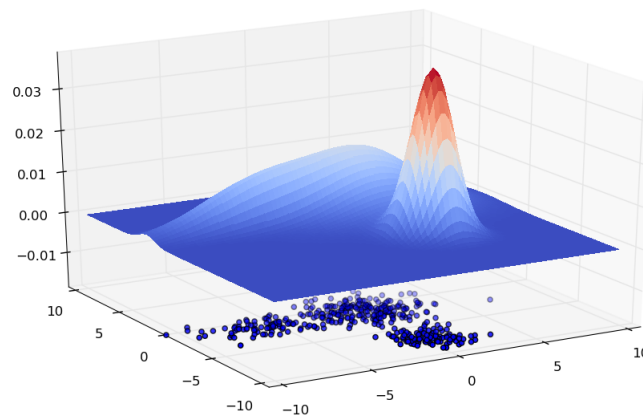


Projet de TER de substitution

Introduction à la statistique non paramétrique : Régression non paramétrique par noyau



Réalisé par :
BENIDIRI Lahlou

Encadreur :
Ph.Frédéric Richard

Remerciements

En premier lieu, je voudrais remercier vivement mes parents pour leurs soutien permanent, ainsi que tous mes proches.

Je remercie en particulier le professeur **Ph.Richard Frédéric**, mon encadreur pour son aide, sa disponibilité, sa rigueur scientifique et son sens d'écoute et d'échange.

Tout le corps professoral du M1 MAS qui nous a fait bénéficier d'une formation pluridisciplinaire de très haut niveau et très adapté aux réalités de la Statistique.

Table des matières

1	Introduction	4
2	Estimation non paramétrique de la densité de probabilité :(5)	5
2.1	Critère d'erreur :(2)	6
2.2	Quelques définitions :(2)	6
2.3	Estimation par la méthode d'histogramme :	7
2.4	Propriétés	11
2.5	Choix du paramètre	12
2.6	Conclusion	13
2.7	Estimation par la méthode de noyau :	13
2.8	Noyaux usuels(5)	15
2.9	Définitions :(7)	15
2.10	Propriétés de l'estimateurs à noyau, Espérance, variance de l'estimateur	16
2.11	Comportement asymptotique du biais et de la variance(5)	17
2.11.1	L'erreur quadratique moyenne intégrée $MISE(\hat{f}_h(x))$:(5)	18
2.11.2	L'erreur quadratique moyenne intégrée asymptotique $AMISE(\hat{f}_h(x))$:.	18
2.12	Comportement asymptotique et vitesse de convergence de l'estimateur à noyau	18
2.13	choix de paramètre de lissage h	19
2.14	Conclusion	21
3	Régression non paramétrique par la méthode de noyau(11)	22
3.1	Présentation générale des fonctions de lissage	22
3.2	L'estimation de nadaraya watson[NW](4)	22
3.3	Propriété de l'estimateur	23
3.3.1	Optimalité asymptotique(2)	23
3.4	Conclusion	24
4	Application	25
4.1	Le noyau en tant que fonction de pondération (9)	26
4.2	Conclusion	30
4.2.1	Code-R de la régression du noyau(9)	31
5	Conclusion générale	32

6	Abstract	33
7	Références	34

1 Introduction

La régression non paramétrique est une forme d'analyse de la régression dans lequel le prédicteur, ou fonction d'estimation, ne prend pas de forme prédéterminée, mais est construit selon les informations provenant des données.

Cette notion exige des tailles d'échantillons plus importantes que celles de la régression basée sur des modèles paramétriques parce que les données doivent fournir la structure du modèle ainsi que les estimations du modèle.

Dans le cadre de notre projet de substitution, nous allons consacrer tout un chapitre sur " Estimation non paramétrique de la densité de probabilité". En s'appuyant sur deux méthodes principales celle d'histogramme et celle de noyau.

La deuxième section comportera un cours sur "la régression par noyau" dans le cadre unidimensionnel. Ceci nous conduira à traiter un exemple illustratif sur l'application de cette notion dont un but prédictif.

le dernier chapitre consistera à traiter des données réelles avec la régression non paramétrique par noyau, dans un cadre d'une application qui serve à prédire le débit d'une rivière à partir d'un bassin versant afin de conclure avec une conclusion générale.

2 Estimation non paramétrique de la densité de probabilité :(5)

Dans nombreuses application la densité f est inconnue et on dispose d'un n -échantillon i.i.d X_1, \dots, X_n issu d'une v.a X admettant f comme densité de probabilité, le problème de statisticien consiste à utiliser cet échantillon pour construire un estimateur qui soit le plus proche possible de f .

De nombreuses estimations ont été proposés(5) :

- L'estimation par la méthode d'histogramme.
- L'estimation par la méthode de noyau.
- L'estimation par les séries orthogonales.

La figure suivante montre un exemple dans lequel les observations (représentées par les points bleus dans le plan horizontal) sont issues de R^2 et la densité qui a généré ces points est la surface courbe. Le sommet (en rouge) sur cette surface correspond à la région du plan horizontal où les observations (les points bleus) sont les plus denses.

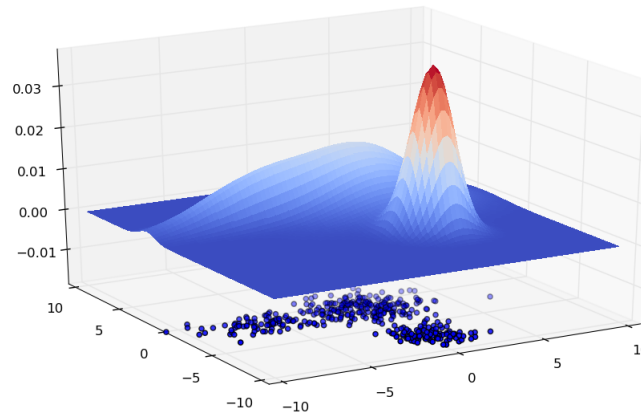


FIGURE 1 – Exemple de données de R^2 et de densité estimée

Nous examinons dans la suite les deux premières méthodes non paramétriques mentionnées, l'estimation par histogramme et l'estimation par noyaux.

2.1 Critère d'erreur :(2)

Pour mesurer les performances théoriques des estimateurs et identifier le meilleur. Il est nécessaire de spécifier un critère d'erreur. Nous considérons la densité f et son estimateur .

a) L'erreur quadratique intégrée (ISE) :

$$\begin{aligned} ISE(\hat{f}_n(x)) &= \int (f(x) - \hat{f}_n(x))^2 dx \\ &= \int f(x)^2 dx - 2 \int f(x) \hat{f}_n(x) dx + \int \hat{f}_n^2(x) dx \end{aligned}$$

b) L'erreur quadratique moyenne (MSE) :

$$\begin{aligned} MSE(\hat{f}_n(x)) &= E[(f_n(x) - \hat{f}_n(x))] \\ &= Var \hat{f}_n(x) + biais^2(\hat{f}_n(x)). \end{aligned}$$

c) L'erreur quadratique moyenne intégrée. (MISE) :

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

2.2 Quelques définitions :(2)

Définition 1 : On dit que l'estimateur est sans biais si :

$$E[\hat{f}_n] = f_n$$

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

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

Définition 3 : On dit qu'un estimateur \hat{f}_n est asymptotiquement uniformément sans biais si :

$$\lim_{n \rightarrow \infty} \sup [\mathbb{E}(\hat{f}_n(x) - f(x))]$$

Définition 4 : On dit qu'un estimateur est convergent en moyenne quadratique si :

$$\lim_{n \rightarrow \infty} MSE(\hat{f}_n(x)) = \lim_{n \rightarrow \infty} E[(\hat{f}_n(x) - f(x))^2] = 0$$

Définition 5 : On dit qu'un estimateur est convergent en moyenne quadratique intégrée si :

$$\lim_{n \rightarrow +\infty} MISE(\hat{f}_n(x)) = \lim_{n \rightarrow \infty} \int MSE(\hat{f}_n)$$

Définition 6 : On dit qu'un estimateur \hat{f}_n de f est asymptotiquement normale si :

$$\hat{f}_n \rightarrow N(E[\hat{f}_n], var[\hat{f}_n])$$

2.3 Estimation par la méthode d'histogramme :

Étant donné un ensemble de N observations vectorielles $M_N = x_1, \dots, x_N$ décrites par d variables, donc $M_N \subset \mathbb{R}^d$.

Soient : x_1, x_2, \dots, x_n sont des réalisations de la variables aléatoires i.i.d X_1, X_2, \dots, X_n de densité f inconnue sur l'intervalle fini $[a, b]$.

Pour estimer d'une manière non paramétrique, il est naturel de se donner k intervalles (Classes) $I_j, j = 1, 2, \dots, k$ avec $I_j = [a_{j-1}, a_j]$. Pour construire l'histogramme nous devons choisir une origine x_0 et une largeur d'intervalle h , la largeur h qui contrôle principalement la qualité de lissage.

Supposons que la droite est coupée en intervalles, la densité estimée est :

$$\hat{f}_h(x_i) = \frac{1}{n} \frac{\text{cas } (x_i \text{ est dans la meme classe que } x)}{\text{largeur de l'intervalle contenant } x}$$

L'estimateur de f sur $[a_j, a_{j+1}]$ est donnée par : $\hat{f}_h(x) = \frac{n_j}{n \cdot h}$

Remarque1 :

Nous pouvons vérifier que cette estimation est elle-même une fonction de densité ; en faisant la somme des densités sur la totalité des intervalles de h , et en tenant compte du fait que la densité est nulle à l'extérieur de h nous obtenons la valeur 1.

Le module Seaborn de Python nous fournit aussi des fonctions pour des graphiques utiles pour l'analyse statistique. Par exemple, la fonction `distplot` permet non seulement de visualiser l'histogramme d'un échantillon (Figure 3), mais aussi d'estimer la distribution dont l'échantillon est issu. Comme le montre la figure (4) :

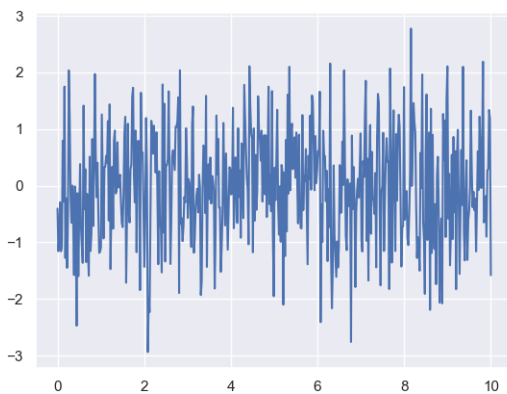


FIGURE 2 – Données générées avec Seaborn

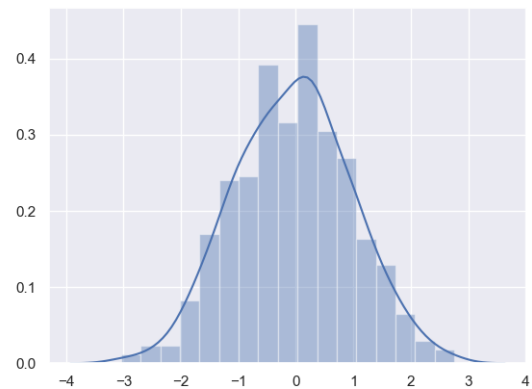


FIGURE 3 – Estimation de la distribution

Remarque 2 :

Découpage de ce domaine h en intervalles de même volume V . Nous reviendrons sur le choix du nombre d'intervalles (ou de leur largeur).

La figure suivante montre un ensemble d'observations de $h \subset \mathbb{R}^2$ et un découpage de ce domaine en intervalles carrés.

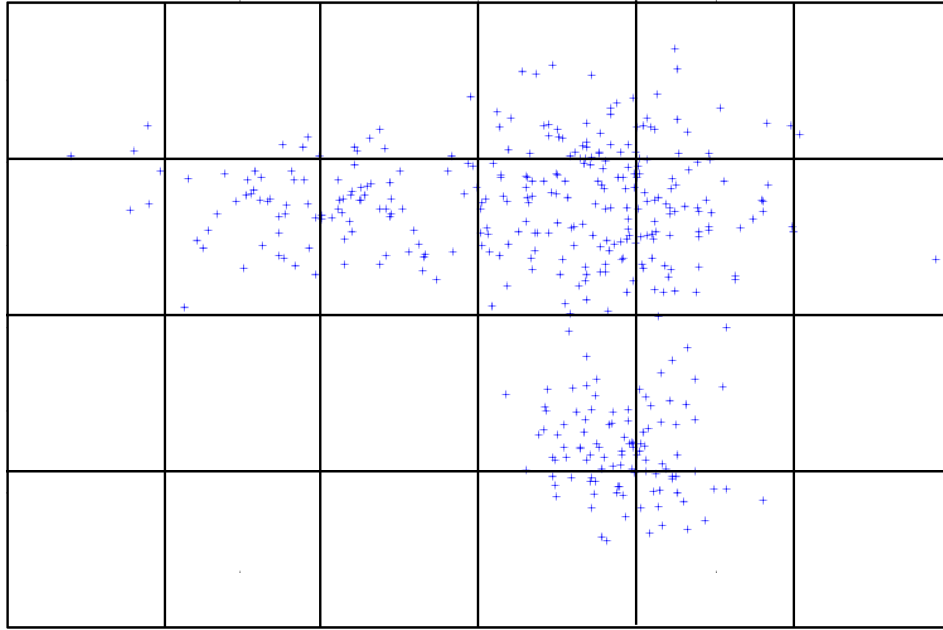


FIGURE 4 – *Exemple de découpage d'un domaine bidimensionnel h en intervalles $2D$*

Soit pour des données unidimensionnelles, un histogramme est facile à représenter, voir par ex. la figure Fig 6. Pour des données bidimensionnelles une représentation par des colonnes en 3D est moins lisible car certaines colonnes cachent d'autres.

Afin de comprendre les caractéristiques de cette méthode d'estimation de densités, il est nécessaire de s'intéresser aux aspects suivants :(12)

1. Comment varie l'estimation lors d'un décalage spatial du domaine h et donc de son découpage en intervalles.
2. Quelle est la relation entre la qualité d'estimation (la proximité entre l'estimation et la vraie valeur) et la résolution de l'estimation (la précision spatiale).
3. Comment varie l'estimation lorsqu'elle est réalisée à partir d'un autre échantillon M'_N de même effectif N et issu de la même distribution.

Le décalage spatial du domaine C et, donc, de son découpage en intervalles a un effet significatif sur l'estimation. Par exemple, la densité peut être nettement sous-évaluée dans une zone dense si elle est centrée sur une jonction de la grille, donc partagée entre plusieurs intervalles voisins.

Dans le but d'étudier la relation entre la qualité d'estimation et la résolution de l'estimation, considérons un exemple unidimensionnel de données générées suivant une distribution uniforme dans l'intervalle $[0, 1]$. Considérons d'abord un découpage fixé et une taille de l'échantillon qui augmente. La figure suivante illustre(12) cette situation.

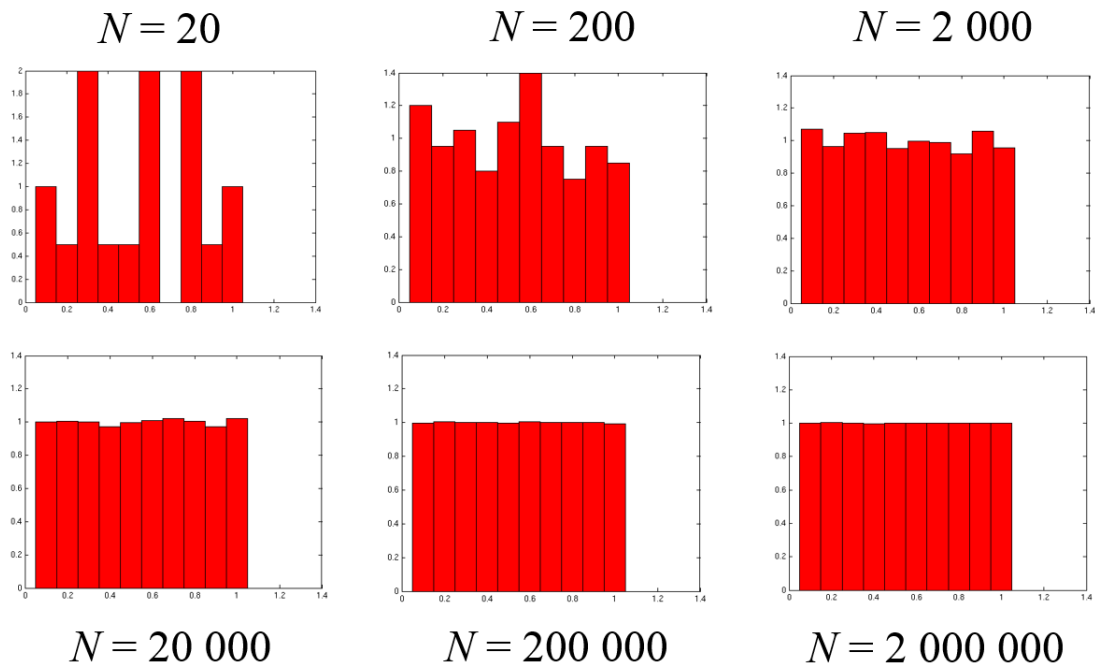


FIGURE 5 – Nombre d'intervalles fixé, la taille de l'échantillon augmente

On constate dans ce cas une amélioration progressive de l'estimation (réduction de la variance) avec l'augmentation de la taille de l'échantillon. L'histogramme ressemble de plus en plus à la représentation d'une distribution uniforme.

Considérons maintenant une taille d'échantillon fixe ($N = 2000$) et un nombre d'intervalles qui augmente, cas illustré dans la figure suivante : (12)

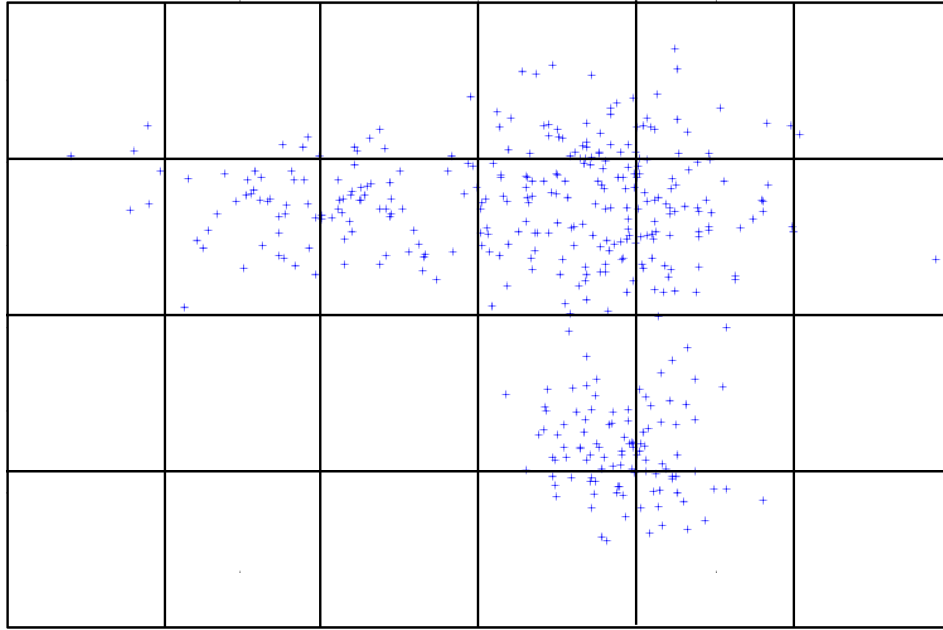


FIGURE 6 – Exemple de découpage d'un domaine bidimensionnel h en intervalles $2D$

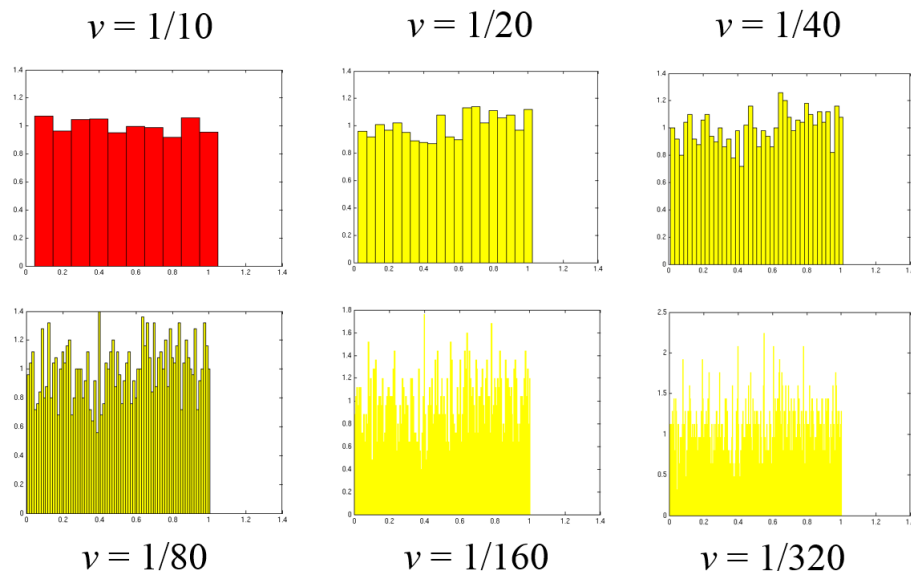


FIGURE 7 – figure
Échantillon fixé, le nombre d'intervalles augmente

Dans ce cas, la résolution spatiale s'améliore avec l'augmentation du nombre d'intervalles mais l'estimation se détériore (la variance augmente).

L'histogramme ressemble de moins en moins à la représentation d'une distribution uniforme.

Enfin, la figure suivante montre un autre échantillon M^N de même effectif N et issu de la même distribution. L'estimation varie de façon significative dans les intervalles où le nombre d'observations est relativement faible.

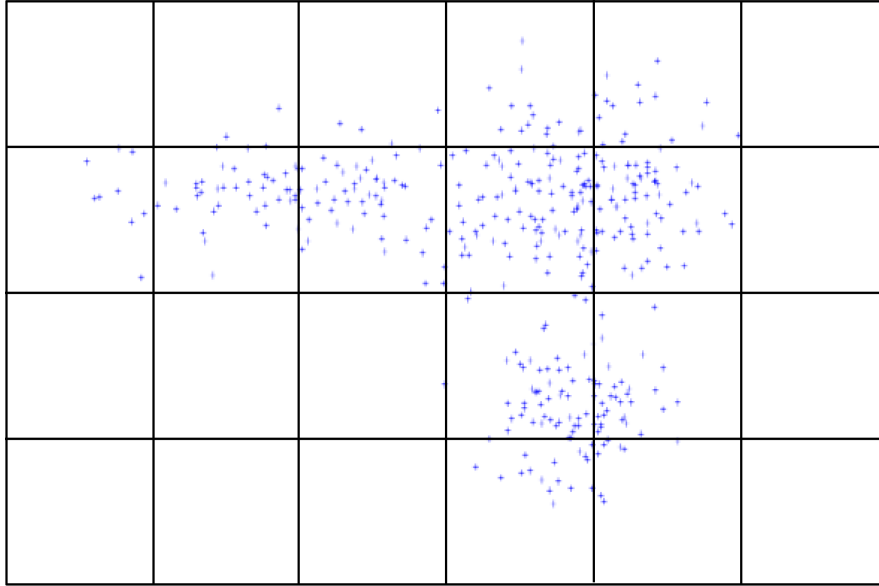


FIGURE 8 – Un autre échantillon issu de la même fonction de densité f produira des estimations par histogramme différentes.

2.4 Propriétés

(5)

Ces histogrammes illustrent une caractéristique des méthodes d'estimations de densité lorsqu'elles ont un faible biais, elles ont tendance à voir une grande variabilité (lissage important).

Pour construire un bon histogramme, il faut trouver un juste équilibre entre le nombre d'observation n et le nombre de classe pour mesurer les performances de f_h , on peut utiliser l'erreur quadratique moyenne intégrée $MISE((x))$.

On a :

$n_j \rightarrow B(n, P_j)$ est la loi binomiale pour $x \in [a_j, a_{j+1}]$, on déduit que :

$$E[\hat{f}_h(x)] = E\left(\frac{n_j}{n \cdot h}\right) = \frac{1}{n \cdot h} \cdot E(n_j) = \frac{p_j}{h}$$

Et :

$$\begin{aligned} Var(\hat{f}_h(x)) &= Var\left(\frac{n_j}{n \cdot h}\right) \\ &= \frac{1}{n \cdot h^2} \cdot p_j(1 - P_j) \quad \text{Car } n_j \rightarrow B(n, P_j). \end{aligned}$$

L'erreur quadratique moyenne de $\hat{f}_h(x)$ est :

$$MSE(\hat{f}_h(x)) = var(\hat{f}_h(x)) + biais^2(f_h(x))$$

On a :

$$\begin{aligned} Biais(\hat{f}_h(x)) &= \mathbb{E}(\hat{f}_h(x)) - f_h(x) \\ &= \frac{P_j}{h} - f(x) \\ &= \frac{1}{h} \int_{a_j}^{a_{j+1}} f(t)dt - f(x). \end{aligned}$$

En faisant le développement de Taylor à $f(t)$. On obtient :

$$\begin{aligned} \text{Biais}(\hat{f}_h(x)) &= \frac{1}{h} \int_{a_j}^{a_{j+1}} [f(x) + (t-x)f'(x) + o(h^2)]dt - f(x) \\ &= \frac{f''(x)}{2} \cdot [h - 2(x - a_j)]^2 + o(h^2) \end{aligned}$$

De plus si $x \in [a_j, a_{j+1}]$:

$$\begin{aligned} \text{var}(\hat{f}_h(x)) &= \frac{P_j(1 - P_j)}{h^2} \\ &= \frac{\int_{a_j}^{a_{j+1}} f(t)dt}{n \cdot h^2} - \frac{P_j^2}{n \cdot h^2} \\ &= \frac{f(x)}{n \cdot h} + o(n^{-1}). \end{aligned}$$

Pour que $\hat{f}_h(x)$ soit peu variable il faut que $nh \rightarrow \infty$ donc :

$$\begin{aligned} \text{MSE}(\hat{f}_h(x)) &= \text{var}(\hat{f}_h(x)) + \text{biais}^2(f_h(x)) \\ &= \frac{f(x)}{n \cdot h} + \frac{f''^2(x)}{4} [h - 2(x - a_j)]^2 + o(h^2) + o(n^{-1}). \end{aligned}$$

Finalement en intégrant par rapport à x . On montre que :

$$\text{MISE}(\hat{f}_h(x)) = \int_{a_j}^{a_{j+1}} \text{MSE}(\hat{f}_h(x))dx = \frac{1}{n \cdot h} + h^2 \frac{\int_{a_j}^{a_{j+1}} f'(x)dx}{12} + o(h^3) + o(n^{-1}) \dots (*)$$

Dans (*), nous pouvons voir que le paramètre h est relié directement de terme $h^2 \frac{\int_{a_j}^{a_{j+1}} f'(x)dx}{12}$ provenant de carré de biais intégré et que ce paramètre est inversement proportionnel à la variance intégrée.

h donne un histogramme peu biaisé tandis qu'en H et N déterminent un histogramme moins variable.

2.5 Choix du paramètre

En pratique le choix de h est en fonction de la taille de l'échantillon n . Les règles les plus utilisées sont : **la règle de Scott** : la valeur qui minimise l'erreur quadratique est (MIS).

$$h_{opt} = \left[\frac{S}{\int f'(t)dt} \right]^{\frac{1}{3}} \cdot n^{-\frac{1}{3}}$$

En prenant pour f la densité de la loi normale $N(n, \sigma)$. Nous pouvons montrer que :

$$h_{opt} = 3.431 \cdot S \leq n^{-\frac{1}{3}}$$

En estimant \sum Par S.P, on obtient ainsi la règle de Scott(5) :

$$h_{opt} = 3.431 \cdot S \cdot n^{-\frac{1}{3}}$$

La règle de Friedman : (5)

$$h_{opt} = 2 \cdot I_q \cdot n^{-\frac{1}{3}}$$

avec

$$I_q = Q_3 - Q_2$$

.

2.6 Conclusion

L'estimation de la densité varie de façon significative lors d'un décalage spatial du domaine h et donc de son découpage en intervalles. l'estimation obtenue pour la densité peut être significativement différente.

En fixant le nombre d'intervalles (un volume v fixé par intervalle), la qualité d'estimation s'améliore avec l'augmentation de l'effectif N mais représente une moyenne sur un volume v assez large. Avec une taille N d'échantillon fixée, la résolution s'améliore avec la l'augmentation du nombre d'intervalles (la diminution de v) mais la qualité d'estimation se dégrade fortement.

Nous pouvons constatés des estimations différentes à partir de deux échantillons de même effectif issus de la même distribution.

Enfin, nous avons présentés les deux règles les plus utilisées pour le choix de paramètre h .

2.7 Estimation par la méthode de noyau :

La méthode d'estimation à noyau de la densité de probabilité, initiée par Rosenblatt en 1956 et développée par Parzen en 1962. Ce dernier a étudié ces propriétés statistiques fondamentales.

Définition :

Soit x_1, \dots, x_n un échantillon de taille n et de loi $f(x)$ sur R de fonction de répartition $F(x)$.

$$F : R \rightarrow [0, 1]$$

Définie pour tout $x \in R$.

$$F_n = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{x_i \leq x}$$

On peut écrire d'une manière équivalente :

$$F_n = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{x \in [x_i, \infty]}$$

La fonction de répartition empirique F_n est un estimateur simple de $F(x)$ et :

$$n.F_n \rightarrow B(n, F(x))$$

Et B est la loi Binomiale.

$$E(F_n(x)) = F(x)$$

Et

$$Var(F_n(x)) = \frac{1}{n} [1 - F(x)]F(x)$$

A partir de la définition d'une densité de probabilité et de la fonction de répartition, et en utilisant la fonction de répartition empirique, la fonction f peut s'écrire en ces points de continuité :

$$f_h(x) = \lim_{x \rightarrow 0} = \frac{F_n(x+h) - (F_n(x-h))}{2.h}$$

$$\hat{f}_h(x) = \frac{1}{2.n.h} \sum \mathbb{K}(x_i)_{[x-h, x+h]} \dots \dots (*)$$

En posant

$$W(n) = \begin{cases} \frac{1}{2} & \text{si } 1 \leq x \leq 1 \\ 0 & \text{sinon.} \end{cases}$$

Nous pouvons réécrire (*) sous la forme :

$$\hat{f}_h(x) = \frac{1}{2.n.h} \sum_{i=1}^n W \frac{x - x_i}{h}$$

Nous venons de définir un estimateur à noyau dit de resamblatt (uniforme). Parzen a étudié une classe générale d'estimation, remplaçant la fonction par une fonction noyau k (kernel) satisfaisant la condition :

$$\int k(x) d(x) = 1$$

Généralement k est une densité de probabilité par analyse avec la définition d'estimation de parzen est :

$$\hat{f}_h(x) = \frac{1}{2.n.h} \sum_{i=1}^n \hat{k} \frac{x - x_i}{h}$$

On a h paramètre de lissage, k est une fonction noyau définie sur R. La figure suivante montre un exemple simple dans lequel les quelques observations (représentées par les points bleus dans le plan horizontal) sont issues de R^2 . La fonction noyau employée est la loi normale et le paramètre de lissage est suffisamment faible pour que les noyaux individuels soient en général séparés dans la densité représentée par la surface courbe.

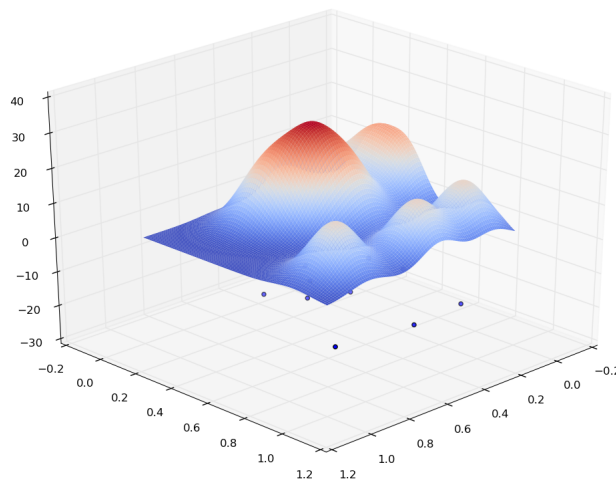


FIGURE 9 – Illustration d'une estimation de densité par noyaux

Soit k noyau h remplissent les conditions suivantes :

$$\int_R k(y) d(y) = 1, \quad \int_R y.k(y) d(y) = 0, \quad \int_R y^2 k(y) d(y) = \sigma_k^2 \dots (1, 1)$$

On peut vérifier que $\hat{f}_k(x)$ est une densité de probabilité avec :

$$\hat{f}_k(x) \geq 0 \text{ si } x \in B$$

Et

$$\begin{aligned} \int_R \hat{f}_k(x) &= \int_R \frac{1}{n.h} \sum_{i=1}^n k \frac{x - x_i}{h} \\ &= \frac{1}{n.h} \sum_{i=1}^n \int k \frac{x - x_i}{h} \\ &= \frac{1}{n} \sum_{i=1}^n \int k(y) dy \\ &= \frac{1}{n} \sum_{i=1}^n 1 = 1 \end{aligned}$$

2.8 Noyaux usuels(5)

Les noyaux les plus utilisés dans l'estimation de la densité de probabilité sont données dans le tableau suivant :

Noyaux	$K(\mu)$	
Uniforme	$\frac{1}{2} \text{ si } \mu \in [-1, 1]$	
Triangulaire	$(1 - \mu), \mu \leq 1$	
Gaussien	$\frac{1}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2}}, \mu \in \mathbb{R}$...(1,2)
Quadratique	$\frac{15}{16}(1 - \mu^2)^2, \mu \in [-1, 1]$	
Epanechnikov	$\frac{3}{4}(1 - \mu^2), \mu \in [-1, 1]$	
Gbrique	$\frac{35}{32}(1 - \mu^2)^3, \mu \in [-1, 1]$	

Les figures suivantes montrent deux exemples de fonctions noyaux qui satisfont ces conditions (1,1) :

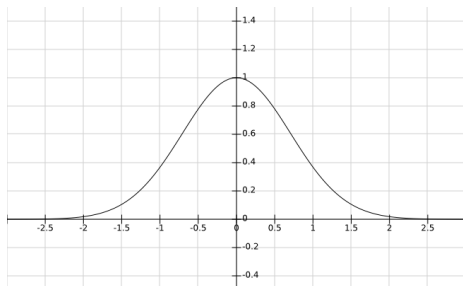


FIGURE 10 – Exemple d'un noyau gaussien

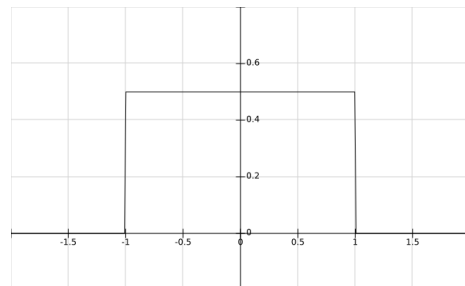


FIGURE 11 – Exemple d'un noyau uniforme

2.9 Définitions :(7)

K est une fonction non négative définie de \mathbb{R} dans \mathbb{R} vérifiant :

1. $\int K(x)dx = 1.$
2. $\int K(x)dx < \infty.$
3. $\sup_x |K(x)| < \infty.$

2.10 Propriétés de l'estimateurs à noyau, Espérance, variance de l'estimateur

- **L'espérance de l'estimateur** : L'espérance mathématique de l'estimateur de $\hat{f}_h(x)$ est :

$$\begin{aligned}\mathbb{E}(\hat{f}_h(x)) &= \frac{1}{n.h} \mathbb{E}[\sum_{i=1}^n k(\frac{x_i - x}{h})] \\ &= \frac{1}{n.h} \sum_{i=1}^n \mathbb{E}[k(\frac{x_i - x}{h})] \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} K(\frac{\mu - x}{h}) f(\mu) d\mu, \text{ avec : } \mu = x_i.\end{aligned}$$

En posant : $y = \frac{\mu - x}{h} \Rightarrow dy = \frac{d\mu}{h} \Rightarrow d\mu = h.dy.$

$$\mathbb{E}(\hat{f}_h(x)) = \int_{-\infty}^{+\infty} K(y) f(x + h.y) dy$$

En faisant le développement de taylor à l'ordre 2 au point $h = 0$ de $f(x + h.y)$ on obtient :

$$f(x + h.y) = f(x) + \frac{h.y}{1!} f'(x) + \frac{h^2.y^2}{2!} f''(x) + o(h^2).$$

d'où :

$$\begin{aligned}\mathbb{E}(\hat{f}_h(x)) &= \int_{-\infty}^{+\infty} K(y) [f(x) + \frac{h.y}{1!} f'(x) + \frac{h^2.y^2}{2!} f''(x)] dy + o(h^2). \\ &= \int_{-\infty}^{+\infty} K(y) f(x) dy + \int_{-\infty}^{+\infty} h.y.K(y) f'(x) dy + \int_{-\infty}^{+\infty} \frac{h^2.y^2}{2!} K(y) f''(x) dy + o(h^2) \\ &= f(x) \int_{-\infty}^{+\infty} K(y) dy + h f'(x) \int_{-\infty}^{+\infty} y.K(y) dy + \frac{h^2}{2!} f''(x) \int_{-\infty}^{+\infty} y^2.K(y) dy + o(h^2) \\ &= f(x) + \frac{h^2}{2!} f''(x) \mu_2(K) + o(h^2).\end{aligned}$$

- **Le biais de l'estimateur** $\hat{f}_h(x)$: est donnée par :

$$\begin{aligned}biais(\hat{f}_h(x)) &= \mathbb{E}(\hat{f}_h(x)) - f(x) \\ &= \int_{-\infty}^{+\infty} K(y) f(x + h.y) dy - f(x) \\ &= \frac{h^2}{2!} f''(x) \mu_2(K) + o(h^2).\end{aligned}$$

Avec : $\mu_2(K) = \int_{-\infty}^{+\infty} y^2 K(y) dy$

- **La variance de l'estimateur $\hat{f}_h(x)$** : est donnée par :

$$\begin{aligned}
var(\hat{f}_h(x)) &= \mathbb{E}[(\hat{f}_h(x))^2] - [\mathbb{E}(\hat{f}_h(x))]^2 \\
&= var\left(\frac{1}{n.h} \sum_{i=1}^n k\left(\frac{x_i - x}{h}\right)\right) \\
&= \frac{1}{n^2.h^2} \sum_{i=1}^n var\left(k\left(\frac{x_i - x}{h}\right)\right) \\
&= \frac{1}{n^2.h^2} \sum_{i=1}^n [\mathbb{E}\left(\left(\frac{x_i - x}{h}\right)^2\right) - (\mathbb{E}\left(\frac{x_i - x}{h}\right))^2] \\
&= \frac{1}{n.h} \int_{-\infty}^{+\infty} K(y)^2 f(x + h.y) dy + \frac{1}{n.h} \left[\int_{-\infty}^{+\infty} K(y) f(x + h.y) dy \right]^2 \\
&= \frac{f(x)}{n.h} \int_{-\infty}^{+\infty} K(y)^2 dy - \frac{f'(x)}{n} \int_{-\infty}^{+\infty} y K(y)^2 dy - \frac{1}{n} [f(x) + biais(\hat{f}_h(x))]^2.
\end{aligned}$$

- **La MSE de l'estimateur $\hat{f}_h(x)$** : est donnée par :

$$\begin{aligned}
MSE(\hat{f}_h(x)) &= var(\hat{f}_h(x)) + (biais(\hat{f}_h(x)))^2 \\
&= \frac{f(x)}{n.h} \int_{-\infty}^{+\infty} K(y)^2 dy + \left[\frac{h^2}{2} f''(x) \cdot \mu_2(K) \right]^2 + O\left(\frac{1}{n.h}\right) + o(h^4) \\
&= \frac{f(x)}{n.h} \int_{-\infty}^{+\infty} K(y)^2 dy + \frac{h^4}{4} f''(x)^2 \left[\int_{-\infty}^{+\infty} y K(y) dy \right]^2 + o\left(\frac{1}{n.h} + h^4\right)
\end{aligned}$$

2.11 Comportement asymptotique du biais et de la variance(5)

- a) **Théorème 1 (Parzen)** : convergence du biais.

Si on a :

1. $\lim h = 0$.
2. $\sup_x |K(x)| < \infty$.
3. $\int_{-\infty}^{+\infty} K(y) dy = 1$. et $\int_{-\infty}^{+\infty} y K(y) dy = 0$.

Alors l'estimateur $\hat{f}_h(x)$ est asymptotiquement sans biais .

$$\lim \mathbb{E}(\hat{f}_h(x)) = f(x).$$

- b) **Théorème 2** : convergence de la variance.(2)

1. $\int_{-\infty}^{+\infty} K(y) dy = 1$.
2. $\sup_x |K(x)| < \infty$.
3. $\int_{-\infty}^{+\infty} y K(y) dy = 0$.
4. $\lim_{n \rightarrow \infty} n.h = \infty$ alors $var(\hat{f}_h(x))$ tend vers 0, lorsque $n.h \rightarrow \infty$

- c) **Théorème 3** : Si $\lim_{n \rightarrow \infty} h = 0$ et $\lim_{n \rightarrow \infty} n.h = \infty$ et K est un noyau satisfait aux conditions suivantes :

- $\sup_x |K(x)| < \infty$, $\int_{-\infty}^{+\infty} y K(y) dy = 0$.
- $\int_{-\infty}^{+\infty} K(y) dy = 1$. Alors l'estimateur $\hat{f}_h(x)$ est consistant en moyenne quadratique : $\lim_{n \rightarrow \infty} (MSE(\hat{f}_h(x), f(x))) = 0$.

2.11.1 L'erreur quadratique moyenne intégrée $MISE(\hat{f}_h(x))$: (5)

$$\begin{aligned}
MISE(\hat{f}_h(x), f(x)) &= \int_{-\infty}^{+\infty} (MSE(\hat{f}_h(x), f(x))) dx \\
&= \int_{-\infty}^{+\infty} \text{var}(\hat{f}_h(x)) + (\text{biais}(\hat{f}_h(x)))^2 \\
&= \int_{-\infty}^{+\infty} \left[\frac{f(x)}{n.h} \int_{-\infty}^{+\infty} K(y)^2 dy + \left[\frac{h^2}{2} f''(x) \cdot \mu_2(K) \right]^2 + O\left(\frac{1}{n.h} + h^4\right) \right] \\
&= \frac{1}{n.h} \int_{-\infty}^{+\infty} K(y)^2 dy \cdot \int_{-\infty}^{+\infty} f(x) dx + \frac{h^4}{4} \mu_2(K) \int_{-\infty}^{+\infty} f''(x) dx + O\left(\frac{1}{n.h} + h^4\right) \\
&= \frac{1}{n.h} \int_{-\infty}^{+\infty} K(y)^2 dy + \frac{h^4}{4} \mu_2(K) \int_{-\infty}^{+\infty} f''(x) dx + O\left(\frac{1}{n.h} + h^4\right) \\
&= \frac{1}{n.h} R(k) + \frac{h^4}{4} \sigma_k^2 R(f'') + o\left(\frac{1}{n.h} + h^4\right).
\end{aligned}$$

Avec : $R(k) = \int_{-\infty}^{+\infty} K(y)^2 dy$.

2.11.2 L'erreur quadratique moyenne intégrée asymptotique $AMISE(\hat{f}_h(x))$:

$$\begin{aligned}
AMISE(\hat{f}_h(x), f(x)) &= MISE(\hat{f}_h(x), f(x)) - o\left(\frac{1}{n.h} + h^4\right) \\
&= \frac{1}{n.h} R(k) + \frac{h^4}{4} \sigma_k^2 R(f'')
\end{aligned}$$

2.12 Comportement asymptotique et vitesse de convergence de l'estimateur à noyau

Proposition : Si f est bornée et K est un noyau de carrée intégrable ie : $\int_{-\infty}^{+\infty} K^2(y) dy < \infty$, alors $\forall x \in \mathbb{R}$:

$$\text{var}(\hat{f}_h(x)) = \text{var}\left(\frac{1}{n.h} \sum_{i=1}^n k\left(\frac{x_i - x}{h}\right)\right).$$

Preuve :

$$\begin{aligned}
\text{var}(\hat{f}_h(x)) &= \text{var}\left(\frac{1}{n.h} \sum_{i=1}^n k\left(\frac{x_i - x}{h}\right)\right) \\
&= \frac{1}{n^2.h^2} \text{var}\left(k\left(\frac{x_i - x}{h}\right)\right), \quad \text{car les v.a sont i.i.d.} \\
&= \frac{1}{n^2.h^2} \text{var}\left(k\left(\frac{x_i - x}{h}\right)\right), \quad \text{car les v.a ont la meme loi.}
\end{aligned}$$

On déduit la majoration suivante :

$$\begin{aligned}
\text{var}(\hat{f}_h(x)) &< \frac{1}{n^2.h^2} \mathbb{E}\left(k^2\left(\frac{x_i - x}{h}\right)\right) \\
&\leq \frac{1}{n.h^2} \int_{-\infty}^{+\infty} K^2\left(\frac{\mu - x}{h}\right) f(\mu) d\mu \\
&\leq \frac{\|f\|_\infty}{n.h^2} \int_{-\infty}^{+\infty} K^2\left(\frac{\mu - x}{h}\right) d\mu, \quad \text{on pose } v = \frac{\mu - x}{h} \\
&\leq \frac{\|f\|_\infty}{n.h} \int_{-\infty}^{+\infty} K^2(v) dv
\end{aligned}$$

Avec : $\|f\|_\infty = \sup_x(f)$.

Etude de biais : Soit K un noyau symétrique $\int_{-\infty}^{+\infty} \mu K(\mu) d\mu = 0$, et f est de classe C^2 telle que f'' est bornée :

$$\begin{aligned} \text{biais}(\hat{f}_h(x)) &= \mathbb{E}(\hat{f}_h(x)) - f(x) = \mathbb{E}\left(\frac{1}{n.h} \sum_{i=1}^n k\left(\frac{x_i - x}{h}\right)\right) - f(x) \\ &= \frac{1}{h} \mathbb{E}\left(k\left(\frac{x_i - x}{h}\right)\right) - f(x) \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} k\left(\frac{\mu - x}{h}\right) f(\mu) d\mu - f(x) \\ &= \int_{-\infty}^{+\infty} K(v) f(x + hv) dv - f(x) \\ &= \int_{-\infty}^{+\infty} K(v) [f(x + hv) - f(x)] dv \end{aligned}$$

Car le noyau est une densité de probabilité, et F est supposée de classe C^2 , On peut appliquer la formule de Taylor à $f(x + hv) - f(x)$ ce qui nous donne :

$$f(x + hv) - f(x) = hv f'(x) + \frac{(vh)^2}{2!} f''(x + thv), \quad t \in [0, 1].$$

$$\begin{aligned} \text{biais}(\hat{f}_h(x)) &= \int_{-\infty}^{+\infty} K(v) [vh f'(x) + \frac{(vh)^2}{2!} f''(x + thv)] dv \\ &= \int_{-\infty}^{+\infty} hv K(v) f'(x) dv + \frac{h^2}{2!} \int_{-\infty}^{+\infty} v^2 K(v) f''(x + thv) dv \end{aligned}$$

On a f'' est bornée :

$$\text{biais}(\hat{f}_h(x)) \leq \frac{h^2}{2!} \|f''\|_\infty \int_{-\infty}^{+\infty} v^2 K(v) dv.$$

Définition : Soit T un intervalle de \mathbb{R} et deux réels $B, L > 0$, la classe de holder $\sum(B, L)$ sur T est définie comme l'ensemble des fonction $g : T \rightarrow \mathbb{R}$ telle que g est :

$l = \lfloor B \rfloor$ fois dérivable et $|g^l(x) - g^l(y)| \leq L|x - y|^{B-L}$, $\forall x, y \in Y$.

Définition : Soit $l \in \mathbb{N}$, On dit que $K : \mathbb{R} \rightarrow \mathbb{R}$ est un noyau d'ordre l si la fonction $\begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ x \rightarrow x^j K(x), \end{cases}$ sont intégrable pour j et vérifiant : $\int_{-\infty}^{+\infty} K(\mu) d\mu = 1$, et $\int_{-\infty}^{+\infty} \mu^j K(\mu) d\mu = 0$. pour $j = 1, 2, \dots, l - 1$.

Proposition : Si la densité f est dans une classe de holder $\sum(B, L)$ sur $T \rightarrow \mathbb{R}$ et si K est un noyau d'ordre $l = \lfloor B \rfloor$ et telle que

$$\int_{-\infty}^{+\infty} |\mu|^B K(\mu) d\mu < \infty.$$

Alors pour tout h positive ($h > 0$) on a :

$$|\text{biais}(\hat{f}_h(x))| \leq \sqrt{c_2} h^B$$

avec : $\sqrt{c_2} = \frac{1}{l!} \int_{-\infty}^{+\infty} |\mu|^B K(\mu) d\mu$, et l c'est l'ordre de $f(x + hv) - f(x)$.

On peut aussi déduire une majoration pour le MSE :

$$MSE(\hat{f}_h(x)) \leq \frac{c_1}{n.h} + c_2 h^{2B}$$

Avec : $c_1 = \|f\|_\infty \int_{-\infty}^{+\infty} K^2(\mu) d\mu$.

Pour déterminer le paramètre de lissage h on dérivant le $MSE(\hat{f}_h(x))$ on obtient :

$$h^* = \left(\frac{c_1}{2Bc_2}\right)^{\frac{1}{2B+1}} n^{-\frac{1}{2B+1}}.$$

2.13 choix de paramètre de lissage h

On peut constater à travers les deux figures en ci-dessous, qu'une valeur trop faible pour h produit des résultats sans intérêt, la densité étant non nulle seulement à proximité immédiate des observations de D_N , alors

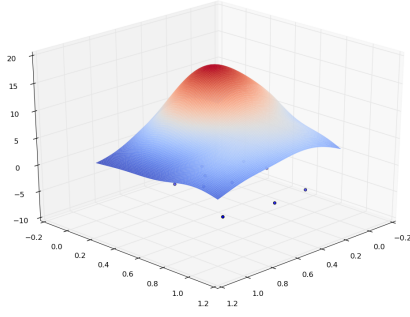


FIGURE 12 – Estimation pour $h^2 = 0,05$

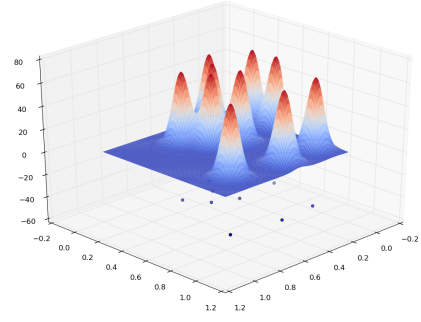


FIGURE 13 – Estimation pour $h^2 = 0,2$

qu'une valeur trop élevée produit un lissage excessif avec perte des détails.

Le choix du paramètre h est donc critique pour obtenir des résultats satisfaisants. Parmi les méthodes de choix de la valeur optimale de h , nous présentons ci-dessous les plus importantes.

- a) **Méthode de validation croisée** : Pour sélectionner le *fenêtre* h , on va minimiser l'erreur quadratique moyenne intégrée (MISE) :

$$h^* = \arg \min(MISE(\hat{f}_h(x))).$$

- b) **Méthode de validation croisée de la vraisemblance** : Pour un estimateur $\hat{f}_h(x)$ de $f(x)$ et une largeur h . La sélection pour validation croisée de la vraisemblance est une approche classique fondée sur un critère non asymptotique du maximum de vraisemblance, elle consiste de maximiser par rapport à h , la vraisemblance pour l'échantillon (X_i) est définie par :

$$L(v(h)) = \prod_{i=1}^n \hat{f}_h(x_i)$$

où :

$$\hat{f}_h(x_i) = \frac{1}{(n-1)h} \sum_{j=1, j \neq i} K\left(\frac{x_j - x_i}{h}\right).$$

- c) **Méthode de plug in itérée** : La décision d'un choix optimal pour le paramètre de lissage h s'oppose la spécification d'un critère d'erreur qui puisse être optimiser. Dans ce cas on cherche à minimiser l'erreur quadratique moyenne intégrée asymptotique AMISE :

$$MISE(\hat{f}_h(x)) = \frac{h^4}{4} \sigma_K^4 R(f'') + \frac{R(k)}{n \cdot h}$$

On obtient le h optimale :

$$h^* = \left[\frac{R(k)}{\sigma_K^4 R(f'')} \right]^{\frac{1}{5}} \cdot n^{-\frac{1}{5}}.$$

2.14 Conclusion

On conclut que l'estimateur non-paramétrique par la méthode du noyau de la densité est :

$$\hat{f}_h(x) = \frac{1}{2.n.h} \sum_{i=1}^n \hat{k} \frac{x - x_i}{h}$$

où K est un noyau (kernel en anglais) et h un paramètre nommé fenêtre, qui régit le degré de lissage de l'estimation.

Nous avons vu que lorsque K est choisi, nous pouvons calculer en fonction du choix de la fenêtre h_n . Le meilleur choix théorique obtenu pour h_n et la vitesse de convergence de l'estimateur dépendent de la régularité de la densité. Nous avons indiqué aussi précédemment les méthodes de choix optimale de h que nous pouvons présenté comme suit (5) :

1. Méthode de validation croisée.
2. Méthode de validation croisée de la vraisemblance.
3. Méthode de Plug in itérée.

3 Régression non paramétrique par la méthode de noyau(11)

3.1 Présentation générale des fonctions de lissage

Soit un échantillon aléatoire composée des couples (x_i, y_i) , $i = 1, \dots, n$, ou les x_i représentent les valeurs observées de la variable explicative X et les y_i représentent celles de la variable dépendante Y . Alors, le modèle de régression non paramétrique univariée est donné par(1) :

$$y_i = \mu(x_i) + \epsilon_i, i = 1, \dots, n, \dots(1, 3)$$

Où les ϵ_i sont les erreurs aléatoires, non corrélées entre elles, de moyenne 0 et de variance σ^2 et ou $\mu(x)$ est la fonction de régression que l'on cherche à estimer.

La méthode du noyau estime la fonction de régression en calculant une moyenne pondérée des observations y_i pour toutes les valeurs x_0 du domaine. La pondération accordée à un point i de l'échantillon dépend du type de noyau utilisé ainsi que de la distance entre x_i et x_0 . Un noyau $d(t)$ est une fonction continue sur R et décroissante en $|t|$. Pour obtenir l'estimateur de la fonction de régression $\mu(x)$ du modèle (1.3), à l'aide de la méthode du noyau, on calcule la moyenne pondérée des observations y_i à chaque point x_0 du domaine de X de la façon suivante :

$$\hat{m}(x) = \frac{\sum_{i=1}^n k\left(\frac{x_0 - x_i}{h}\right) y_i}{\sum_{i=1}^n k\left(\frac{x_0 - x_i}{h}\right)} \dots(1, 4)$$

Où λ est le paramètre de lissage prenant des valeurs positives (Hastie et Tibshirani, 1990, p.19). Ce dernier permet de contrôler la grandeur de la fenêtre, centrée à x_0 , à l'intérieur de laquelle les poids accordés aux x_i sont non nuls. Le tableau (1.2) présente quelques noyaux parmi les plus utilisés.

$$\hat{m}(x_0) = \frac{\sum_{i=1}^n K\left(\frac{x_0 - x_i}{h}\right) y_i}{\sum_{i=1}^n K\left(\frac{x_0 - x_i}{h}\right)}.$$

3.2 L'estimation de nadaraya watson[NW](4)

On suppose d'un échantillon composé de n couples indépendants se v.a $((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$.

Dans le modèle de régression non paramétrique on suppose l'existence d'une fonction $m(\cdot)$, qui exprime la valeur moyenne de la variable réponse Y en fonction de la variable d'entrée X.

$$Y = m(x) + \epsilon_i, \quad i = 1, \dots, n.$$

Avec : ϵ_i : Bruit blanc.

La fonction de lien est :

$$\begin{aligned} m(x) &= \mathbb{E}(Y/X = x) = \frac{\int_R y f_{X,Y}(x, y) dy}{f_X(x)} \\ &= \frac{\int_R y f_{X,Y}(x, y) dy}{\int_R f(x, y) dy} = \frac{r(x)}{f_X(x)} \end{aligned}$$

Où : $f_{X,Y}$: est la densité conjointe du couple (X, Y) .

f_X : est la densité marginale de X.

Définition 1 : Un estimateur $\hat{m}_n(x)$ de $m(x)$ est dite linéaire de la régression si :

$$\hat{m}_n(x) = \sum_{i=1}^n w_{i,n}(x) \cdot y_i.$$

Où $w_{i,n}(x)$: est la fonction poids ne dépend pas de y_i .

Définition 2 : Supposons que l'on dispose d'un échantillon $((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$, de v.a réelles de même loi que (X, Y) , On se propose de construire un estimateur de $m(x)$ à partir d'observation $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

Le premier estimateur est l'estimateur à noyau de $[N, W]$. Il est construit à partir d'une fonction noyau $K(\cdot)$ et une largeur h , de manière analogue à l'estimateur à noyau de la densité $f_X(\cdot)$, introduit par [P.R]. L'estimation à noyau de $[N, W]$ est :

$$\hat{m}_n(x) = \begin{cases} \frac{\sum_{i=1}^n y_i K(\frac{x-x_i}{h})}{\sum_{i=1}^n K(\frac{x-x_i}{h})}, & \text{si } \sum_{i=1}^n K(\frac{x-x_i}{h}) \neq 0 \\ \frac{1}{n} \sum_{i=1}^n y_i, & \text{sinon.} \end{cases}$$

Remarque : L'estimateur à noyau d'une fonction bivariee $f_{X,Y}$, est donnée par :

$$\hat{f}_{X,Y}(x, y) = \frac{1}{n \cdot h^2} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) K\left(\frac{y-y_i}{h}\right).$$

Proposition : Si le noyau K est symétrique nous obtenons les résultats suivants :

$$\hat{m}_n^{NW}(x) = \frac{\int_R y f_{X,Y}(x, y) dy}{\hat{f}_X(x)}.$$

3.3 Propriété de l'estimateur

a) **La variance de l'estimateur :**

$$\begin{aligned} \text{var}(\hat{m}_n^{NW}(x)) &= \frac{1}{f_X(x)} \int_R y^2 f_{X,Y}(x, y) dy - (m(x))^2 \\ &= \frac{1}{n \cdot h} \frac{\sigma^2(x)}{f_X(x)} \int_R K^2(v) dv (1 + o(1)). \end{aligned}$$

Avec : $o(1)$ tend vers 0 si h tend vers 0.

Proposition : Nous supposons que $E(y^2) < \infty$, alors à chaque point de continuité des fonctions $f_X(x), m(x)$ et $\sigma(x)$ telle que $f_X(x) > 0$, Nous avons :

$$\text{var}(\hat{m}_n^{NW}(x)) = \frac{1}{n \cdot h^2} \frac{\sigma^2(x)}{f_X(x)} \int_R K^2(x) dx (1 + o(1)).$$

avec : $o(1)$ tend vers 0 si h tend vers 0.

b) **Le biais de l'estimateur :** Le traitement du biais est purement analytique et repose sur le developpement de taylor, Il nous faut supposer certain propriétés de régularité sur la fonction $f_X(\cdot)$ qui determinent l'ordre du biais asymptotique en fonction du paramètre de lisage (h).

Supposons q et $f_X(\cdot)$, sont de classe C^q , et le noyau est d'ordre $(q-1)$ à support compact.

$\mu_0(K) = \int_R K(\mu) d\mu = 1$, et $\mu_j(K) = \int_R \mu^j K(\mu) dx = 0$, $1 \leq j \leq (q-1)$, $\int \mu^q K(\mu) d\mu < \infty$.

Lorsque h tend vers 0 et $(n \cdot h)$ tend vers ∞ , Nous avons :

$$\begin{aligned} \text{biais}(\hat{m}_n^{NW}(x)) &= E[\hat{m}_n^{NW}(x)] - m(x) \\ &= \frac{h^q}{q!} m^q(x) + q m^{q-1} \cdot \frac{f'(x)}{f_X(x)} \int_R \mu^q K(\mu) d\mu \cdot (1 + o(1)). \end{aligned}$$

3.3.1 Optimalité asymptotique(2)

On suppose que $h \rightarrow 0$ et $(n \cdot h) \rightarrow \infty$, lorsque $n \rightarrow \infty$ désigne par $K(q)$, La classe des noyaux d'orde $(q-1)\infty$ à support compact. Nous supposons que le noyau $k \in K(q)$.

$$\text{biais}(\hat{m}_n^{NW}(x)) = \frac{h^q}{q!} [b(x, q)] (1 + o(1))$$

Avec : $b(x, q) = m^q(x) + qm^{q-1} \cdot \frac{f'(x)}{f_X(x)} \int_R \mu^q K(\mu) d\mu$.

Et :

$$\begin{aligned} \text{var}(\hat{m}_n^{NW}(x)) &= \frac{1}{n \cdot h} \frac{\sigma^2(x)}{f_X(x)} \int_R K^2(\mu) d\mu (1 + o(1)) \\ &= \frac{1}{n \cdot h} [v^2(x)] (1 + o(1)) \end{aligned}$$

Théorème 1 : Sous les hypothèses des propositions précédentes on obtient :

$$\begin{aligned} \text{MSE}(\hat{m}_n^{NW}(x)) &= \text{biais}^2(\hat{m}_n^{NW}(x)) + \text{var}(\hat{m}_n^{NW}(x)) \\ &= \left(\frac{1}{n \cdot h} [v^2(x)] + \left[\frac{h^{2q}}{(q!)^2} \cdot b(x, q)^2 \right] \right) (1 + o(1)). \end{aligned}$$

Le MSE asymptotique est donnée par :

$$\text{AMSE}(\hat{m}_n^{NW}(x)) = \frac{1}{n \cdot h} [v^2(x)] + \left[\frac{h^{2q}}{(q!)^2} \cdot b(x, q)^2 \right].$$

Le paramètre de lissage h est :

$$h_{opt}^{AMSE} = \arg \min_h [\text{AMSE}(h)].$$

d'où :

$$h^* = \frac{(q!)^2 v^2(x)}{2q(b(x, q))^2} \cdot n^{-\frac{1}{2q+1}}$$

Théorème 2 : L'erreur quadratique moyenne intégrée (MISE) est :

$$\begin{aligned} \text{MISE}(\hat{m}_n^{NW}(x)) &= \int \text{MSE}(\hat{m}_n^{NW}(x)) dx \\ &= \left[\frac{1}{n \cdot h} \int_R [v^2(x)] + \left[\frac{h^{2q}}{(q!)^2} \cdot \int_R b(x, q)^2 \right] \right] (1 + o(1)). \end{aligned}$$

Le AMISE asymptotique est donnée par :

$$\text{AMISE}(\hat{m}_n^{NW}(x)) = \frac{1}{n \cdot h} \int_R [v^2(x)] + \left[\frac{h^{2q}}{(q!)^2} \cdot \int_R b(x, q)^2 \right].$$

Le paramètre de lissage h associé est donnée par :

$$h^* = \left[\frac{(q!)^2 (q-1)! \int_R v^2(x) dx}{2 \int_R (b(x, q))^2 dx} \right]^{\frac{1}{2q+1}} \cdot n^{-\frac{1}{2q+1}}$$

3.4 Conclusion

Dans cette partie nous allons présenter la théorie entourant la régression non paramétrique par noyau notamment l'estimateur de Nadaraya Watson, nous avons aussi mis l'accent sur le paramètre du lissage . Dans la partie suivante, nous allons aborder un exemple illustratif pour mieux comprendre cette notion de régression par noyau.

4 Application

Dans cette section, les valeurs du noyau sont utilisées pour dériver des poids afin de prédire les sorties à partir d'entrées données. Les étapes du calcul des pondérations et de leur utilisation pour prédire la variable de sortie, y à partir de la variable expliquée x .

Et pour mieux comprendre le fonctionnement de la régression du noyau. Nous allons discuter un exemple d'application (8) sur la régression non paramétrique que nous avons défini précédemment, dont l'objectif est de prédire le débit d'une rivière à partir d'un bassin versant. Les données(10) utilisées sont présentées en ci-dessous :

Area	River Flow
11	2337
22	2750
33	2301
44	2500
56	2100
67	1100
78	1000
89	1642
100	1932
50	1700
70	1750
90	2000

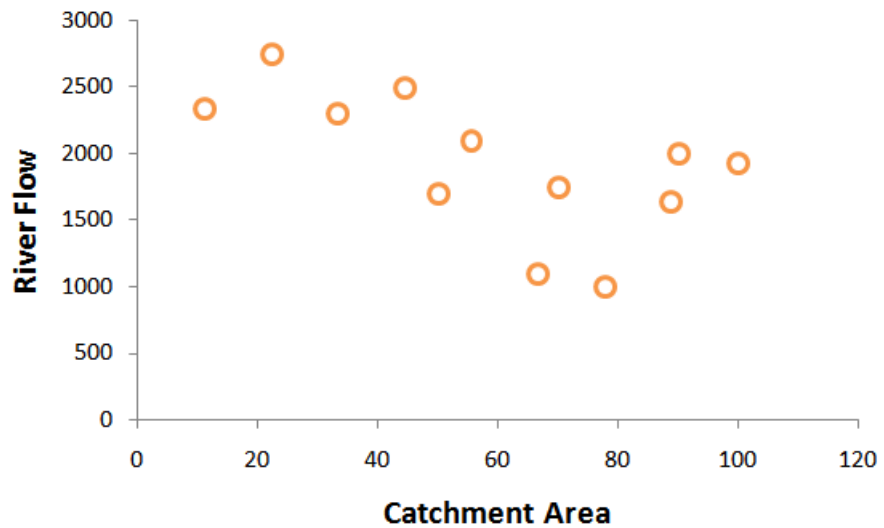


FIGURE 14 – Exemple d'ensemble de données (10) pour construire une régression par noyau

Il existe une relation non linéaire entre le bassin versant (en miles carrés) et le débit de la rivière (en pieds cubes par seconde). La sortie, y est le débit de la rivière et l'entrée, x est le bassin versant dans cet exemple. Dans ce qui suit, nous allons considérer un noyau gaussien suivant :

$$K(x) = \frac{1}{h\sqrt{2\pi}} e^{-0.5\left(\frac{x-x_i}{h}\right)^2}$$

Tout d'abord, nous allons mettre en pratique l'estimation de la densité par la méthode d'histogramme pour estimer la distribution de ces données, comme l'indique la figure suivante :

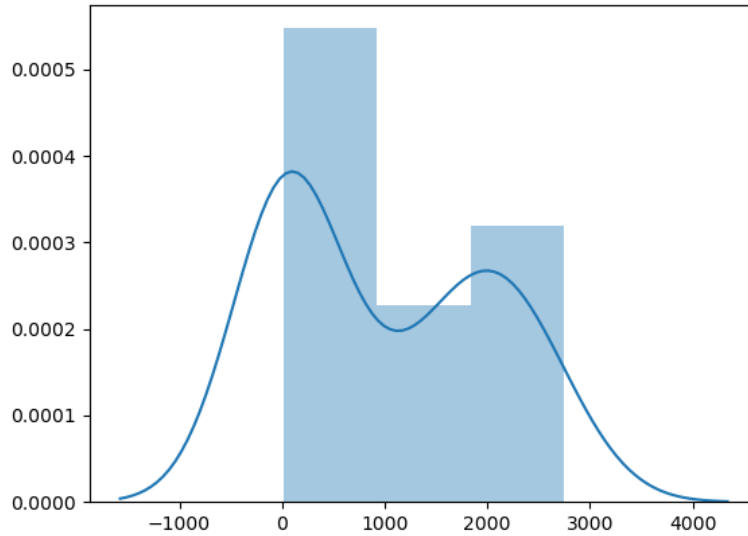


FIGURE 15 – Estimation de densité par la méthode d'histogramme
On observe que les données sont dispersées et s'éloignent fortement de la loi normale

Nous constatons en ci-dessus que le prédicteur, ou fonction d'estimation, ne prend pas de forme prédéterminée, mais est construit selon les informations provenant des données

La régression non paramétrique exige des tailles d'échantillons plus importantes que celles de la régression basée sur des modèles paramétriques parce que les données doivent fournir la structure du modèle ainsi que les estimations du modèle. Dans ce qui suit, nous allons effectuer la régression non paramétrique notamment la régression par l'estimateur de Nadiriya Waston.

4.1 Le noyau en tant que fonction de pondération (9)

Dans un premier temps, les noyaux sont estimés comme décrit dans les sections précédentes en utilisant une valeur de bande passante pour toutes les entrées. Ensuite, on les utilise pour calculer les poids. Dans cet exemple, une valeur de bande passante de 10 est utilisée pour expliquer la régression des noyaux. Cependant, la valeur de la bande passante doit être optimisée pour s'adapter aux données de manière appropriée.

Il est important de comprendre que les noyaux sont développés à toutes les valeurs de x_i . Le calcul fondamental derrière la régression des noyaux consiste à estimer la somme pondérée de toutes les valeurs y observées pour une valeur prédictive donnée, x_i . Les poids ne sont rien d'autre que les valeurs du noyau, échelonnées entre 0 et 1, coupant la ligne perpendiculaire à l'axe x à une valeur x_i donnée (comme le montre la figure ci-dessous pour cet exemple)

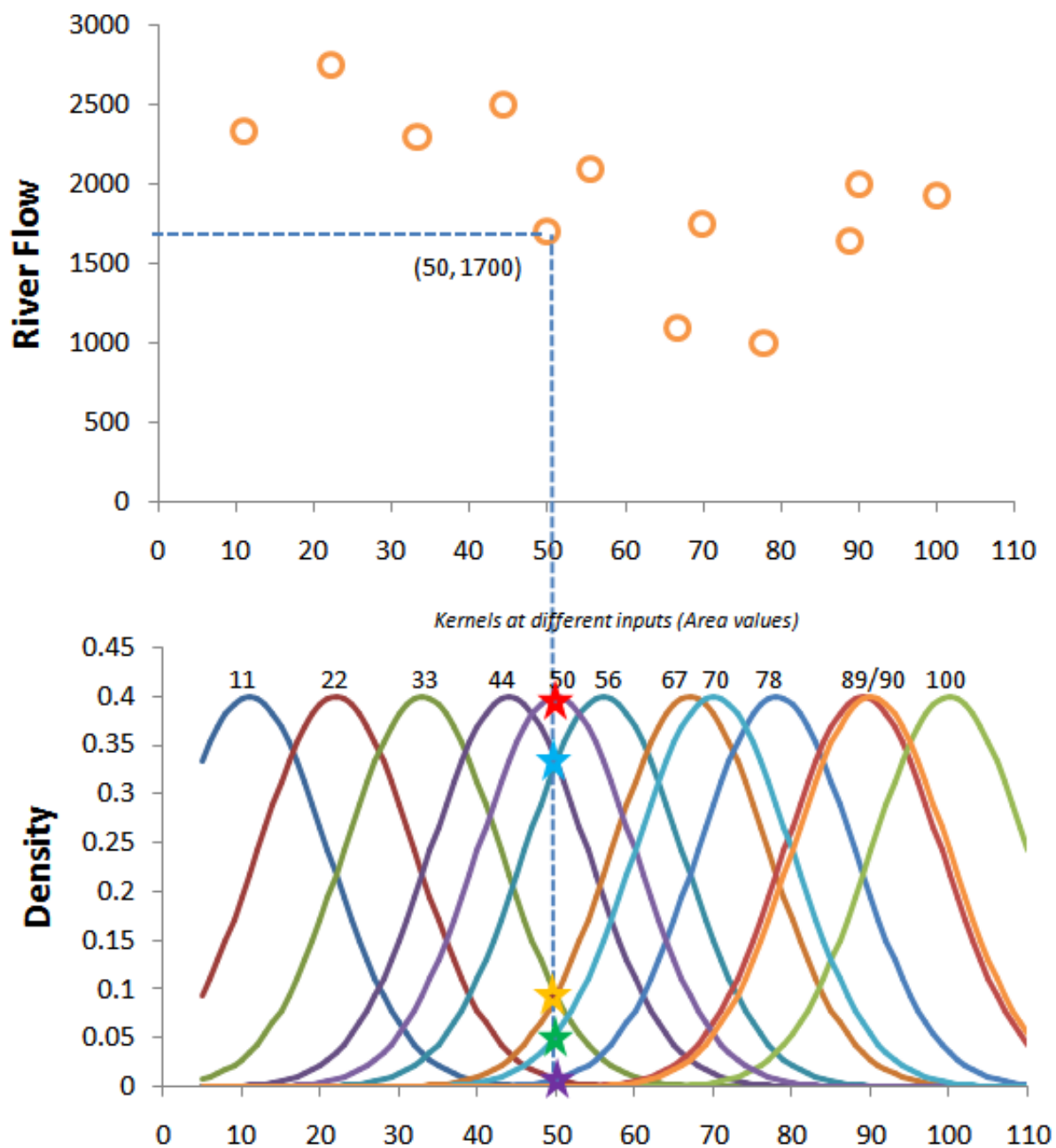


FIGURE 16 – Diagramme de dispersion montrant les données d’entrée-sortie et les noyaux à toutes les entrées en utilisant une bande passante de 10

Intersections	Kernel values
★	K(50)
★	K(44),K(56)
★	K(33),K(67)
★	K(70)
★	K(11),K(22),K(78),K(89),K(90),K(100)

FIGURE 17 – Intersection de la variable prédictive, $zone = 50$ avec les noyaux

Les marques en étoile ci-dessus indiquent les points d’intersection et les valeurs de noyau correspondantes.

Les valeurs du noyau sont échelonnées entre 0 et 1 pour les utiliser comme poids. L'équation suivante est utilisée pour mettre à l'échelle les valeurs du noyau entre 0 et 1.

Où w_i est le poids pour l'entrée i et n est le nombre total de points de données. La sortie y pour $x = 50$ est calculée comme la somme pondérée de toutes les valeurs y observées comme indiqué dans l'équation ci-dessous.

$$W_i = \frac{K_i}{\sum_{i=1}^n K_i} \dots (1, 5)$$

Où y_{11} est le débit de la rivière pour $Area = 11$ et w_{11} est le poids correspondant et il en va de même pour les autres termes de l'équation.

$$Y_{50} = Y_{11} \times W_{11} + Y_{22} \times W_{22} + \dots + Y_{100} \times W_{100} = 2006.40$$

Le tableau suivant montre les étapes du calcul des poids à partir des valeurs du noyau pour l'entrée $x = 50$. La dernière colonne du tableau fournit le résultat final, c'est-à-dire le débit fluvial calculé pour une zone x_j miles carrés.

A	B	C	D	E	F
Kernels	Density	Col-B/Sum of Col-B = weight	River flow_Data	Col-D × Col-C	Sum of Col-F = Ymodel

FIGURE 18 – Les étapes de calcul

De même, pour toutes les valeurs de x , les poids sont estimés comme mentionné dans le tableau ci-dessus et utilisés pour calculer les valeurs correspondantes de y . Les valeurs prévues du débit de la rivière pour les différents apports de surface sont indiquées dans le tableau ainsi que dans le graphique ci-dessous.

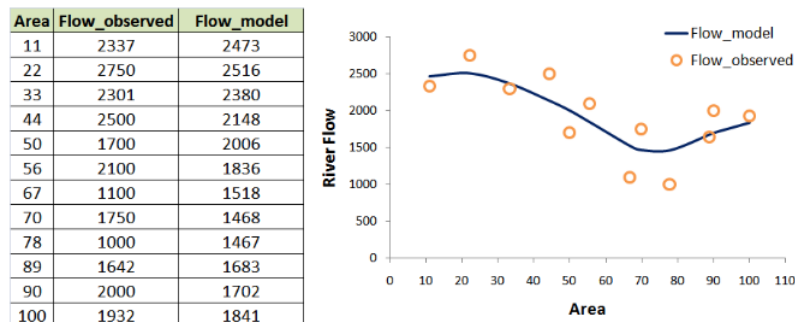


FIGURE 19 – Débit de rivière calculé à partir de la régression des noyaux

Nous allons effectuer la régression sur nos données en s'appuyant sur une fonction écrite en R, comme le montre la figure ci-dessous :

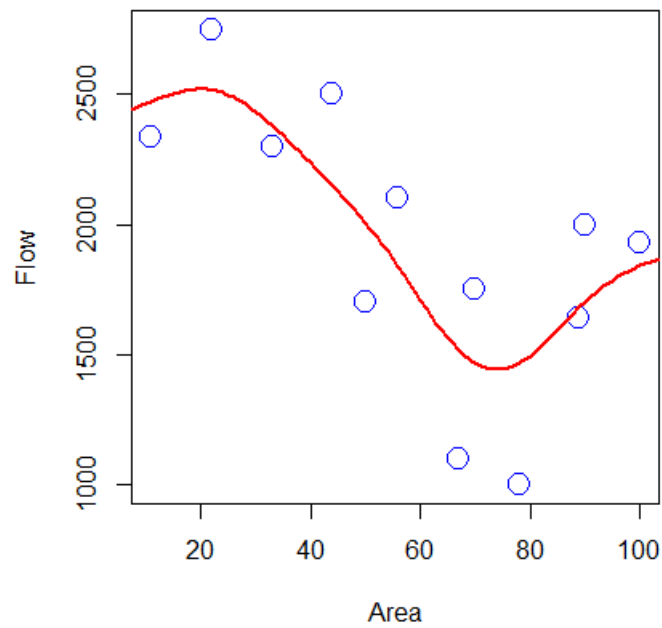


FIGURE 20 – Une régression effectuée sur les données par noyau

Maintenant, nous allons mettre en pratique l'estimateur de Nadaraya Watson :

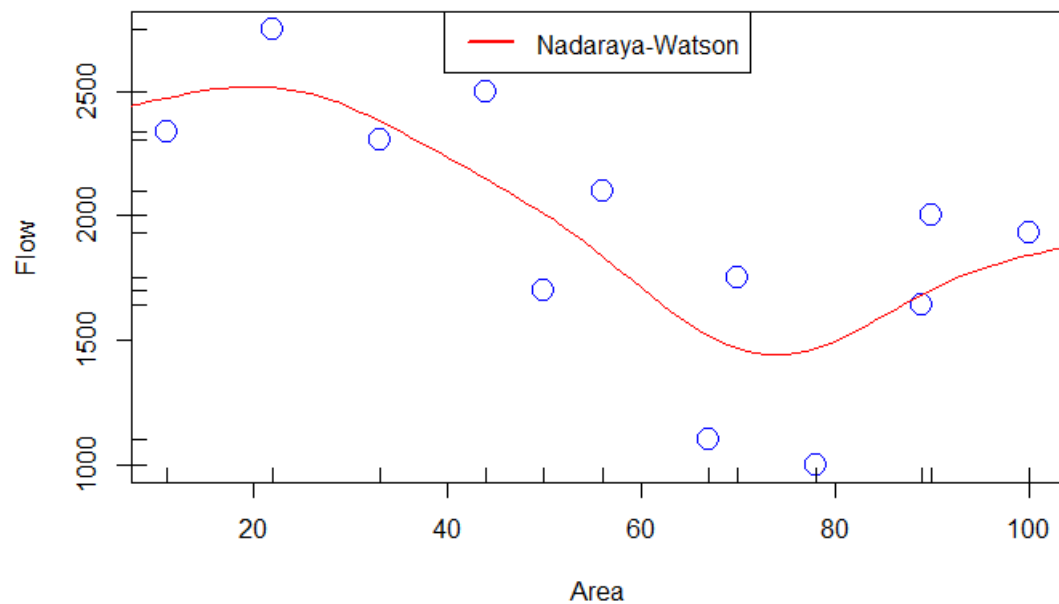


FIGURE 21 – Débit de rivière calculé à partir de l'estimateur NW

En le comparant avec le graphe de la régression par le noyau gaussien, ils sont presque identiques. (la différence n'est pas significative). En quelques sortes nous pouvons confirmer que l'estimateur de Nadaraya Watson admet une bonne précision pour un paramètre $h = 10$.

Le calcul ci-dessus est basé sur une valeur de bande passante de 10. La largeur de bande du noyau influence les prédictions avec une énorme marge. Dans le but de montrer l'importance du choix de parametre nous avons effectué la regression par la kernel pour des valeurs differentes de h : les figures ci-dessous l'illustrent :

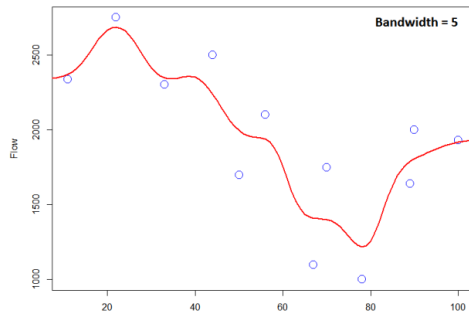


FIGURE 22 – Sortie de la régression du noyau pour une valeur de $h = 5$

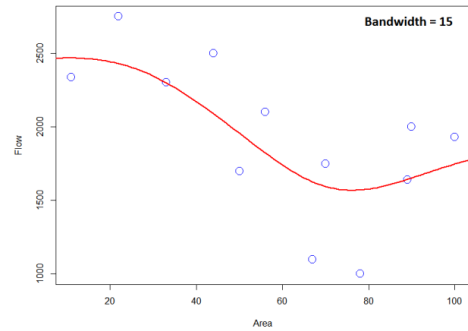


FIGURE 23 – Sortie de la régression du noyau pour une valeur $h = 15$

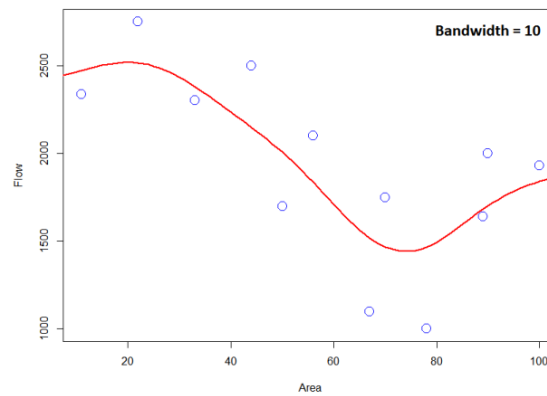


FIGURE 24 – Sortie de la régression du noyau pour une valeur $h = 10$

Une faible valeur de la bande passante est trop importante pour les données en raison du noyau plus étroit qui donne un poids élevé au " y_i " de " x_i " dans l'ensemble de données d'entraînement lorsqu'il est prévu pour x_i . Une faible largeur de bande produit une grande variance dans la sortie qui ne semble pas réaliste, alors qu'une valeur élevée de largeur de bande lisse la sortie, ce qui ne révèle pas la véritable relation entre l'entrée et la sortie. Par conséquent, pour obtenir un compromis entre biais et variance, la largeur de bande doit être optimisée dans la régression du noyau.

4.2 Conclusion

Dans cette section nous avons appliqués la régression non paramétrique avec les noyaux sur un échantillon des données réelles. Pour plusieurs paramètres de lissage h . La méthode des noyaux nous a permis de calculer le débit de la rivière en chaque entré x .

4.2.1 Code-R de la régression du noyau(9)

```
MNW ← function(x, X, Y, h, K = dnorm)
```

Arguments

x : evaluation points

X : vector (size n) with the predictors

Y : vector (size n) with the response variable

h : bandwidth

K : kernel

Matrix of size n x length(x)

```
Kx ← sapply(X, function(Xi)K((x - Xi)/h)/h)
```

Weights

```
W ← Kx/rowSums(Kx)
```

Means at x ("drop" to drop the matrix attributes)

```
drop(W * Y)
```

```
xGrid ← seq(5, 110, 1)
```

```
h ← 10
```

```
ykernel ← NULL
```

```
for(xesti in xGrid)
```

```
xx ← xesti - x
```

```
K ← gaussianKernel(xx, b)
```

```
Ksum ← sum(K)
```

```
weight ← K/Ksum
```

```
yk ← sum(weight * y)
```

```
xkyk ← c(xesti, yk)
```

```
ykernel ← rbind(ykernel, xkyk)
```

```
n ← 24
```

Kernel regression

```
data ← data.frame(Area = c(11, 22, 33, 44, 50, 56, 67, 70, 78, 89, 90, 100)
```

```
RiverFlow = c(2337, 2750, 2301, 2500, 1700, 2100, 1100, 1750, 1000, 1642, 2000, 1932))
```

```
X ← data$Area; Y ← data$RiverFlow
```

function to calculate Gaussian kernel

```
gaussianKernel ← function(x, h)
```

```
K <- -(1/((sqrt(2 * pi)))) * exp(-0.5 * (x/h)2)
```

```
return(K)
```

Estimation de l'esperance exponentielle :(Nadariya Waston) :

```
ykernel ← NULL
```

```
plot(X, Y, xlab = "Area", ylab = "Flow", col = "blue", cex = 2)
```

```
rug(X, side = 1); rug(Y, side = 2)
```

```
lines(xGrid, (xGrid), col = "red")
```

```
lines(xGrid, MNW(x = xGrid, X = X, Y = Y, h = h), col = "red", )
```

```
legend("top", legend = c("Nadaraya - Watson", lwd = 2, col = 2)
```


5 Conclusion générale

Dans ce travail, nous avons présenté deux méthodes les plus connues d'estimation non paramétrique de la densité de probabilité (Histogramme et Noyau). Ensuite nous avons étudié la régression non paramétrique par noyau (dans le cadre uni-dimensionnel) avec la théorie entourant, l'étude des propriétés de convergence des estimateurs qui en résultent. Afin de clôturer ce travail avec une application sur la régression par la kernel. Les perspectives qui en découlent lors de cette étude :

- Les méthodes d'estimation non paramétriques nous permis d'estimer la fonction de densité f inconnue selon les informations provenant des données.
- L'impact du choix du paramètre de lissage sur la convergence de l'estimateur, mais aussi sur les résultats de la prédiction.
- L'application nous a démontré que la méthode de noyau est utilisée dans un objectif prédictif.

Mots clés : Estimation non paramétrique, Histogramme, Noyau, Régression, Nadaraya Watson.

6 Abstract

In this work, we presented two of the best known methods for nonparametric probability density estimation (Histogram and Kernel). Then we studied nonparametric kernel regression (in the one-dimensional framework) with the surrounding theory, studying the convergence properties of the resulting estimators. In order to close this work with an application on kernel regression. The perspectives which result from this study :

- The nonparametric estimation methods allow us to estimate the unknown density function f according to the information from the data.
- The impact of the choice of the smoothing parameter on the convergence of the estimator, but also on the prediction results.
- The application has shown us that the kernel method is used for a predictive purpose.

Keywords : Nonparametric estimation, Histogram, Kernel, Regression, Nadaraya Watson.

7 Références

Références

- [1] Ph.Frédéric Richard Cours sur la Régression , Amétice,2019-2020.
- [2] Ph.Pudlo Piere Cours de Statistiques», Amétice, 2019-2020.
- [3] Catherine Matias Introduction à la statistique non paramétrique, CNRS, evry, 2014.
- [4] « Dr.NATHALIE VANDAL » Thèse sur La régression non paramétrique
Théorie et Application à une étude portant sur la densité mammaire 2005.
- [5] Google, Cours STAT 2150 "Statistique non paramétrique : "Méthodes de lissage",
Année académique 2008 - 2009
- [6] Revue, L.LÉGERJ. J. DAUDIN Étude d'un modèle de régression non paramétrique.
- [7] Wikipédia « <https://fr.wikipedia.org/wiki/Statistiques>.»
- [8] « Ph.D Niranjan Pramanik», <https://towardsdatascience.com/kernel-regression-made-easy-to-understand-86caf2d2b844>
- [9] Kernel regression estimation, Google Scholar
- [10] Kaggle.
- [11] François Husson, Introduction à l'estimation non paramétrique. 2016.
- [12] Google, <http://cedric.cnam.fr/vertigo/Cours/ml/coursEstimationDensite.html>