-1)h} \sum_{j \neq 1}^n \int K\left(\frac{x_i - x_j}{h}\right) f(x) dx \right] \\ &= \frac{1}{n} \int f(z) \int K\left(\frac{x - z}{h}\right) f(z) dx dz. \end{aligned}$$

D'autre part,

$$\begin{aligned} E \left[\int f_h(x) f(x) \right] &= E \left[\frac{1}{nh} \sum_{i=1}^n \int K\left(\frac{x_i - x_j}{h}\right) f(x) dx \right] \\ &= \frac{1}{n} \int f(z) \int K\left(\frac{x - z}{h}\right) f(x) dx dz \end{aligned}$$

Ce qui implique que :

$$E \left[\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i) \right] = E \left[\int f_h(x) f(x) \right]$$

Finalement, un estimateur sans biais de $MISE(h) - R(f)$ est donnée donc par $UCV(h)$.

En utilisant l'équation précédente, le critère $UCV(h)$ devient :

$$UCV(h) = \frac{R(K)}{nh} + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[\int \frac{1}{n^2 h^2} K\left(\frac{x-x_i}{h}\right) K\left(\frac{x-x_j}{h}\right) dx - \frac{2}{n(n-1)h} K\left(\frac{x_i-x_j}{h}\right) \right] \quad (2.27)$$

Nous noterons h_{UCV} l'estimateur de h qui minimise $UCV(h)$.

Soit x_1, x_2, \dots, x_n un échantillon i.i.d issu d'une variable aléatoire X de

fonction de densité f .

Utilisant le noyau gaussien on obtient :

$$UCV(h) = \frac{1}{2n^2 h \sqrt{\pi}} \left(n + 2 \sum_{i=1}^n \sum_{j=1, j \neq i}^n \exp\left(-\left(\frac{x_i - x_j}{2h}\right)^2\right) \right) - \frac{2}{\sqrt{2\pi} n(n-1)h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \exp\left(-\frac{(x_i - x_j)^2}{2h^2}\right) \quad (2.28)$$

2.2 Validation croisée biaisée

Un critère de validation croisée biaisée, a été introduit par Scott et Terrell [54] en 1987 pour remédier aux problèmes de validation croisée non biaisée. Il s'agit d'introduire un biais dans le UCV fin de réduire sa variance. L'Erreur Quadratique Intégrée Moyenne Asymptotique s'écrit sous la forme :

$$AMISE = \frac{h^4}{4} \sigma_K^4 R(f'') + \frac{R(K)}{nh} \quad (2.29)$$

Le paramètre de lissage basé sur la méthode de validation croisée biaisée est la valeur h qui minimise un estimateur du $AMISE$. On peut estimer le $AMISE$ si l'on estime $R(f'')$

Un estimateur naturel de ce terme est donné par $R(f_h'')$ où f_h est l'estimateur de la densité qui utilise la méthode du noyau. Scott et Terrell [54]

Supposant que le noyau k satisfait aux conditions suivantes :

$$\int K''(u)du = 0, \mu_1(K'') = \int uK''(u) = 0, \mu_2(K'') = \int u^2K''(u) = 2 \quad (2.30)$$

On obtient le développement asymptotique :

$$E[R(f_h'')] = R(f'') + \frac{R(K'')}{nh^5} + O(h^2). \quad (2.31)$$

Proposition 1 : (Scott et Trell) [54]

Soit x_1, x_2, \dots, x_n un n-échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f . Pour un noyau K on obtient :

$$BCV(h) = \frac{R(K)}{nh} + h^4 \frac{\mu_2^2(K)}{4n^2} \sum_i \sum_{j, j \neq i} K_h^{(2)} K_h^{(2)}(x_i - x_j). \quad (2.32)$$

Proposition 2

Soit x_1, x_2, \dots, x_n un n-échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f . En choisissant le noyau gaussien on obtient :

$$BCV(h) = \frac{1}{2nh\sqrt{\pi}} + \frac{1}{64n^2\sqrt{\pi}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[\left(\frac{x_i - x_j}{h}\right)^4 - 12\left(\frac{x_i - x_j}{h}\right)^2 + 12 \right] \exp \left[-\frac{(x_i - x_j)^2}{4h^2} \right]. \quad (2.33)$$

Cette méthode nous donne plusieurs minimums locaux pour la fonctionnelle cible à minimiser. Cependant, d'après plusieurs simulations, les auteurs proposent de choisir la valeur inférieure parmi les minimums locaux.

Conclusion

Dans cette chapitre, nous concluons que le paramètre de lissage h est très sensible, nous devons donc choisir avec soin h , en utilisant l'une des méthodes mentionnées précédemment.

Chapitre 3

Régression non paramétrique

le modèle régression consiste à estimer les relation entre l'une des variables dépendant et une ou plusieurs variables indépendantes sous une forme algébrique en utilisant les données disponibles sur les variables en formulant une équation ou une courbe qui explique cette relation. La relation est connue sous le nom d'équation de régression, et régression est défini par :

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

3.1 Estimation de la fonction de régression par noyau

3.1.1 La méthode du noyau

Notre but dans ce chapitre est de construire l'estimateur à noyau de la fonction de régression par la méthode Nadaraya-Watson et d'étudier ses propriétés .

La procédure statistique générale la plus largement utilisée est la régression (linéaire). Les modèles de régression sont des outils puissants pour modéliser une variable cible y en fonction d'un ensemble de prédicteurs x , permettant de prédire les valeurs de y et de construire des tests et des estimations d'intervalle pour les prédictions et les paramètres. Les modèles de régression sont également susceptibles de rencontrer les mêmes problèmes que tout autre modèle paramétrique.

Considérons le modèle de régression linéaire simple,

$$y_i = b_0 + b_1 x_i + \epsilon_i \quad i = 1, \dots, n \quad (3.1)$$

Une alternative plus générale à 3.1 est le modèle de régression non paramétrique avec les erreurs ϵ généralement considérées comme indépendantes et identiquement (approximativement gaussienne) avec une moyenne nulle et une variance σ^2 . Si ce modèle est une bonne représentation de la réalité, les estimations des moindres carrés de b_i peuvent être calculées.

Mais que faire si le modèle linéaire 3.1 n'est pas approprié ?

L'ajustement d'un modèle linéaire à une relation non linéaire peut donner des résultats mauvaises, impliquant un degré de certitude qui n'est pas réaliste. Une alternative plus générale à 3.1 est le modèle de régression non paramétrique.

$$y_i = r(x_i) + \epsilon_i \quad i = 1, \dots, n \quad (3.2)$$

La fonction de régression $r(x)$ est l'espérance conditionnelle $r(x) = E(Y/X = x)$, avec $E(\epsilon/X = x) = 0$, et $V(Y/X = x) = \sigma^2(x)$ pas nécessairement constant.

Le grand problème est l'estimation de la fonction r à partir de les n observations $(X_i; Y_i) \quad i = 1, \dots, n$ qui suivent la même loi que $(X; Y)$.

Supposons que $(X; Y)$ a une densité f sur \mathbb{R} et que :

$$f_X : x \mapsto f_X(x) = \int f(x, y) dy > 0 \quad (3.3)$$

la fonction de densité de X

Donc pour tout $x \in \mathbb{R}$:

$$\begin{aligned}
 r(x) &= E(Y/X = x) \\
 &= \int y f(y/x) dy \\
 &= \int y \frac{f(x, y)}{f_X(x)} dy
 \end{aligned} \tag{3.4}$$

Un estimateur par noyau de $f(x, y)$ est : $\hat{f}_n(x, y) = \frac{1}{nh_x h_y} \sum_1^n K_x\left(\frac{x-x_i}{h_x}\right) K_y\left(\frac{y-y_i}{h_y}\right)$ tandis qu'un estimateur par noyau de $f_X(x)$ est : $\hat{f}_X(x) = \frac{1}{nh_x} \sum_1^n K_x\left(\frac{x-x_i}{h_x}\right)$ En substituant dans 3.4 et en notant que :

$$\int K_y(u) du = 1 \text{ et que } \int u K_y(u) du = 0$$

On obtient l'estimateur à noyau de Nadaraya-Watson,

$$r_{nw} = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}$$

3.1.2 Définition

Soit $K : \mathfrak{R} \rightarrow \mathfrak{R}$ un noyau ($K \geq 0$, et $\int K = 1$), $h > 0$ une fenêtre, l'estimateur de Nadaraya-Watson de régression de Y sur X est donné par

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

où

$$W_{n,i}(x) = \frac{K\left(\frac{X_i-x}{h}\right)}{\sum_{j=1}^n K\left(\frac{X_j-x}{h}\right)} \mathbf{1}_{\sum_{j=1}^n K\left(\frac{X_j-x}{h}\right) \neq 0}$$

C'est l'estimateur à noyau introduit par Nadaraya-Watson (Nadaraya, 1964 et Watson, 1964).

Cet estimateur est parfois appelé estimateur à noyau d'évaluation) [10]

La construction de cet estimateur dépend de deux paramètres, le paramètre de lissage h

dont le choix est crucial pour obtenir de bonnes propriétés asymptotiques et le noyau K dont on ne peut pas négliger le rôle pour la réduction du biais.

2.1 Proposition

Si K est un noyau d'ordre 1 alors $\forall x \in \mathfrak{R}$

$$\hat{r}_n(x) = \begin{cases} \frac{\sum_{i=1}^n Y_i K(\frac{X_i-x}{h})}{\sum_{i=1}^n K(\frac{X_i-x}{h})} \neq 0 \\ 0 \text{ sinon} \end{cases}$$

Démonstration

$\hat{r}_{n,X}(x) = 0$ est équivalent à $\sum_{i=1}^n K(\frac{X_i-x}{h}) = 0$.

Supposons donc que $\sum_{i=1}^n K(\frac{X_i-x}{h}) \neq 0$. Alors

$$\begin{aligned} \hat{r}_n(x) &= \int \frac{y \hat{f}_n(x, y)}{\hat{f}_n(x)} dy \\ &= \frac{1}{\hat{f}_n(x)} \int y \frac{1}{nh^2} \sum_{i=1}^n K(\frac{X_i-x}{h}) K(\frac{Y_i-y}{h}) dy \\ &= \frac{nh}{\sum_{i=1}^n K(\frac{X_i-x}{h})} \frac{1}{nh^2} \sum_{i=1}^n K(\frac{X_i-x}{h}) \int y K(\frac{Y_i-y}{h}) dy \\ &= \frac{1}{\sum_{i=1}^n K(\frac{X_i-x}{h})} \sum_{i=1}^n K(\frac{X_i-x}{h}) \frac{1}{h} \int y K(\frac{Y_i-y}{h}) dy \\ &= \frac{1}{\sum_{i=1}^n K(\frac{X_i-x}{h})} \sum_{i=1}^n K(\frac{X_i-x}{h}) Y_i \end{aligned} \tag{3.5}$$

Pour la dernière ligne, on a utilisé le fait que

$$\frac{1}{h} \int y K(\frac{Y_i-y}{h}) dy = \frac{1}{h} \int (Y_i - uh) K(u) h du = Y_i \int K(u) du - h \int u K(u) du = Y_i$$

3.1.3 Propriétés asymptotiques de Nadraya Watson

L'étude asymptotique du biais de la variance de l'estimateur de Nadraya-Watson. En vu de la décomposition suivante :

$$E(\hat{r}_n(x, h_n) - r(x))^2 = Var(\hat{r}_n(x, h_n) + (E\hat{r}_n(x, h_n) - r(x))^2 \quad (3.6)$$

1-Etude asymptotique de la variance

Proposition

Sous les hypothèses de la proposition(3.2) etsi $EY^2 < \infty$, alors en chaque point de continuité des fonction $r(x)$, $f(x)$ et $\sigma^2(x) = Var(Y/X = x)$

On a

$$Var[r_n(x, h_n)] = \frac{1}{nh_n} \left\{ \frac{\sigma^2(x)}{f(x)} \int_{\mathfrak{R}} K^2(u) du \right\} (o(1) + 1). \quad (3.7)$$

où $f(x) > 0$.

Preuve

Soit la fonction $\psi(x) = \int_{\mathfrak{R}} y^2 f(x, y) dy$, en se basant sur le Théorème de Bochner on a.

Théorème de Bonchner :

Soit $K : (\mathfrak{R}^m, \beta^m) \rightarrow (\mathfrak{R}, \beta)$ une fonction mesurable, où β^p est la tribu borélienne de \mathfrak{R}^p , vérifiant :

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

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

et

$\|z\|^m |K(z)| \rightarrow 0$ quand $\|z\| \rightarrow \infty$. Par ailleurs, soit

$g : (\mathfrak{R}^m, \beta^m) \rightarrow (\mathfrak{R}, \beta)$ une fonction telle que

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

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

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

Si g est uniformément continue alors la convergence ci dessus est uniforme.

$$\begin{aligned}
 \text{var}(\phi(x, h_n)) &= \frac{1}{nh_n^2} \left\{ E \left[Y^2 K^2 \left(\frac{x - X}{h_n} \right) \right] - \left[E Y K \left(\frac{x - X}{h_n} \right) \right]^2 \right\} \\
 &= \frac{1}{nh_n} \left\{ \int_{\mathfrak{R}} K^2(u) \psi(x - h_n u) du - h_n \left(\int_{\mathfrak{R}} K(u) f(x - u h_n) r(x - h_n u) du \right)^2 \right\} \\
 &= \frac{1}{nh_n} \psi \int_{\mathfrak{R}} K^2(u) du (1 + o(1)),
 \end{aligned} \tag{3.8}$$

$$E[f_n(x, h_n) - E f_n(x, h_n) \phi_n(x, h_n) - E(\phi_n(x, h_n))] = \frac{1}{nh_n \phi(x) \int_{\mathfrak{R}} K^2(u) du} (1 + o(1))$$

et

$$\text{var} f_n(x, h_n) = \frac{1}{nh_n} f(x) \int_{\mathfrak{R}} K^2(u) du (1 + o(1)).$$

Posons

$$B_n = \begin{pmatrix} f_n(x, h_n) \\ \phi_n(x, h_n) \end{pmatrix}$$

et

$$A(x) = \begin{pmatrix} -r(x) \\ \frac{1}{|f(x)|^2} \end{pmatrix}.$$

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

$$\Sigma := \frac{1}{nh_n} \begin{pmatrix} f(x) & \phi(x) \\ \phi(x) & \psi(x) \end{pmatrix} \int_{\mathfrak{R}} K^2(u) du (1 + o(1)).$$

En remarquant, que

$$\begin{aligned}
 \text{var} r_n(x, h_n) &= A \Sigma A^t \\
 &= \frac{1}{nh_n} \left(\frac{\psi(x)}{|f(x)|^2} - \frac{(\phi(x))^2}{|f(x)|^3} \right) \int_{\mathfrak{R}} K^2(u) du (1 + o(1)),
 \end{aligned}$$

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

$$varr_n(x, h_n) = \frac{1}{nh_n} \left\{ \frac{\sigma^2(x)}{f(x)} \int_{\mathbb{R}} K^2(u) du \right\} (o(1) + 1).$$

2- Etude asymptotique du biais

3.4.2 Proposition

Sous les hypothèses de la proposition (3.4.1) et

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

$$Er_n(x, h_n) = \frac{E[\Phi_n(x, h_n)]}{E[f_n(x, h_n)]} + O\left(\frac{1}{nh_n}\right).$$

b) Si $EY^2 < \infty, nh^2 \rightarrow \infty$, alors

$$Er_n(x, h_n) = \frac{E[\Phi_n(x, h_n)]}{E[f_n(x, h_n)]} + O\left(\frac{1}{\sqrt{nh_n}}\right)$$

Preuve :

En multipliant les deux expressions de l'identité suivante

$$\frac{1}{f_n(x, h_n)} = \frac{1}{Ef_n(x, h_n)} - \frac{f_n(x, h_n) - Ef_n(x, h_n)}{[Ef_n(x, h_n)]^2} + \frac{[f_n(x, h_n) - Ef_n(x, h_n)]^2}{f_n(x, h_n)[Ef_n(x, h_n)]^2}$$

par $\Phi_n(x)$ et en passant à l'espérance nous obtenons

$$\begin{aligned} Er_n(x, h_n) &= \frac{E\Phi_n(x)}{Ef_n(x, h_n)} - (Ef_n(x, h_n))^{-2} E[(\Phi_n(x, h_n) \\ &\quad - E[\phi_n(x, h_n)])(f_n(x, h_n) - Ef_n(x, h_n)) \\ &\quad + E(f_n(x, h_n))^{-1} (Ef_n(x, h_n))^{-2} \phi_n(x, h_n) [f_n(x, h_n) - Ef_n(x, h_n)]^2] \\ &= \frac{E\phi_n(x, h_n)}{Ef_n(x, h_n)} + [c_n^1(x) + c_n^2(x)] (Ef_n(x, h_n))^{-2} \end{aligned}$$

où

$$c_n^{(1)}(x) = E[(\phi_n(x, h_n) - E[\phi_n(x, h_n)])(f_n(x, h_n) - E[f_n(x, h_n)])],$$

$$c_n^{(2)}(x) = E(f_n(x, h_n)^{-1} \Phi_n(x, h_n) [f_n(x, h_n) - E f_n(x, h_n)]^2).$$

Comme $var[f_n(x, h_n)] \sim \frac{1}{nh_n} f(x) \int_{\mathfrak{R}} K^2(t) dt$ (cf la preuve de la proposition (3.2) et

$var[\Phi_n(x, h_n)] \sim \frac{1}{nh_n} \int_{\mathfrak{R}} y^2 f(x, y) dy \int_{\mathfrak{R}} K^2(t) dt$ (cf la proposition précédent), alors

$$|c_n^1(x)| \leq (var\phi(x, h_n))^{\frac{1}{2}} (var f_n(x, h_n))^{\frac{1}{2}} = O\left(\frac{1}{nh_n}\right). \quad (3.9)$$

Pour montrer le cas b :

$$\begin{aligned} |c_n^2(x)| &\leq E \max_{1 \leq i \leq n} |Y_i| [E[f_n(x, h_n) - E(x, h_n)]^2], \\ &\leq \left[\sum_{i=1}^n E Y_i^2 \right]^{\frac{1}{2}} [E[f_n(x, h_n) - E[f_n(x, h_n)]]^4]^{\frac{1}{2}} \\ &= \sqrt{n} (E Y^2)^{\frac{1}{2}} O\left(\frac{1}{nh_n}\right) \\ &= O\left(\frac{1}{\sqrt{nh_n}}\right) \end{aligned}$$

3.1.4 Statistiques de l'estimateur Nadaraya-Watson

Lorsque nous voulons calculer l'espérance et la variance, nous rencontrons des difficultés au numérateur et au dénominateur Nadraya Watson. Nous allons donc analyser le numérateur et le dénominateur séparément. Défini par

$$r(x) = \int y f(x, y) dy = m(x) f(x),$$

$$\hat{r}_h(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i$$

L'estimation de la courbe de régression est donc donnée par :

$$\hat{m}_h(x) = \frac{\hat{r}_h(x)}{\hat{f}_h(x)}.$$

Tout d'abord, nous étudierons les numérateurs pour l'espérance et la variance :

$$\begin{aligned} E[\hat{r}_h(x)] &= E[n^{-1} \sum_{i=1} K_h(x - X_i) Y_i] \\ &= E[K_h(x - X) Y] \\ &= \int \int y K_h(x - u) f(y \setminus u) f(u) dy du \\ &= \int K_h(x - u) f(u) \left(\int y f(y \setminus u) dy \right) du \\ &= \int K_h(x - u) f(u) (E[Y \setminus X = u]) du \\ &= \int K_h(x - u) f(u) m(u) du \\ &= \int K_h(x - u) r(u) du \quad (3.3.1) \end{aligned}$$

Similaire à l'estimation de la densité avec des noyaux, nous notons que si

$$E[\hat{r}_h(x)] = r(x) + \frac{h^2}{2} r'' \mu_2 + o(h^2) \quad (h \rightarrow 0)$$

Par conséquent, $r_h(x)$ est asymptotiquement sans biais pour $h \rightarrow 0$. Pour calculer la variance $r_h(x)$ soit $s^2(x) = E[Y^2 \mid X = x]$, Ainsi

$$\begin{aligned}
 \text{Var}[\hat{r}_h(x)] &= \text{Var}\left[n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i\right] \\
 &= n^{-1} \text{Var}[K_h(x - X) Y] \\
 &= n^{-1} \left\{ \int K_h^2(x - u) s^2 f(u) du - \left(\int K_h(x - u) r(u) du \right)^2 \right\} \\
 &= n^{-1} h^{-1} \int K^2(u) s^2(x + uh) f(x + uh) du + o((nh)^{-1}) \\
 &= (nh)^{-1} f(x) s^2(x) \|K\|_2^2 + o((nh)^{-1}) \quad (nh \rightarrow \infty) \quad (3.3.2)
 \end{aligned}$$

Lorsque nous combinons (3.3.1) et (3.3.2), nous obtenons la formule du Erreur quadratique moyenne de $\hat{r}_h(x)$

$$\text{MSE}[\hat{r}_h(x)] = \frac{1}{nh} f(x) s^2(x) \|K\|_2^2 + \frac{h^4}{4} (r''(x) \mu_2(K))^2 + o(h^4) + o((nh)^{-1})$$

$$(h \rightarrow 0, nh \rightarrow \infty) \quad (3.3.3)$$

Donc, si on laisse $h \rightarrow 0$ tel que $nh \rightarrow \infty$, on a

$$\text{MSE}[\hat{r}_h(x)] \rightarrow 0$$

Par conséquent, l'estimation est cohérente

$$\hat{r}_h(x) \xrightarrow{p} r(x)$$

Le dénominateur de $\hat{m}_h(x)$, l'estimation de la densité du noyau $h(x)$, est

également pour les mêmes asymptotiques de h cohérentes.

$$\hat{m}_h(x) = \frac{\hat{r}_h(X)}{f_h(x)} \xrightarrow{p} \frac{r(x)}{f(x)} = m(x) \quad (h \rightarrow 0, nh \rightarrow \infty).$$

Convergence de l'estimateur Nadaraya-Watson :

$\hat{m}_h(x)$ est une estimation qui Converge de la courbe de régression $m(x)$, si $h \rightarrow 0, nh \rightarrow \infty$

Conclusion

Dans cette chapitre , nous concluons que le théorème Nadraya Watson est important pour estimer la fonction de régression dans l'estimation non paramétrique.

Conclusion

Dans ce mémoire, on a travaillé sur la méthode d'estimation à noyau, qui permettant de construire un estimateur non paramétrique de la fonction de densité et la fonction régression. Cette méthode d'estimation est simple et peut être très utile pour étudier des phénomènes lorsqu'on désire comprendre et analyser la relation qui existe entre deux variables à l'absence d'hypothèse ou information sur les caractéristiques de la distribution de base de la population. A travers les résultats obtenus, nous concluons que : le choix de noyau K a une influence faible sur la qualité de l'estimateur, par contre le paramètre de lissage h a un grand influence, et dont le choix est très intéressant.

Bibliographie

- [1] Bismut, J. M. (1973). Conjugate convex functions in optimal stochastic control. *Journal of Mathematical Analysis and applications*, 44(2), 384-404
- [2] Briand, P. (2001). *équations Différentielles Stochastiques Rétrogrades*. Mars
- [3] Chala, A, Khalout, R (2018). A risk-sensitive stochastic maximum principle for fully coupled forward-backward stochastic differential equations with applications. *Asian J Control*.
- [4] Chala, Adel . (2013). *Contribution à L'étude Des Controles optimales stochastique*. université de biskra. –+
- [5] D. Bosq and J.Lecoutre., M. (1987). *Théorie de l'estimation fonctionnelle*. Paris *Economica* .
- [6] D. Roussas, G. G.(2000) *Asymptotic normality of the kernel estimate of a probability density function under association*. *Statist. Probab. Lett. Statistics* .Cambridge.
- [7] Härdle, W. (1990) *Applied Nonparametric Regression*. Cambridge Univ. Press.
- [8] Hodges, J., and Lehmann, E .(1956). The efficiency of some nonparametric competitors of the t-test. *Annals Mathematics statistzcs* 27(1) .
- [9] J.Lecoutre. (1982) *Contribution à l'estimation non paramétrique de la régression*. PhD thesis, Université de Pierre et Marie Curie-ParisVI-France.
- [10] J.S.Simonoff.,(1996) *Smoothing Methods in Statistics*. Springer-Verlag .
- [11] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*.
- [12] Scott, D. W. (1992) *Multivariate Density Estimation :Theory, Practice, and Visualization*.Wiley Interscience, New York, .

Annexe : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont ex-

MSE Mean Squared Error en(Anglais) et (la française) Erreur Quadrati

$MISE$ Mean Integred Squared Error

$A - MISE$ Asymptotique-Mean Integred Squared Error

$p.s$ presque sur

UCV Un Validation Croisée

pliquées ci-dessous : reg regessgramme

$E(X)$ l'espérance de Mathématique

P propabilité

tq Telle que

$i.e$ C'est à dire

(Ω, A, P) espace probabilité

Résumé

Dans ce travail, nous avons utilisé deux méthodes, la méthode des graphes et la méthode du noyau, pour trouver la fonction de densité dans l'estimation non paramétrique. D'autre part, nous avons étudié la fonction de régression en utilisant la méthode complète de la théorie Nadraya Watson.

Abstract

In this work, we used two methods, the graph method and the kernel method, to find the density function in the nonparametric estimation. On the other hand, we studied the regression function by using the comprehensive kernel method of the theory Nadraya Watson.

ملخص

في هذا العمل استخدمنا طريقتين طريقة الرسم البياني والنواة لايجاد دالة الكثافة في التقدير اللامعلمي ومن جهة اخرى دراسنا دالة الانحدار وذلك بالاستعمال طريقة نواة الشاملة لنظرية ندرايا واتسون .