

Régression non-paramétrique avec erreurs dans les variables

Youssef Baccar & Yasmine Rekik & Amine Sefraoui

April 30, 2023

Contents

1	Introduction et objectif du mémoire	3
2	Introduction au problème	4
2.1	Problème de déconvolution	4
2.2	Problèmes de régression non-paramétrique avec erreurs dans les variables	4
2.3	Estimateurs à noyaux pour la déconvolution de densité	6
3	Méthode à noyaux pour la régression non-paramétrique avec erreurs dans les variables	8
3.1	Modèle sans erreurs dans les variables	9
3.2	Modèle avec erreurs dans les variables	10
3.3	Propriétés asymptotiques	12
3.3.1	Consistance	12
3.3.2	Vitesse de convergence	13
4	Cross Validation, SIMEX	14
4.1	Critère sur la non-convergence (faible) de \hat{m} au taux $(\epsilon_n)_{n \in \mathbb{N}}$	15
5	Applications numériques	16
5.1	Méthode à noyaux avec Python	16
5.2	Préparation des données	20
5.3	Metrique et performance du modèle (train set)	23
5.4	Utilisation d'une autre fonction noyau, qui n'est pas dans \mathbb{L}^1	27
5.5	Convergence optimale	32
5.6	Application à une base de données	36
6	Annexe	40
6.1	Preuve du Lemme 1	40
6.2	Preuve du Lemme 2	40
6.3	Preuve du Théorème 2	42
6.4	Preuve du Théorème 3	49

1 Introduction et objectif du mémoire

Ce travail s'inscrit dans le cadre de notre mémoire de Master 1 portant sur la régression non paramétrique avec des erreurs dans les variables. Dans ce manuscrit nous présentons, dans un premier temps, la méthode de déconvolution qui permet d'estimer une fonction de densité. Celle-ci constitue un pièce maîtresse et nous permettra d'étudier ensuite les méthodes à noyaux (régression non paramétrique) plus en détail.

Afin d'expliquer brièvement l'intérêt de ces méthodes nous en donnons la description suivante : La régression non paramétrique est une forme de prédiction dans laquelle la fonction d'estimation ne prend pas de forme prédéterminée mais est construite selon les données en faisant intervenir l'espérance conditionnelle (a contrario du modèle linéaire, par exemple, où la fonction est affine). Plus concrètement, on se situe dans un cadre où l'on dispose d'une variable d'entrée (explicative) notée X associée à une autre variable de sortie (à expliquer) notée Y , toutes deux recensées pour le même individu dans une base de données de plusieurs observations. Notre but est de parvenir à prédire Y pour une valeur donnée de X , et pour ce faire, on divise l'échantillon de données en deux ensembles. Le premier est celui d'apprentissage qui servira à l'estimation, et le second celui de validation dont le but est de vérifier le niveau d'exactitude de ces prédictions en les comparant aux vraies valeurs observées.

Ces méthodes interviennent dans des domaines d'études divers et variés tels que :

- **L'économie** : Souvent utilisée pour estimer des fonctions de production et des courbes de demande, par exemple pour expliquer la relation entre le prix d'un produit Y et la demande des consommateurs X .
- **L'écologie** : Pour évaluer la relation entre la température de l'eau et la densité des populations de poissons dans un lac.
- **La finance** : Une étude pourrait utiliser la régression non paramétrique pour exprimer la relation entre le taux d'intérêt et le prix des obligations

Ou encore l'épidémiologie, la sociologie ou l'agriculture, etc...

Dans cet ouvrage, nous commençons par l'étude théorique de ce modèle dans les trois premières parties, pour ensuite basculer vers une approche pratique avec une application numérique sur une base de données d'un projet immobilier de Casablanca au Maroc, il s'agira de prédire le prix d'un bien immobilier connaissant sa superficie.

Dans l'ensemble du manuscrit, E dénotera l'espérance et \perp l'indépendance.

2 Introduction au problème

2.1 Problème de déconvolution

Expliquons d'abord ce que c'est qu'un problème de déconvolution: Notre objectif est d'estimer une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ que l'on suppose dans $\mathbb{L}^1(\mathbb{R})$ alors que tout accès empirique est limité à une certaine quantité h qui est le produit de convolution de f et d'une certaine densité de probabilité g que l'on appelle densité d'erreur.

$$h(x) = f * g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy, \quad x \in \mathbb{R} \quad (1)$$

Par conséquent, f ne peut être estimée à partir de certaines observations qu'indirectement. La stratégie est donc d'estimer d'abord h puis d'appliquer une procédure de déconvolution sur la version empirique \hat{h} de h pour estimer f .

Cependant, pour pouvoir faire cela, nous devons supposer que la densité d'erreur g est connue. Bien que cette condition ne soit pas toujours réaliste en pratique, une connaissance complète de g sera supposée dans toute la suite.

De plus, les méthodes de Fourier sont très utiles ici. En effet, la transformation de Fourier d'une fonction g intégrable est défini par :

$$g^{ft}(t) = \int_{\mathbb{R}} e^{itx} g(x) dx$$

Elle change ainsi le produit de convolution en simple multiplication. Donc, plus concrètement:

$$(1) \Leftrightarrow h^{ft} = f^{ft} \cdot g^{ft}$$

Ainsi, comme schéma approximatif pour la construction d'estimateurs de déconvolution nous donnons:

1. En premier lieu, on estime h^{ft} à partir des informations empiriques directes, on obtient alors $\widehat{h^{ft}}$
2. Puis, on calcule $\widehat{h^{ft}}(t)$ et, en supposant g^{ft} non nul, on le divise par $g^{ft}(t)$. On obtient donc un estimateur $\widehat{f^{ft}}(t)$. Notez que g^{ft} est calculable car g est connu. De plus, on suppose que si $\widehat{h^{ft}}(t) \neq 0$ alors $g^{ft}(t) \neq 0$
3. Enfin, on régularise $\widehat{f^{ft}}$ pour que sa transformation de Fourier inverse existe (Par le théorème d'inversion de Fourier) et on prend \hat{f} comme l'estimateur de déconvolution de f .

2.2 Problèmes de régression non-paramétrique avec erreurs dans les variables

Dans le cadre de notre projet, nous nous intéressons à une application particulière des problèmes de déconvolution qui est la régression non paramétrique avec erreurs dans les variables:

On suppose que l'on observe des réalisations de n couples de variables $(X_1, Y_1), \dots, (X_n, Y_n)$ i.i.d. de même loi que (X, Y) et on souhaite étudier le lien entre X et Y .

Les Y_i sont les variables à expliquer, les X_i sont les covariables ou le design et on se place dans le cadre d'un design aléatoire. On suppose également que la variable Y est intégrable. Il se trouve que la fonction de régression

$$m(x) = \mathbb{E}(Y_j \mid X_j = x) \quad (2)$$

est le meilleur prédicteur de Y_j basé sur la covariable X_j , au sens \mathbb{L}^2 .

En effet, on peut facilement montrer le résultat suivant:

Lemme 1.

$$E |Y_j - m(X_j)|^2 \leq E |Y_j - f(X_j)|^2 \quad (3)$$

Pour toute fonction mesurable $f: \mathbb{R} \rightarrow \mathbb{R}$ avec $E(f^2(X_j)) < \infty$.

Preuve. Ce lemme sera démontré en annexe. □

On obtient donc le modèle de régression suivant :

$$Y_j = m(X_j) + \epsilon_j, \quad j = 1, \dots, n \quad \text{où} \quad E(\epsilon_j \mid X_j) = 0$$

Les variables aléatoires ϵ_j sont les erreurs de régression.

De plus, parmi nos hypothèses, on suppose que les X_1, \dots, X_n sont continuellement distribuées et ont une densité (au sens de Lebesgue) f_X dite densité de design.

On considère $x \in \mathbb{R}$ tel que $f_X(x) \neq 0$ et notre but est d'estimer:

$$m(x) = \mathbb{E}(Y \mid X = x) = \int y f_{Y|X=x}(y) dy = \int y \frac{f_{(X,Y)}(x, y)}{f_X(x)} dy$$

(d'après la formule de Bayes) à partir des données $(X_1, Y_1), \dots, (X_n, Y_n)$ et où $f_{(X,Y)}$ est la densité du couple (X, Y)

Pour faire cela, on se base sur la loi forte des grands nombres, la quantité à estimer étant une espérance conditionnelle.

Le principe est le suivant, plus la distance entre X_j et x est petite, plus l'observation (X_j, Y_j) est informative pour $m(x)$. C'est pourquoi Y_j doit être pondéré en fonction de la proximité de X_j et x .

On introduit alors l'estimateur de poids de régression général:

$$\hat{m}_0(x) = \sum_{j=1}^n Y_j \cdot P_{n,j}(x, X_1, \dots, X_n)$$

$$\text{où } \sum_{j=1}^n P_{n,j}(x, X_1, \dots, X_n) = 1 \quad p.s \quad \forall x \in \mathbb{R}$$

Cependant, dans de nombreuses situations réelles, les covariables X_1, \dots, X_n ne peuvent être observé qu'avec un certain bruit additif indépendant. On change alors le schéma d'observation et on suppose que l'on observe les données i.i.d $(W_1, Y_1), \dots, (W_n, Y_n)$ où :

$$\begin{cases} W_j = X_j + \delta_j \\ Y_j = m(X_j) + \epsilon_j \end{cases} \quad (4)$$

Il s'agit du modèle de régression avec erreurs dans les variables.

Les contaminations des données $(\delta_j)_j$ sont des variables aléatoires i.i.d. qui ont pour densité g supposée connue. De plus, $\delta_j \perp X_j$ et $\delta_j \perp \epsilon_j \forall j$ et donc $\delta_j \perp Y_j \forall j$.

Dans ce cas, pour accéder à m , on doit d'abord estimer la densité de design f_X avec une procédure de déconvolution. C'est pourquoi on aura besoin des estimateurs de déconvolution de densité.

Si on note h la densité des W_j , on a $h = f_X * g = \int f_X(x-y)g(y)dy$ puisque la densité de la somme de deux variables aléatoires indépendantes est le produit de convolution des densités marginales de ces deux variables aléatoires. On a :

$$\begin{aligned} h^{ft}(t) &= \int e^{itx} h(x) dx = \mathbb{E}(e^{itW_1}) \\ &= \psi_{W_1}(t) = \psi_{X_1}(t) \psi_{\delta_1}(t) = f_X^{ft}(t) g^{ft}(t) \end{aligned}$$

où $\psi_{W_1}, \psi_{X_1}, \psi_{\delta_1}$ représentent (respectivement) les fonctions caractéristiques de W_1, X_1, δ_1 .

On peut estimer $\psi_{W_1}(t)$ par $\widehat{\psi_{W_1}}(t) = \frac{1}{n} \sum_{j=1}^n e^{itW_j}$ (LFGN + une transformation mesurable de variables aléatoires iid reste iid).

Ainsi, en supposant que g^{ft} ne s'annule pas, un estimateur sans biais et consistant de $f_X^{ft}(t)$ est donné par :

$$\widehat{f_X^{ft}}(t) = \frac{1}{n} \sum_{j=1}^n \frac{e^{itW_j}}{g^{ft}(t)} \quad (5)$$

Mais on est intéressé par f_X et non f_X^{ft} . On ne peut donc pas directement utiliser l'inversion de Fourier et choisir $\widehat{f_X}(x) = \frac{1}{2\pi} \int e^{-itx} \widehat{\psi_{X_1}}(t) dt$ car on ne sait pas si $\widehat{\psi_{X_1}}$ est intégrable ou de carré intégrable sur \mathbb{R} . On doit donc régulariser $\widehat{\psi_{X_1}}$ avant l'inversion de Fourier et ce, pour également éviter les risques d'overfitting. Pour faire face à ce problème, on introduit les estimateurs à noyaux pour la déconvolution de densité.

2.3 Estimateurs à noyaux pour la déconvolution de densité

La méthode d'estimation par noyaux est une approche qui repose sur la fonction de répartition empirique :

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

Ainsi en supposant que la densité f , que l'on cherche à estimer, est continue alors on a :

$$\begin{aligned} f(x) &\approx F'(x) \\ &= \lim_{b \rightarrow 0} \frac{F(x+b) - F(x)}{b} \\ &= \lim_{b \rightarrow 0} \frac{F(x+b) - F(x-b)}{2b} \end{aligned}$$

Donc pour un b petit (proche de 0) on a :

$$f(x) \approx \frac{F(x+b) - F(x-b)}{2b}$$

d'où la pertinence de l'estimateur :

$$\begin{aligned} \hat{f}_n(x) &\approx \frac{\hat{F}_n(x+b) - \hat{F}_n(x-b)}{2b} \\ &= \frac{1}{2b} \left(\frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x+b} - \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x-b} \right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2b} 1_{]x-b, x+b]}(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2b} 1_{]-1, 1]} \left(\frac{X_i - x}{b} \right) \\ &= \frac{1}{nb} \sum_{i=1}^n \frac{1}{2} 1_{]-1, 1]} \left(\frac{X_i - x}{b} \right) \end{aligned}$$

Dans cet esprit, nous exposons la méthode d'estimation par noyaux pour la déconvolution de densité.

Définition 1. Une fonction $K : \mathbb{R} \rightarrow \mathbb{R}$ est appelée noyau (kernel) si K est intégrable et $\int_{\mathbb{R}} K(y) dy = 1$

On définit alors, en généralisant cette démarche aux différent noyaux, un estimateur à noyau de la densité des W_j dès qu'on se donne un noyau K et un paramètre dit fenêtre $b > 0$ par :

$$\hat{h}(x) = \frac{1}{nb} \sum_{j=1}^n K\left(\frac{x - W_j}{b}\right) \quad (6)$$

Si $K \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ alors $\hat{h} \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. Sa transformation de Fourier existe et on obtient alors :

$$\begin{aligned}
\hat{h}^{ft}(t) &= \int e^{itx} \hat{h}(x) dx \\
&= \int e^{itx} \times \frac{1}{nb} \times \sum_{j=1}^n K\left(\frac{x - X_j}{b}\right) dx \\
&= \frac{1}{nb} \sum_{j=1}^n \int e^{itx} \times K\left(\frac{x - X_j}{b}\right) dx
\end{aligned}$$

On opère le changement de variable : $z = \frac{x - W_j}{b}$ et on obtient

$$\begin{aligned}
\hat{h}^{ft}(t) &= \frac{1}{n} \sum_{j=1}^n \int e^{it(zb + W_j)} K(z) dz \\
&= \left(\frac{1}{n} \sum_{j=1}^n e^{itW_j} \right) \times \int e^{i(tb)z} K(z) dz \\
&= \hat{\psi}_{W_1}(t) \times K^{ft}(tb)
\end{aligned}$$

Ainsi,

$$\widehat{f_X}^{ft}(t) = \frac{\widehat{\psi_{W_1}(t)} K^{ft}(tb)}{g^{ft}(t)}$$

Et on obtient l'estimateur à noyau de déconvolution de densité :

$$\widehat{f_X}(x) = \frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \frac{1}{n} \sum_{j=1}^n \frac{e^{itW_j}}{g^{ft}(t)} dt \quad (7)$$

où g^{ft} ne s'annule pas et K^{ft} à support compact.

Nous pouvons également établir certaines propriétés concernant l'esperance et la variance de notre estimateur.

Lemme 2. Si f_X est bornée et continue, $f_X^{ft} \in \mathbb{L}^1(\mathbb{R})$, $g^{ft}(t) \neq 0 \ \forall t \in \mathbb{R}$, $K \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ et K à support compact. On a $\forall x \in \mathbb{R}$:

- $\mathbb{E}(\hat{f_X}(x)) = (K_b * f_X)(x)$, où $K_b = \frac{1}{b} K(\frac{\cdot}{b})$
- $var(\hat{f_X}(x)) \leq \frac{1}{2\pi} \|f_X * g\|_{\infty} \frac{1}{n} \int \left| \frac{K^{ft}(tb)}{g^{ft}(t)} \right|^2 dt$

Preuve. Ce lemme sera démontré en appendice. □

3 Méthode à noyaux pour la régression non-paramétrique avec erreurs dans les variables

N'oublions pas que notre but est toujours d'estimer la fonction de régression dans le modèle avec erreurs dans les variables (4). Mais avant cela, prenons le modèle avec des variables non contaminées.

3.1 Modèle sans erreurs dans les variables

Nous supposons que les X_j ne sont pas contaminées. Nous avons déjà introduit plus haut l'estimateur de poids de régression général \hat{m}_0 mais nous avons laissé les poids $W_{j,n}$ indéfinis. Comme pour la déconvolution de densité, les méthodes à noyaux sont très utilisées. On introduit alors l'estimateur de Nadaraya-Watson. Pour cet estimateur, en se donnant un noyau K d'ordre 1 (cf. définition 4) et une fenêtre $b > 0$, les poids sont choisis comme suit :

$$W_{j,n}(x; X_1, \dots, X_n) = \frac{K\left(\frac{x-X_j}{b}\right)}{\sum_{k=1}^n K\left(\frac{x-X_k}{b}\right)} \quad (8)$$

En effet, cette intuition nous vient du fait que $E(Y \mid X = x) = m(x) = \int y \frac{f_{(X,Y)}(x,y)}{f_X(x)} dy$, et donc qu'un estimateur potentiel est:

$$\hat{m}(x) = \int y \frac{\widehat{f_{n,(X,Y)}}(x,y)}{\widehat{f_{n,X}}(x)} dy$$

où $\widehat{f_{n,(X,Y)}}(x,y) = \frac{1}{nb^2} \sum_{j=1}^n K\left(\frac{x-X_j}{b}\right) K\left(\frac{y-Y_j}{b}\right)$ estimateur de la densité du couple (X,Y) par la méthode à noyaux.
et $\widehat{f_{n,X}}(x) = \frac{1}{nb} \sum_{j=1}^n K\left(\frac{x-X_j}{b}\right)$ estimateur de la densité marginale de X par la méthode à noyaux.
Donc,

$$\begin{aligned} \hat{m}(x) &= \int y \frac{\widehat{f_{n,(X,Y)}}(x,y)}{\widehat{f_{n,X}}(x)} dy \\ &= \int y \times \frac{1}{b} \times \frac{\sum_{j=1}^n K\left(\frac{x-X_j}{b}\right) \times K\left(\frac{y-Y_j}{b}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{b}\right)} dy \\ &= \frac{1}{\sum_{j=1}^n K\left(\frac{x-X_j}{b}\right)} \times \frac{1}{b} \times \sum_{j=1}^n \int y K\left(\frac{y-Y_j}{b}\right) \times K\left(\frac{x-X_j}{b}\right) dy \end{aligned}$$

On fait le changement de variable $z = \frac{y-Y_j}{b}$

$$\begin{aligned} \hat{m}(x) &= \frac{1}{\sum_{j=1}^n K\left(\frac{x-X_j}{b}\right)} \times \sum_{j=1}^n K\left(\frac{x-X_j}{b}\right) \times \underbrace{\int (zb + Y_j) K(z) dz}_{=Y_j \text{ car } \int K=1 \text{ et } \int zK(z)dz=0} \\ &= \frac{\sum_{j=1}^n Y_j K\left(\frac{x-X_j}{b}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{b}\right)} \end{aligned}$$

On remarque que la somme des poids est bien égal à 1. Ainsi, l'estimateur de Nadaraya-Watson est défini comme suit :

$$\hat{m}(x) = \frac{\sum_{j=1}^n Y_j \cdot K\left(\frac{x-X_j}{b}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{b}\right)} \quad (9)$$

On remarque qu'en multipliant le numérateur et le dénominateur par $\frac{1}{nb}$, on obtient au dénominateur l'estimateur (6) de $f_X(x)$ à partir des données i.i.d. X_1, \dots, X_n .

D'autre part, si on pose $p(x) = m(x)f_X(x)$, on a :

$$\begin{aligned}
E\left(\frac{1}{nb} \sum_{j=1}^n Y_j \cdot K\left(\frac{x - X_j}{b}\right)\right) &= \frac{1}{b} E(Y_1 K\left(\frac{x - X_1}{b}\right)) \\
&= \frac{1}{b} E(E(Y_1 | X_1) \cdot K\left(\frac{x - X_1}{b}\right)) \\
&= \frac{1}{b} \int E(Y_1 | X_1 = t) \cdot K\left(\frac{x - t}{b}\right) f_X(t) dt \\
&= \frac{1}{b} \int m(t) \cdot K\left(\frac{x - t}{b}\right) \cdot f_X(t) dt \\
&= \int m(x - sb) \cdot f_X(x - sb) \cdot K(s) ds \\
&= \int p(x - sb) \cdot K(s) ds
\end{aligned}$$

où on a effectué le changement de variables $s = \frac{x-t}{b}$

En supposant que p est continue bornée, on a que : il existe $M \in \mathbb{R}$ tel que $p(x - sb)K(s) \leq MK(s)$

Or, $\int_{\mathbb{R}} K(y) dy = 1$, donc par convergence dominée :

$$\lim_{b \rightarrow 0} \int p(x - sb)K(s) ds = \int p(x)K(s) ds = p(x) \int K(s) ds = p(x)$$

Donc pour b proche de 0, le numérateur se comporte en moyenne comme

$$p(x) = m(x)f_X(x).$$

3.2 Modèle avec erreurs dans les variables

Construction de notre estimateur :

Considérons désormais les données contaminées $(W_1, Y_1), \dots, (W_n, Y_n)$ du modèle (4). On se propose de reprendre l'estimateur de Nadaraya-Watson (9) en l'adaptant aux données contaminées. Ainsi, on remplace le dénominateur par l'estimateur à noyau de déconvolution de densité (7) en utilisant les données W_1, \dots, W_n . Le dénominateur est donc une version empirique de la densité $f_X(x)$. C'est-à-dire

$$\hat{f}_X(x) = \frac{1}{2\pi n} \sum_{j=1}^n \int \exp(-itx) \exp(itW_j) \frac{K^{ft}(tb)}{g^{ft}(t)} dt$$

D'autre part, le numérateur doit également être changé pour qu'il n'utilise que les données contaminées W_1, \dots, W_n dans sa construction.

On propose de prendre

$$\hat{p}(x) = \frac{1}{2\pi n} \sum_{j=1}^n Y_j \int \exp(-itx) K^{ft}(tb) \frac{\exp(itW_j)}{g^{ft}(t)} dt \quad (10)$$

comme estimateur de p . En effet, on cherche \hat{p} tel que: $(\hat{p} * g)(x) = \frac{1}{nb} \sum_{j=1}^n Y_j \cdot K\left(\frac{x - W_j}{b}\right)$
Et donc tel que :

$$\begin{aligned}
(\hat{p} * g)^{ft}(t) &= \int \exp(itx) \frac{1}{nb} \sum_{j=1}^n Y_j K\left(\frac{x - W_j}{b}\right) dx \\
&= \frac{1}{nb} \sum_{j=1}^n Y_j \int \exp(itx) \cdot K\left(\frac{x - W_j}{b}\right) dx \\
&= \frac{1}{nb} \sum_{j=1}^n Y_j \int \exp(it(zb + W_j)) \cdot K(z) dz \\
&= \frac{1}{n} \sum_{j=1}^n Y_j \exp(itW_j) \int \exp(itbz) \cdot K(z) dz \\
&= \frac{1}{n} \sum_{j=1}^n Y_j \exp(itW_j) \cdot K^{ft}(tb)
\end{aligned}$$

Ainsi,

$$\hat{p}^{ft}(t) = \frac{1}{n} \sum_{j=1}^n Y_j \cdot \exp(itW_j) \frac{K^{ft}(tb)}{g^{ft}(t)}$$

et donc:

$$\hat{p}(x) = \frac{1}{2\pi} \int \exp(-itx) \frac{1}{n} \sum_{j=1}^n Y_j \exp(itW_j) \frac{K^{ft}(tb)}{g^{ft}(t)} dt$$

ou encore,

$$\hat{p}(x) = \frac{1}{2\pi n} \sum_{j=1}^n Y_j \int \exp(-itx) K^{ft}(tb) \frac{\exp(itW_j)}{g^{ft}(t)} dt$$

Puis, comme $\hat{m}(x) = \frac{\hat{p}(x)}{\hat{f}_X(x)}$, on en déduit l'estimateur final suivant :

$$\hat{m}(x) = \frac{\sum_{j=1}^n Y_j \int \exp(-itx) \int \exp(itW_j) \frac{K^{ft}(tb)}{g^{ft}(t)} dt}{\sum_{j=1}^n \int \exp(-itx) \exp(itW_j) \frac{K^{ft}(tb)}{g^{ft}(t)} dt} \quad (11)$$

Espérance de l'estimateur :

On trouve que: $\mathbb{E}[\hat{p}(x)] = \frac{1}{2\pi} \int \exp(-itx) p^{ft}(t) K^{ft}(tb) dt$.
En effet,

$$\begin{aligned}
E(\hat{p}(x)) &= \frac{1}{2\pi n} E \left(\sum_{j=1}^n Y_j \int \exp(-itx) K^{ft}(tb) \frac{e^{itW_j}}{g^{ft}(t)} dt \right) \\
&= \frac{1}{2\pi} E \left(Y_1 \int \exp(-itx) K^{ft}(tb) \frac{e^{itW_1}}{g^{ft}(t)} dt \right) \quad \text{car les variables sont i.i.d} \\
&= \frac{1}{2\pi} \int e^{-itx} \frac{K^{ft}(tb)}{g^{ft}(t)} E(Y_1 \exp(it(X_1 + \delta_1))) dt \quad \text{D'après le theoreme de Fubini (fonctions mesurable)} \\
&= \frac{1}{2\pi} \int e^{-itx} \frac{K^{ft}(tb)}{g^{ft}(t)} \times \underbrace{E(e^{it\delta_1})}_{=g^{ft}(t)} \times E(Y_1 e^{itX_1}) dt \\
&= \frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \times E(E(Y_1 | X_1) \times e^{itX_1}) dt \\
&= \frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \times \left(\int m(s) \times e^{its} \times f(s) ds \right) dt \\
&= \frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \times p^{ft}(t) dt
\end{aligned}$$

Or, $p^{ft}(t)K^{ft}(tb) = (p * \frac{K(\cdot/b)}{b})^{ft}(t)$, en supposant alors que p^{ft} et K^{ft} bornées, on trouve que $(p * \frac{K(\cdot/b)}{b})^{ft} \in L^1$ d'où,

$$\mathbb{E}[\hat{p}(x)] = (p * K_b)(x)$$

où $K_b = \frac{K(\cdot/b)}{b}$. Et par convergence dominée, on obtient:

$$\lim_{b \rightarrow 0} \mathbb{E}[\hat{p}(x)] = p(x)$$

3.3 Propriétés asymptotiques

Intéressons nous désormais au comportement asymptotique de notre estimateur de régression par noyau de déconvolution (11).

3.3.1 Consistance

On souhaite savoir si notre estimateur est bien consistant, c'est-à-dire si $\hat{m}(x)$ converge vers $m(x)$ lorsque n tend vers $+\infty$.

On définit l'erreur d'estimation quadratique ponctuelle comme suit :

$$R(x) = |\hat{m}(x) - m(x)|^2$$

On admet le théorème suivant :

Théorème 1 (Consistance forte). *Soit $x \in \mathbb{R}$. On considère le problème de régression avec erreurs dans les variables défini dans (4) sous les conditions suivantes:*

- $g^{ft}(t) \neq 0 \quad \forall t \in \mathbb{R}$
- Les fonctions $p = m.f_X$ et f_X sont bornés sur \mathbb{R} et continues en $xp \in \mathbb{L}^1(\mathbb{R})$

- La fonction noyau K continu et bornée satisfait $K \in \mathbb{L}^1(\mathbb{R})$, $\int K = 1$ et $\text{supp}(K^{ft}) = [-1, 1]$
- On choisit une fenêtre b telle que $b = b_n \searrow 0$ et $b \cdot \min_{|t| \leq \frac{1}{b}} |g^{ft}(t)| \geq n^{-\xi}$ pour un certain $\xi \in]0, \frac{1}{2}[$ et n suffisamment grand
- Le moment d'ordre $2s$ de Y_1 existe pour un $s > \frac{1}{1-2\xi}$
- $|m(y)| \leq C_1$ et $f_X(y) \geq C_2 \forall y \in \mathbb{R}$ pour des constantes $C_1, C_2 > 0$

Alors, l'estimateur (11) satisfait

$$\lim_{n \rightarrow +\infty} \hat{m}(x) = m(x) \text{ p.s.}$$

3.3.2 Vitesse de convergence

On s'intéresse maintenant à la vitesse de convergence de $\hat{m}(x)$ vers $m(x)$. Pour cela, on définit la convergence au sens faible.

Définition 2. On dit que $\hat{m}(x)$ atteint le taux de convergence faible $(\epsilon_n)_{n \geq 0}$, où $(\epsilon_n)_{n \geq 0}$ est une suite qui décroît vers 0 si

$$\lim_{C \rightarrow \infty} \left(\limsup_{n \rightarrow \infty} \sup_{(m, f_X) \in \mathcal{P}} P[|\hat{m}(x) - m(x)|^2 \geq C \cdot \epsilon_n] \right) = 0$$

où $\mathcal{P} = \mathcal{M} \times \mathcal{F}$ est l'ensemble des couples de fonctions de régression et densités design admissibles.

Nous avons également besoin des définitions suivantes :

Définition 3. On dit que p appartient à la classe de Hölder $\mathcal{F}_{\beta, C, \delta, x}$ si $|p(x)| \leq C \forall x \in \mathbb{R}$, p est $\lfloor \beta \rfloor$ fois dérivable et

$$|p^{(\lfloor \beta \rfloor)}(y) - p^{(\lfloor \beta \rfloor)}(\tilde{y})| \leq C |y - \tilde{y}|^{\beta - \lfloor \beta \rfloor} \quad \forall y, \tilde{y} \in [x - \delta, x + \delta]$$

Définition 4. Soit $\ell \in \mathbb{N}^*$. On dit que le noyau K est d'ordre ℓ si $\forall j \in \{1, \dots, \ell\}$, $u \rightarrow u^j K(u)$ est intégrable et $\int u^j K(u) du = 0$

Définition 5. La densité d'erreur g est dite "ordinary smooth" si :

$$C_1(1 + |t|)^{-\alpha} \leq |g^{ft}(t)| \leq C_2(1 + |t|)^{-\alpha} \quad \forall t \in \mathbb{R}$$

Pour $C_2 > C_1 > 0$ et $\alpha > 0$

Définition 6. La densité d'erreur g est dite "supersmooth" si :

$$C_1 \exp(-d_1 |t|^\gamma) \leq |g^{ft}(t)| \leq C_2 \exp(-d_2 |t|^\gamma) \quad \forall t \in \mathbb{R}$$

Pour $C_2 > C_1 > 0$, $0 < d_2 < d_1$ et $\gamma > 0$

Théorème 2. On considère le modèle (4). On suppose que les conditions

$$|m(x)| \leq C_1 \text{ et } f_X(x) \geq C_2 \text{ pour des constantes } C_1, C_2 > 0$$

sont inclus dans la définition de \mathcal{P} .

On suppose également que $f_X(\cdot) \text{Var}(Y_1 | X_1 = \cdot)$ et $m(\cdot)^2 f_X(\cdot)$ sont uniformément bornés sur \mathbb{R} . On suppose aussi que p et f_X satisfont la condition de Hölder pour $\beta > 0$, que $p \in \mathbb{L}^1(\mathbb{R})$. En utilisant l'estimateur (11) avec K borné et continue et qui satisfait $K \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, $\text{supp}(K^{ft}) = [-1, 1]$, K est un noyau d'ordre $\lfloor \beta \rfloor$ tel que $\int |K(z)| |z|^\beta dz < \infty$. On a

- Pour g ordinary smooth, on prend $b \asymp n^{\frac{-1}{2\beta+2\alpha+1}}$ et on a alors :

$$\lim_{C \rightarrow \infty} \left(\limsup_{n \rightarrow \infty} \sup_{(m, f_X) \in \mathcal{P}} P \left[|\hat{m}(x) - m(x)|^2 \geq C \cdot n^{-\frac{2\beta}{2\beta+2\alpha+1}} \right] \right) = 0$$

- Pour g supersmooth, on prend $b \asymp c_b^{-\frac{1}{\gamma}} \log(n)^{-\frac{1}{\gamma}}$, avec $c_b \in]0, \frac{1}{2d_1}[$ et on a alors :

$$\lim_{C \rightarrow \infty} \left(\limsup_{n \rightarrow \infty} \sup_{(m, f_X) \in \mathcal{P}} P \left[|\hat{m}(x) - m(x)|^2 \geq C \cdot \log(n)^{-\frac{2\beta}{\gamma}} \right] \right) = 0$$

Preuve. Ce théorème sera démontré en appendice. □

4 Cross Validation, SIMEX

-Nous avons vu précédemment, que sous certaines hypothèses, on pouvait trouver le paramètre fenêtre optimal b , pour permettre la convergence faible de $\hat{m}(x)$ au taux $(\epsilon_n)_{n \in \mathbb{N}}$. Maintenant, nous introduisons une nouvelle méthode appelée la Cross Validation, qui permet de trouver le meilleur paramètre b de notre modèle. Ce $b > 0$ doit minimiser la quantité suivante:

$$\sum_{j=1}^n (Y_j - \hat{m}_{-j}(X_j))^2 p(X_j)$$

avec p une fonction poids positive, $\hat{m}_{-j}(x)$ est égale à l'estimateur $\hat{m}(x)$ où seules les données $(X_1, Y_1), \dots, (X_{j-1}, Y_{j-1}), (X_{j+1}, Y_{j+1}), \dots, (X_n, Y_n)$ sont prises en compte. C'est à dire:

$$\hat{m}_{-j}(x) = \frac{\sum_{1 \leq l \leq n, l \neq j} Y_l \int \exp(-itx) \exp(itW_l) \frac{K^{ft}(tb)}{g^{ft}(t)} dt}{\sum_{1 \leq l \leq n, l \neq j} \int \exp(-itx) \exp(itW_l) \frac{K^{ft}(tb)}{g^{ft}(t)} dt}$$

On utilise cette stratégie pour éviter qu'il y ait une dépendance entre l'estimateur $\hat{m}(x)$ et la variable X_j .

-Nous allons maintenant voir une autre alternative pour déterminer le b optimal, c'est la méthode SIMEX (SIMulation and EXtrapolation). Etant donnée les $(X_1, Y_1), \dots, (X_n, Y_n)$ et les $W_j = X_j + \delta_j$ introduit auparavant, on va construire de nouvelles données: $W_j^* = W_j + \delta_j^*$ et $W_j^{**} = W_j + \delta_j^{**}$, $j = 1, \dots, n$ où les $\delta_j^*, \delta_j^{**}$ sont indépendants, et suivent la même loi que δ_j de densité g . Et on introduit $\hat{m}^*(x)$ et $\hat{m}^{**}(x)$ les deux estimateurs où l'on remplace les W_j par les W_j^* et W_j^{**} respectivement. La méthode propose de déterminer \hat{b}^* et \hat{b}^{**} qui minimise respectivement les fonctions suivantes:

$$CV^*(b) = \frac{1}{n} \sum_{j=1}^n (Y_j - \hat{m}_{-j}^*(W_j))^2 p(W_j)$$

$$CV^{**}(b) = \frac{1}{n} \sum_{j=1}^n (Y_j - \hat{m}_{-j}^{**}(W_j^*))^2 p(W_j^*)$$

avec toujours les mêmes notations qu'avant, les $\hat{m}_{-j}^*(x)$ et $\hat{m}_{-j}^{**}(x)$, sont les estimateur où la donnée (W_j, Y_j) a été omise. Puis, on propose le b optimal suivant:

$$b^{opt} = \frac{(\hat{b}^*)^2}{\hat{b}^{**}}$$

Cependant, ces deux méthodes sont très coûteuses en pratique, une alternative est proposée, c'est d'estimer l'espérance conditionnelle de $CV^*(b)$ et de $CV^{**}(b)$. Puis on minimise chacune d'elle, en définissant respectivement \hat{b}_1 et \hat{b}_2 , les minimiseurs des deux espérances conditionnelles, on propose de même le b optimal suivant:

$$b^{opt} = \frac{(\hat{b}_1)^2}{\hat{b}_2}$$

4.1 Critère sur la non-convergence (faible) de \hat{m} au taux $(\epsilon_n)_{n \in \mathbb{N}}$

Théorème 3. Dans le cadre de notre modèle de régression, si nos échantillons $(X_1, Y_1), \dots, (X_n, Y_n)$ suivent la loi du couple (X, Y) de densité:

$$f_{(X,Y),\theta}(x, y) = f_1(x)f_2(x) + \theta C \Delta_X(x) \Delta_Y(y),$$

avec $\theta \in \{0, 1\}$. Et,

$$f_0(x) = \frac{1 - \cos(x)}{\pi x^2}, \quad f_1(x) = \frac{1}{\pi(1 + x^2)}, \quad f_2(y) = \frac{1}{2} \exp(-|y|),$$

$$\Delta_X(x) = a_n^{-\beta} \cos(2a_n x) f_0(a_n x), \quad \Delta_Y(y) = \frac{1}{4} \text{sign}(y) \exp(-|y|)$$

avec $(a_n)_n \rightarrow +\infty$, $\beta \geq 1$, et que

$$\|\tau^2\| = \|\text{var}(Y_j^2 | X_j = \cdot)\|_\infty \leq \infty$$

alors il n'existe aucun estimateur de $m(x)$, basé sur les données $(W_1, Y_1), \dots, (W_n, Y_n)$, tel que:

$$\lim_{C \rightarrow +\infty} (\limsup_{n \rightarrow +\infty} \sup_{(m, f_X) \in P} P(|\hat{m}(x) - m(x)|^2 \geq C\epsilon_n) = 0$$

SI: Cas 1: g smooth, $|(g^{ft})'(t)| \leq \text{const.} |t|^{-\alpha}$, et $\epsilon_n = O(n^{\frac{-2\beta}{2\beta+2\alpha+1}})$, $a_n = n^{\frac{1}{(2\beta+2\alpha+1)}}$

OU

Cas 2: g supersmooth, $|(g^{ft})'(t)| \leq \exp(-d|t|^\gamma)$, $d > 0$, et $\epsilon_n = O((\log(n))^{-2\beta/\gamma})$, $a_n = c.(\log(n))^{\frac{1}{\gamma}}$, avec c suffisamment grand.

Preuve. Ce théorème sera démontré en annexe. □

5 Applications numériques

5.1 Méthode à noyaux avec Python

```
import numpy as np
import pandas as pd
import scipy as sp
import matplotlib.pyplot as plt
import sympy as smp
from tqdm import tqdm
from sklearn.model_selection import train_test_split
```

Contexte: On a à notre disposition des données $(X_j, Y_j)_{j=1, \dots, n}$, avec Y_j les données de sortie et X_j les données d'entrée qui suivent une loi continue.

Etant donnée un échantillon iid $(\delta_j)_{j \geq 1}$ suivant une loi continue, on va modifier notre jeux de données comme suit:

On va considérer $(Y_j, W_j)_{j=1, \dots, n}$ avec $W_j = X_j + \delta_j$, qu'on appelle des données contaminées, on suppose que ce sont les données d'entraînement de notre modèle

Et étant donnée un $x \in \mathbb{R}$, le modèle renvoie la valeur de sortie suivante:

$$\hat{m}(x) = \frac{\sum_{i=1}^n Y_i \int_{\mathbb{R}} \exp(-itx) \exp(itW_i) \frac{K^{ft}(tb)}{g^{ft}(t)} dt}{\sum_{i=1}^n \int_{\mathbb{R}} \exp(-itx) \exp(itW_i) \frac{K^{ft}(tb)}{g^{ft}(t)} dt}$$

avec $b > 0$ appelé le paramètre fenêtre, K la fonction noyau, et g la densité des δ_j .

Avant de commencer à simuler le modèle, nous allons exposer une propriété primordiale, pour permettre une bonne efficacité de l'estimateur, il s'agit de la consistance de l'estimateur.

Si toutes ces conditions sont validées:

- $K \in L_1$, continue, bornée
- $\int K = 1$, $\text{supp}(K^{ft}) = [-1, 1]$
- $b = b_n \downarrow 0$
- $b \cdot \min_{|t| \leq \frac{1}{b}} |g^{ft}(t)| \geq n^{-\xi}$, avec $\xi \in [0, 1/2]$

Alors, on a que $\hat{m}(x) \xrightarrow[n \rightarrow +\infty]{} m(x)$ p.s

Pour respecter toutes ces conditions, on va choisir $K(x) = \frac{1 - \cos(x)}{\pi x^2}$. En effet, $K \in L_1$, car

$$\int_{\mathbb{R}} \left| \frac{1 - \cos(x)}{x^2} \right| dx = \int_0^1 \frac{1 - \cos(x)}{x^2} dx + \int_1^\infty \frac{1 - \cos(x)}{x^2} dx$$

Or, $1 - \cos(x) \leq \frac{x^2}{2}, \forall x \in \mathbb{R}$

Donc,

$$\int_0^1 \frac{1 - \cos(x)}{x^2} dx \leq \int_0^1 \frac{x^2/2}{x^2} dx = \frac{1}{2} < +\infty$$

Et d'autre part,

$$\int_1^\infty \frac{1 - \cos(x)}{x^2} dx \leq \int_1^\infty \frac{1}{x^2} dx < +\infty,$$

car $1 - \cos(x) \leq 1$ Alors on a bien $K \in L_1$. Puis on remarque que K est continue sauf en 0, mais on peut prolonger la fonction par continuité en 0, puisque

$$1 - \cos(x) \underset{x \rightarrow 0}{\sim} \frac{x^2}{2}$$

Donc

$$\lim_{x \rightarrow 0} \frac{1 - \cos(x)}{x^2} = \frac{1}{2}$$

Puis, $\frac{1 - \cos(x)}{x^2} \leq \frac{1}{2}$, en réutilisant l'inégalité $1 - \cos(x) \leq \frac{x^2}{2}, \forall x \in \mathbb{R}$, donc K est bornée. On peut vérifier que

$$\int_{\mathbb{R}} \frac{1 - \cos(x)}{x^2} dx = \pi$$

grâce à un résultat d'analyse complexe (Théorème des résidus). D'où, $\int K = 1$. De plus, on peut vérifier facilement que

$$\frac{1}{2\pi} \int_{-1}^1 \exp(-itx)(1 - |x|)dx = \frac{1 - \cos(x)}{\pi x^2} = K(x)$$

D'où, $K^{ft}(x) = (1 - |x|)_+$, et donc K^{ft} est à support dans $[-1, 1]$

Maintenant nous allons nous intéresser à la loi des $(\delta_j)_{j \geq 1}$. Au tout début, on avait opté pour la loi normale $N(0, 1)$, car c'est une loi classique pour modéliser l'erreur. Cependant, en termes de temps d'exécution et de précision du programme, ce n'est la loi la plus adéquate. Nous allons expliquer tout cela. Mais d'abord, nous allons introduire une fonction intégrée de python, appelée quad.

```
from scipy.integrate import quad

def f(x):
    return 3.0*x*x+1.0
I,err= quad(f,0,1)
I,err
```

Output : (2.0, 2.220446049250313e-14)

La fonction quad permet de calculer l'intégrale d'une fonction, par exemple la commande quad(f,0,1) permet de calculer l'intégrale f entre de 0 et 1, la fonction renvoie un tuple, dont la première valeur est la valeur approximée de l'intégrale et la deuxième valeur représente l'erreur d'approximation. Compte tenu de la fonction noyau K choisie précédemment, il va falloir calculer la quantité suivante pour calculer $\hat{m}(x)$:

$$\int_{-\frac{1}{b}}^{\frac{1}{b}} \frac{\exp(-itx) \exp(itW_j)}{g^{ft}(t)} (1 - |tb|) dt$$

Si on choisit la loi normale $N(0, 1)$ pour les variables d'erreur, on a donc

$$g^{ft}(t) = \exp(-t^2/2)$$

et donc la quantité ci dessus devient:

$$\int_{-\frac{1}{b}}^{\frac{1}{b}} \exp(-itx) \exp(itW_j) \exp(t^2/2)(1 - |tb|)dt = \int_{-\frac{1}{b}}^{\frac{1}{b}} \cos(t(W_j - x)) \exp(t^2/2)(1 - |tb|)dt$$

car cette quantité a sa partie imaginaire nulle (pour s'en convaincre il suffit de voir comment a été construit l'estimateur dans le poly). La meilleure manière de calculer numériquement cette intégrale est d'abord de la calculer directement à la main pour éviter les erreurs d'approximation de la fonction quad, et donc permet d'avoir des prédictions plus précises. Le problème est que l'on n'a pas pu la calculer à la main, ce problème est peut-être insoluble. On va essayer de calculer cette intégrale avec la fonction quad (on choisira par exemple dans le code $x = 1$ et $W_j = 2$), et nous allons voir qu'un autre problème embêtant va apparaître.

```
x=1
w=2 #par exemple
b=0.5 #par exemple
h= lambda t: np.cos(t*(w-x))*np.exp(t**2/2)*(1-np.abs(t*b))
I=quad(h,-2,2) #1/b=1/0.5=2
I
```

Output : (1.6701438027890587, 1.9984014443252818e-14)

On voit qu'on obtient une erreur d'approximation faible, mais maintenant un problème survient lorsque l'on prend un b très petit (on prendra par exemple ici $b = 0.1$).

```
x=1
w=2 #par exemple
b=0.1 #par exemple
h= lambda t: np.cos(t*(w-x))*np.exp(t**2/2)*(1-np.abs(t*b))
I=quad(h,-10,10) #1/b=1/0.1=10
I
```

Output : (-9.86637274211825e+18, 763317865.3125)

On voit ici que la valeur de l'intégrale est de l'ordre de 10^{18} et l'erreur d'approximation est également extrêmement grande, et cela est dû au fait que la fonction $\exp(t^2/2)$ explose très vite, et les programmes de calculs d'intégrale implémentés sur python ne s'avère plus très efficace en termes d'erreur quand une fonction a tendance à exploser très vite et donc on est limité pour prendre des b assez proche de 0, et cela est embêtant, car dans certains cas, et pour obtenir une convergence presque sûr de l'estimateur, on a besoin d'un b très proche de 0.

```
x=1
w=2 #par exemple
b=0.01 #par exemple
h= lambda t: np.cos(t*(w-x))*np.exp(t**2/2)
I=quad(h,-100,100) #1/b=1/0.01=100
I
```

Le code ci-dessus nous renvoie même un warning sur la divergence de l'intégrale car le $b = 0.01$ est assez proche de 0.

Pour contourner ce problème, nous devons donc trouver une loi de probabilité de densité g dont la fonction de caractéristique permet un calcul facile à la main de l'intégrale sus-mentionnée:

$$\int_{-\frac{1}{b}}^{\frac{1}{b}} \frac{\exp(-itx) \exp(itW_j)}{g^{ft}(t)} (1 - |tb|) dt$$

Nous allons donc choisir par exemple la loi exponentielle de paramètre 1 dont la fonction caractéristique est:

$$g^{ft}(x) = \frac{1}{1 - ix}$$

On doit donc calculer l'intégrale suivante:

$$\int_{-\frac{1}{b}}^{\frac{1}{b}} \exp(it(W_j - x))(1 - it)(1 - |tb|) dt$$

$$\text{Or } \exp(it(W_j - x))(1 - it)(1 - |tb|) = (\cos(t(W_j - x)) + i \sin(t(W_j - x)))(1 - it)(1 - |tb|)$$

Comme l'intégrale est à valeur dans \mathbb{R} , on a alors que

$$\begin{aligned} \int_{-\frac{1}{b}}^{\frac{1}{b}} \exp(it(W_j - x))(1 - it)(1 - |tb|) dt &= \int_{-\frac{1}{b}}^{\frac{1}{b}} (\cos(t(W_j - x)) + t \sin(t(W_j - x)))(1 - |tb|) dt \\ \int_{-\frac{1}{b}}^0 (\cos(t(W_j - x)) + t \sin(t(W_j - x)))(1 + tb) dt &+ \int_0^{\frac{1}{b}} (\cos(t(W_j - x)) + t \sin(t(W_j - x)))(1 - tb) dt \end{aligned}$$

En développant les fonctions, et en remarquant que cos est paire, et sin est impaire, on obtient:

$$\begin{aligned} &2 \underbrace{\int_0^{\frac{1}{b}} \cos(t(W_j - x)) dt}_{:=D} - 2b \underbrace{\int_0^{\frac{1}{b}} t \cos(t(W_j - x)) dt}_{:=A} - 2b \underbrace{\int_0^{\frac{1}{b}} t^2 \sin(t(W_j - x)) dt}_{:=C} \\ &+ 2 \underbrace{\int_0^{\frac{1}{b}} t \sin(t(W_j - x)) dt}_{:=B} \end{aligned}$$

On a:

$$\begin{aligned} A &= \frac{\sin(\frac{W_j - x}{b})}{(W_j - x)b} + \frac{1}{(W_j - x)^2} \left(\cos(\frac{W_j - x}{b}) - 1 \right) \\ B &= \frac{\sin(\frac{W_j - x}{b})}{(W_j - x)^2} - \frac{\cos(\frac{W_j - x}{b})}{b(W_j - x)} \\ C &= -\frac{\cos(\frac{W_j - x}{b})}{b^2(W_j - x)} + \frac{2 \sin(\frac{W_j - x}{b})}{(W_j - x)^2 b} + \frac{2}{(W_j - x)^3} \left(\cos(\frac{W_j - x}{b}) - 1 \right) \\ D &= \frac{\sin(\frac{W_j - x}{b})}{(W_j - x)} \end{aligned}$$

Maintenant qu'on a la formule explicite de l'estimateur \hat{m} , on va s'occuper du paramètre b pour obtenir la convergence presque sûr. Les seules conditions qu'ils nous restent sont les suivantes:

- $b = b_n \downarrow 0$
- $b. \min_{|t| \leq \frac{1}{b}} |g^{ft}(t)| \geq n^{-\xi}$, avec $\xi \in [0, 1/2]$

Or, $\forall |t| \leq \frac{1}{b}$, on a $|g^{ft}(t)| = \frac{1}{\sqrt{1+t^2}} \geq \frac{1}{\sqrt{1+\frac{1}{b^2}}}$

Donc, $b. \min_{|t| \leq \frac{1}{b}} |g^{ft}(t)| = \frac{b}{\sqrt{1+\frac{1}{b^2}}}$

Et alors il faut trouver un $b > 0$ tel que:

$$\frac{b}{\sqrt{1+\frac{1}{b^2}}} \geq n^{-\xi}, \xi \in [0, 1/2]$$

Maintenant qu'on a toutes les données nécessaires, on va préparer des données pour voir si le modèle marche bien.

5.2 Préparation des données

```
X=np.random.uniform(0,5,15000)
X[0:20] #Les 20 premiers termes de l'échantillon
```

La variable X de la ligne de code représente la variable d'entrée de la base de données. Ma variable de départ doit suivre une loi de densité continue (ici c'est une uniforme [0,5]).

```
delta=np.random.exponential(1,15000)
delta[0:20] #Les 20 premiers termes

W=X+delta
W[0:20] #Données contaminées
Y=5*X
Y[0:20] #Mon objectif est donc d'approximer la fonction 5x
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.33, random_state=42)
W_train=X_train+delta[0:10050]
W_test=X_test+delta[10050:]
len(X_train),len(y_train),len(X_test),len(y_test)
```

Output : (10050, 10050, 4950, 4950)

La variable W du code représente les données contaminées (on a additionné ici par la loi exponentielle de paramètre 1, nous avons expliqué le choix de cette loi précédemment). Puis la variable Y représente la variable de sortie du dataset. Ici on a voulu poser $Y = 5x$. Donc on s'attend à ce que le modèle prédise la fonction $x \rightarrow 5x$. Puis on a décidé de diviser notre jeu de données: 67% du dataset sera réservé à l'entraînement du modèle, et les 33% seront réservé pour tester le modèle. D'après le output, on voit bien que le nombre de données d'entraînement est de 10050 et celui du test est de 4950.

```
len(W_train),len(W_test),len(delta[10050:])
```

Output : (10050, 4950, 4950)

```
b=0.075#on va prendre ce b qui est proche de 0
b/np.sqrt(1+(1/(b**2)))
```

Output : 0.005609246118122567

Dans le code ci-dessus on calcule, $\frac{b}{\sqrt{1+\frac{1}{b^2}}}$

```
100050**(-1/2.1)
```

Output : 0.004158572156471894

Dans la cellule de code ci-dessus on calcule $n^{-\xi}$, pour $\xi = \frac{1}{2.5} \in [0, 1/2]$, et l'entier n représente le nombre de données du train set qui est de 10050 dans ce cas précis.

On a bien: $\frac{b}{\sqrt{1+\frac{1}{b^2}}} \geq n^{-\xi}$, $\xi \in [0, 1/2]$, et en plus, b est proche de 0, alors tous les critères de convergence sont respectés.

Dans le code ci-dessous, on calcule $\hat{m}(3)$:

```
x=3 #on teste pour une valeur
b=0.075
num=[]
denom=[]
for w,y in list(zip(W_train,y_train)):
    D=np.sin((w-x)/b)/(w-x)
    A=np.sin((w-x)/b)/((w-x)*b)+ (1/((w-x)**2))*(np.cos((w-x)/b)-1)
    C=-np.cos((w-x)/b)/(b**2 *(w-x))+2*np.sin((w-x)/b)/((w-x)**2 *b)
    + (2/((w-x)**3))*(np.cos((w-x)/b)-1)
    B=np.sin((w-x)/b)/((w-x)**2) - np.cos((w-x)/b)/(b*(w-x))
    num.append(y*(2*D-2*b*A-2*b*C+2*B))
    denom.append(2*D-2*b*A-2*b*C+2*B)

sum(num)/sum(denom)
```

Output : 14.559850551087953

Puis on calcule $\hat{m}(x)$ pour $x = 1, 2, 3, 4, 5$

```
L=[1,2,3,4,5]
for x in L:
    b=0.075
    num=[]
    denom=[]
    for w,y in list(zip(W_train,y_train)):
        D=np.sin((w-x)/b)/(w-x)
        A=np.sin((w-x)/b)/((w-x)*b)+ (1/((w-x)**2))*(np.cos((w-x)/b)-1)
        C=-np.cos((w-x)/b)/(b**2 *(w-x))+2*np.sin((w-x)/b)/((w-x)**2 *b)
        + (2/((w-x)**3))*(np.cos((w-x)/b)-1)
        B=np.sin((w-x)/b)/((w-x)**2) - np.cos((w-x)/b)/(b*(w-x))
        num.append(y*(2*D-2*b*A-2*b*C+2*B))
        denom.append(2*D-2*b*A-2*b*C+2*B)
    print(sum(num)/sum(denom))
```

Output :

4.952377611458156
9.973514689240305
14.559850551087953
19.70310406543701
24.416684934201307

On remarque déjà que l'estimateur approxime bien la fonction $x \rightarrow 5x$.

On va maintenant prédire les valeurs du train set pour voir si le modèle fit bien avec nos données d'entraînement.

```
y_pred_train=[]
for i in tqdm(range(10050)):
    b=0.075
    num=[]
    denom=[]
    for w,y in list(zip(W_train,y_train)):
        x=X_train[i]
        D=np.sin((w-x)/b)/(w-x)
        A=np.sin((w-x)/b)/((w-x)*b)+ (1/((w-x)**2))*(np.cos((w-x)/b)-1)
        C=-np.cos((w-x)/b)/(b**2 *(w-x))+2*np.sin((w-x)/b)/((w-x)**2 *b)
        + (2/((w-x)**3))*(np.cos((w-x)/b)-1)
        B=np.sin((w-x)/b)/((w-x)**2) - np.cos((w-x)/b)/(b*(w-x))
        num.append(y*(2*D-2*b*A-2*b*C+2*B))
        denom.append(2*D-2*b*A-2*b*C+2*B)
    y_pred_train.append(sum(num)/sum(denom))
```

On a ainsi obtenu la liste `y_pred_train` qui contient toutes les predictions du train set.

```
y_pred_train[0:20] #Les 20 premières valeurs de la liste
```

Output : [8.407530471936305,

4.567211791802308,
21.808362610152187,
13.482010267089722,
4.097375969866031,
4.815785267934783,
13.751672783604612,
15.52685821401836,
5.239565575497821,
5.038506559577921,
21.847678311995733,
18.025534370628975,
23.855697964047454,
11.087762187742914,
14.533736554857358,
5.287672336106021,

8.210429151543458,
13.462494107815745,
10.450084132911002,
21.728596638773404]

5.3 Metrique et performance du modèle (train set)

On va maintenant calculer les quantités suivantes:

$$MSE = \frac{1}{m} \sum (y_{vrai} - y_{pred})^2$$

$$MAE = \frac{1}{m} \sum |y_{vrai} - y_{pred}|$$

$$R2 = 1 - \frac{\sum (y_{vrai} - y_{pred})^2}{\sum (y_{vrai} - \bar{y}_{vrai})^2}$$

Le $R2$ permet d'évaluer la performance du modèle par rapport au niveau de variation présent dans les données.

En fait, plus le $R2$ est proche de 1, et plus notre modèle fit avec les données.

On va importer du package `sklearn.metrics`, des fonctions déjà implémentée pour calculer ses métriques.

```
from sklearn.metrics import *  
from sklearn.metrics import r2_score  
print("MAE:", mean_absolute_error(y_train, y_pred_train))  
print("MSE:", mean_squared_error(y_train, y_pred_train))  
print("R2:", r2_score(y_train, y_pred_train))
```

Output :

MAE: 0.42637673147875305

MSE: 0.3005265832713948

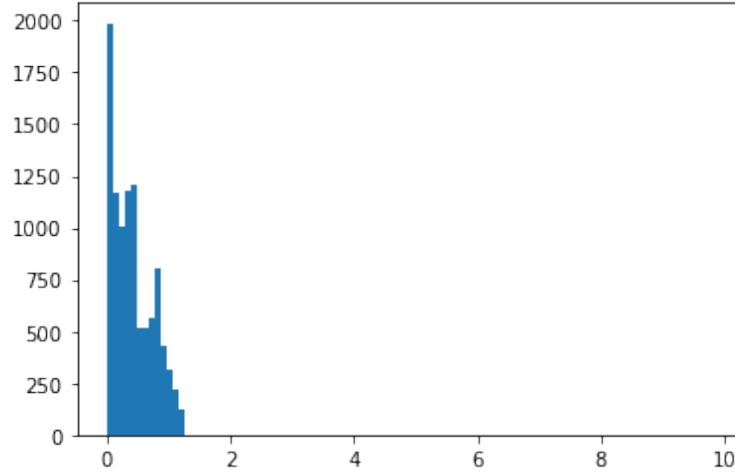
R2: 0.9942468612775656

On a un $R2$ très proche de 1, et donc le modèle est très performant, de même le MAE et le MSE sont très proche de 0.

Maintenant on va proposer un histogramme pour analyser les erreurs du modèle.

Cet histogramme donne le nombre de données en fonction de la valeur de l'erreur (ici égale à $|y_{vrai} - y_{pred}|$).

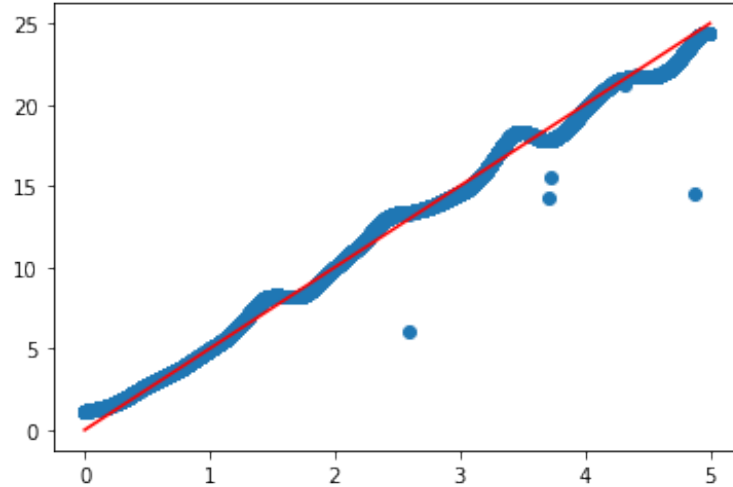
```
err_hist= np.abs(y_train-y_pred_train)  
plt.hist(err_hist, bins=100)  
plt.show()
```



En analysant l'histogramme, on remarque ainsi que au delà d'un écart d'erreur de 2, le nombre de données est quasiment nul, et on remarque que l'immense majorité des données, a une erreur comprise entre 0 et 0.5.

Ensuite, on va tracer nos valeurs prédites en fonction des valeurs d'entrées, on a donc triés par ordre croissant les valeurs d'entrée, et dans le même graphe, on trace la fonction $x \rightarrow 5x$ sur l'intervalle $[0, 5]$, car nos données d'entrée sont à valeurs dans $[0, 5]$.

```
y_pred_train_sort=[]
L=np.argsort(X_train) #contient les indices des valeurs du train par ordre croissant
for i in range(10050):
    y_pred_train_sort.append(y_pred_train[L[i]])
absi=np.linspace(0,5,100)
plt.scatter(np.sort(X_train), y_pred_train_sort)
plt.plot(absi,5*absi,c='r') #Notre modèle a bien prédit la fonction
```

La droite en rouge est le tracé de la fonction $x \rightarrow 5x$, et les points en bleus sont les points dont les abscisses sont les valeurs du X_train c'est à dire de la variable explicative du train set, et les ordonnées sont les prédictions du modèle pour chaque valeur du X_train . Ainsi, on voit bien que l'ensemble des points des prédictions du X_train approxime très bien la fonction $x \rightarrow 5x$. On a alors de bonnes prédictions

Maintenant, on va tester les performances du modèle sur le test set.

```

y_pred_test=[]
for i in tqdm(range(4950)):
    b=0.075
    num=[]
    denom=[]
    for w,y in list(zip(W_train,y_train)):
        x=X_test[i]
        D=np.sin((w-x)/b)/(w-x)
        A=np.sin((w-x)/b)/((w-x)*b)+ (1/((w-x)**2))*(np.cos((w-x)/b)-1)
        C=-np.cos((w-x)/b)/(b**2 *(w-x))+2*np.sin((w-x)/b)/((w-x)**2 *b)
        + (2/((w-x)**3))*(np.cos((w-x)/b)-1)
        B=np.sin((w-x)/b)/((w-x)**2) - np.cos((w-x)/b)/(b*(w-x))
        num.append(y*(2*D-2*b*A-2*b*C+2*B))
        denom.append(2*D-2*b*A-2*b*C+2*B)
    y_pred_test.append(sum(num)/sum(denom))

```

Metrique et performance du modèle (test set) :

```

print("MAE:", mean_absolute_error(y_test, y_pred_test))
print("MSE:", mean_squared_error(y_test,y_pred_test))
print("R2:", r2_score(y_test,y_pred_test))

```

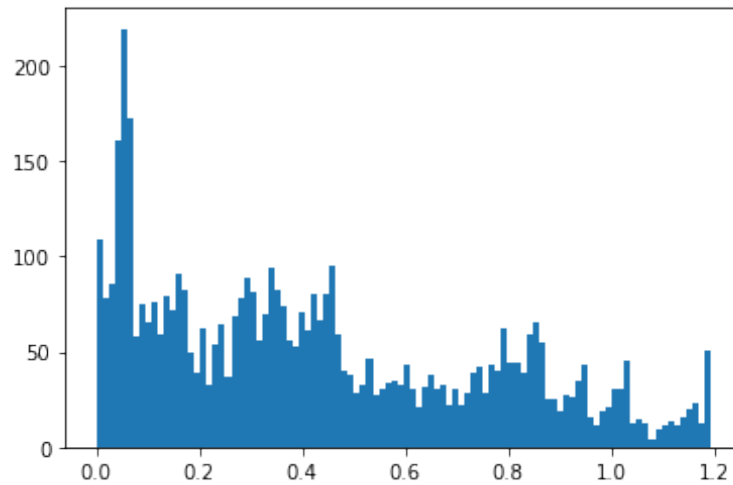
Output :

MAE: 0.4297820637320825

MSE: 0.2902742866314485

R2: 0.9944914140594929

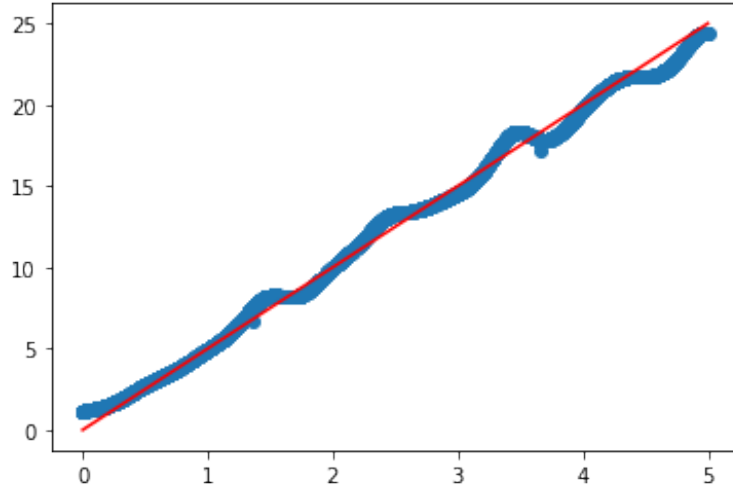
```
err_hist= np.abs(y_test-y_pred_test)
plt.hist(err_hist, bins=100)
plt.show()
```



On voit bien que la très grande majorité des données ont des erreurs très proches de 0.

Maintenant, exactement comme pour le train set, on va comparer nos prédictions à la fonction $x \rightarrow 5x$.

```
y_pred_test_sort=[]
L=np.argsort(X_test) #contient les indices des valeurs du train par ordre croissant
for i in range(4950):
    y_pred_test_sort.append(y_pred_test[L[i]])
absi=np.linspace(0,5,100)
plt.scatter(np.sort(X_test), y_pred_test_sort)
plt.plot(absi,5*absi,c='r') #Notre modèle a bien prédit la fonction
```



Performance très similaire à celle du train set.

5.4 Utilisation d'une autre fonction noyau, qui n'est pas dans \mathbb{L}^1

On va maintenant utiliser la fonction noyau suivante:

$$K(x) = \frac{\sin(x)}{\pi x}, \quad \int K = 1.$$

Les critères de convergence de l'estimateur $\hat{m}(x)$, imposent que K soit dans L_1 . Or nous allons voir que cette fonction noyau n'est pas dans L_1 , mais donne de bien meilleur résultat que la fonction noyau précédente. Puis $x \rightarrow \frac{\sin(x)}{x}$ peut-être prolongée par continuité en 0, car $\frac{\sin(x)}{x} \xrightarrow{x \rightarrow 0} 1$.

De plus $|\sin(x)| \leq |x| \forall x \in \mathbb{R}$, alors K est bornée. La seule condition qu'elle ne vérifie pas, est que $K \in L_1$

Rappel:

Pour une fonction f bien définie, la transformée de Fourier est définie comme suit:

$$f^{ft}(x) = \int_{\mathbb{R}} \exp(itx) f(t) dt$$

et donc, la transformée de Fourier inverse est définie comme suit:

$$(f^{ft})^{-1}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-itx) f(t) dt$$

On a que:

$$\frac{1}{2\pi} \int_{\mathbb{R}} \exp(-itx) 1_{[-1,1]}(t) dt = \frac{1}{2\pi} \int_{-1}^1 (\cos(tx) - i \sin(tx)) dt = \frac{1}{2\pi} \int_{-1}^1 \cos(tx) dt$$

car $t \rightarrow \sin(tx)$ est impaire.

Donc

$$\frac{1}{2\pi} \int_{\mathbb{R}} \exp(-itx) 1_{[-1,1]}(t) dt = \frac{1}{2\pi} \left[\frac{1}{x} \sin(tx) \right]_{-1}^1 = \frac{\sin(x)}{\pi x}$$

On en déduit que $K^{ft}(x) = 1_{[-1,1]}(x)$, et donc $\text{supp}(K^{ft}) = [-1, 1]$.

Dans le même raisonnement qu'avant, on doit calculer l'intégrale suivant:

$$\int_{-\frac{1}{b}}^{\frac{1}{b}} \exp(it(W_j - x))(1 - it) dt$$

Or

$$\exp(it(W_j - x))(1 - it) = (\cos(t(W_j - x)) + i \sin(t(W_j - x)))(1 - it)$$

Comme l'intégrale est à valeur dans \mathbb{R} , on a alors que :

$$\begin{aligned} \int_{-\frac{1}{b}}^{\frac{1}{b}} \exp(it(W_j - x))(1 - it) dt &= \int_{-\frac{1}{b}}^{\frac{1}{b}} \cos(t(W_j - x)) dt + \int_{-\frac{1}{b}}^{\frac{1}{b}} t \sin(t(W_j - x)) dt \\ &= \frac{1}{W_j - x} [\sin(t(W_j - x))]_{-\frac{1}{b}}^{\frac{1}{b}} - \frac{2}{(W_j - x)b} \cos\left(\frac{W_j - x}{b}\right) + \frac{2}{(W_j - x)^2} \sin\left(\frac{W_j - x}{b}\right) \\ &= \frac{2}{W_j - x} \sin\left(\frac{W_j - x}{b}\right) - \frac{2}{(W_j - x)b} \cos\left(\frac{W_j - x}{b}\right) + \frac{2}{(W_j - x)^2} \sin\left(\frac{W_j - x}{b}\right) \end{aligned}$$

On utilise la même densité d'erreur g , et on va prendre cette fois-ci $b = 0.76$, on a choisi ce b par tâtonnement, en calculant $\hat{m}(x)$, pour certains x et pour plusieurs b , on s'est rendu compte que ce dernier donnait les résultats les plus précis, et avec ces conditions on a bien la convergence presque-sûr (à part que $K \in L_1$) .

On calcule $\hat{m}(3)$: (On utilise les mêmes données qu'avant)

```
x=3 #on teste pour une valeur
b=0.76
num=[]
denom=[]
for w,y in list(zip(W_train,y_train)):
    num.append(y*((2/(w-x))*np.sin((w-x)/b)-(1/(w-x))*(2/b)*np.cos((w-x)/b)
    +(2/(w-x)**2)*np.sin((w-x)/b)))
    denom.append((2/(w-x))*np.sin((w-x)/b)-(1/(w-x))*(2/b)*np.cos((w-x)/b)
    +(2/(w-x)**2)*np.sin((w-x)/b))
sum(num)/sum(denom)
```

Output : 15.206322281054584

```
L=[1,2,3,4,5]
for x in L:
    b=0.76
    num=[]
    denom=[]
    for w,y in list(zip(W_train,y_train)):
```

```

num.append(y*((2/(w-x))*np.sin((w-x)/b)-(1/(w-x))*(2/b)*np.cos((w-x)/b)
+(2/(w-x)**2)*np.sin((w-x)/b)))
denom.append((2/(w-x))*np.sin((w-x)/b)-(1/(w-x))*(2/b)*np.cos((w-x)/b)
+(2/(w-x)**2)*np.sin((w-x)/b))

print(sum(num)/sum(denom))

```

Output :

```

4.6727680777776746
9.827207718938233
15.206322281054584
20.405899864865734
24.857035679206113

```

On voit que les valeurs de sortie approxime bien la fonction $x \rightarrow 5x$.

On veut voir maintenant si le modèle fit bien avec le train set.

```

y_pred_train_bis=[]
for i in tqdm(range(10050)):
    b=0.76
    num=[]
    denom=[]
    for w,y in list(zip(W_train,y_train)):
        x=X_train[i]
        num.append(y*((2/(w-x))*np.sin((w-x)/b)-(1/(w-x))*(2/b)*np.cos((w-x)/b)
+(2/(w-x)**2)*np.sin((w-x)/b)))
        denom.append((2/(w-x))*np.sin((w-x)/b)-(1/(w-x))*(2/b)*np.cos((w-x)/b)
+(2/(w-x)**2)*np.sin((w-x)/b))
    y_pred_train_bis.append(sum(num)/sum(denom))

```

Metrique et performance du modèle (train set) :

```

print("MAE:", mean_absolute_error(y_train, y_pred_train_bis))
print("MSE:", mean_squared_error(y_train,y_pred_train_bis))
print("R2:", r2_score(y_train,y_pred_train_bis))

```

Output : MAE: 0.23010521119162217

MSE: 0.0672881860829368

R2: 0.9987118668015916

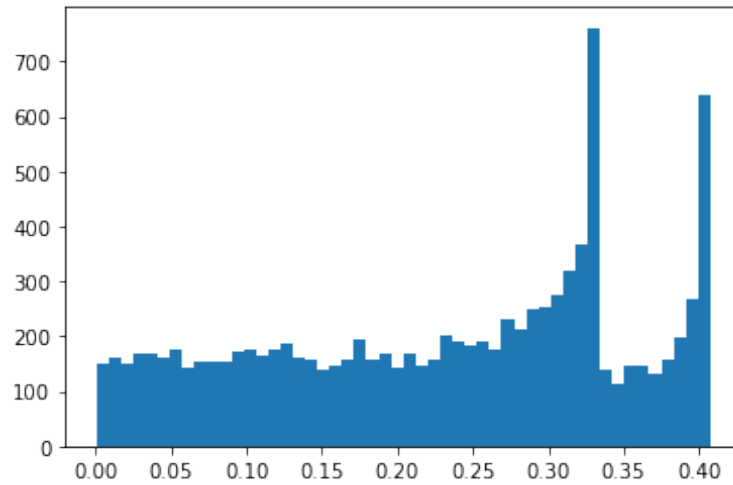
On observe que le MAE et le MSE sont plus proches de 0 en comparaison avec la fonction noyau précédente, de même pour le R2 qui est plus proche de 1.

Puis dans l'histogramme ci-dessous, on remarque que pour toutes les données, il n'y a aucune erreur plus grande que 0.5, toutes les erreurs sont comprises entre $[0, 1/2]$, contrairement dans le cas de la première fonction noyau où pas mal de données avaient une erreur plus grande que 1.

```

err_hist= np.abs(y_train-y_pred_train_bis)
plt.hist(err_hist, bins=50)
plt.show()

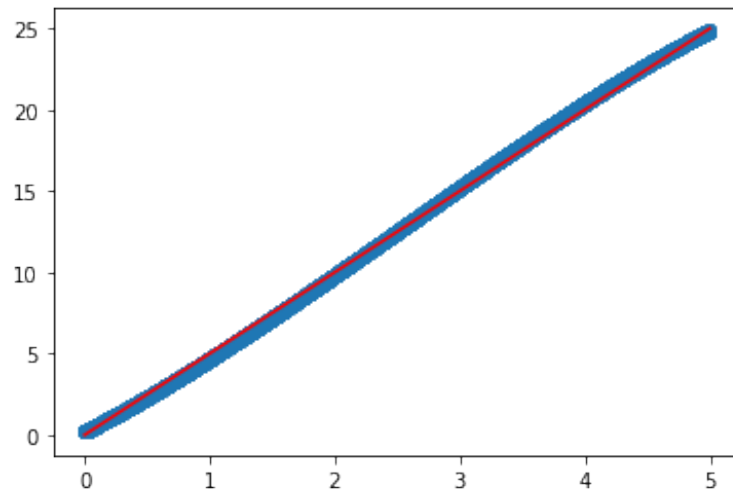
```



```

y_pred_train_sort_bis=[]
L=np.argsort(X_train) #contient les indices des valeurs du train par ordre croissant
for i in range(10050):
    y_pred_train_sort_bis.append(y_pred_train_bis[L[i]])
absi=np.linspace(0,5,100)
plt.scatter(np.sort(X_train), y_pred_train_sort_bis)
plt.plot(absi,5*absi,c='r') #Notre modèle a bien prédit la fonction

```



Dans ce graphe, on voit bien que les prédictions sont plus précises qu'avant, les deux droites sont quasiment superposées.

```

y_pred_test_bis=[]
for i in tqdm(range(4950)):

```

```

b=0.76
num=[]
denom=[]
for w,y in list(zip(W_train,y_train)):
    x=X_test[i]
    num.append(y*((2/(w-x))*np.sin((w-x)/b)-(1/(w-x))*(2/b)*np.cos((w-x)/b)
    +(2/(w-x)**2)*np.sin((w-x)/b)))
    denom.append((2/(w-x))*np.sin((w-x)/b)-(1/(w-x))*(2/b)*np.cos((w-x)/b)
    +(2/(w-x)**2)*np.sin((w-x)/b))
y_pred_test_bis.append(sum(num)/sum(denom))

```

Metrique et performance du modèle (test set) :

```

print("MAE:", mean_absolute_error(y_test, y_pred_test_bis))
print("MSE:", mean_squared_error(y_test,y_pred_test_bis))
print("R2:", r2_score(y_test,y_pred_test_bis))

```

Output : MAE: 0.23279861911562608

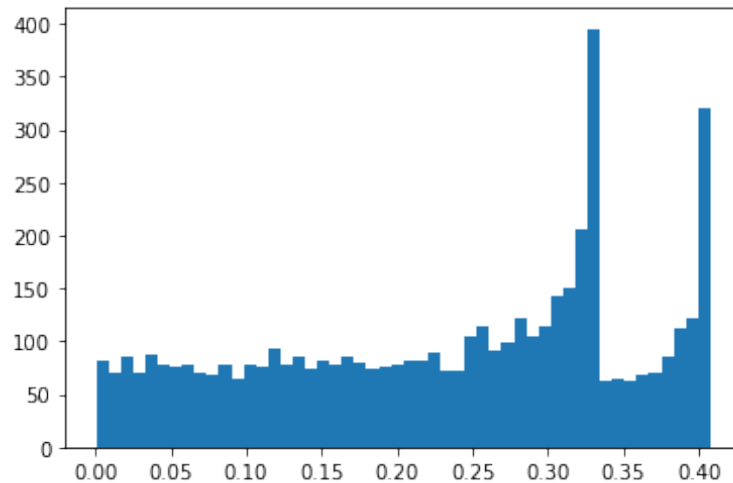
MSE: 0.06845347621435624

R2: 0.9987009464013187

```

err_hist= np.abs(y_test-y_pred_test_bis)
plt.hist(err_hist, bins=50)
plt.show()

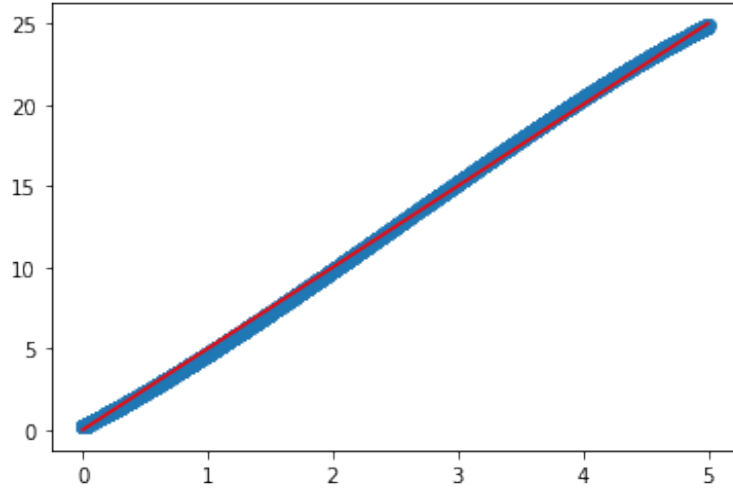
```



```

y_pred_test_sort_bis=[]
L=np.argsort(X_test) #contient les indices des valeurs du train par ordre croissant
for i in range(4950):
    y_pred_test_sort_bis.append(y_pred_test_bis[L[i]])

```



5.5 Convergence optimale

Maintenant nous allons nous intéresser à une propriété intéressante de convergence de \hat{m} , c'est la convergence faible au taux ϵ_n .

On dit que $\hat{m}(x)$ converge faiblement vers $m(x)$ au taux ϵ_n si et seulement si:

$$\lim_{C \rightarrow +\infty} (\limsup_{n \rightarrow +\infty} \sup_{(m, f_X) \in P} P(|\hat{m}(x) - m(x)|^2 \geq C\epsilon_n) = 0$$

Rappelons le théorème 2:

Dans le cadre de notre modèle, on suppose, que $\|\tau^2\| = \|\text{var}(Y_j^2 | X_j = \cdot)\| < +\infty$, et que la fonction noyau K vérifie les conditions suivantes:

- $K \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$
- $\text{supp}(K^{ft}) = [-1, 1]$
- $\int K(z)z^j dz = \delta_{j,0}$, où $\delta_{j,0} = 1$ si $j = 0$, 0 sinon avec $j = 0, \dots, \lfloor \beta \rfloor$, β un réel strictement positif
- $\int |K(z)||z|^\beta dz < +\infty$
- K bornée et continue

Alors

$$\lim_{C \rightarrow +\infty} (\limsup_{n \rightarrow +\infty} \sup_{(m, f_X) \in P} P(|\hat{m}(x) - m(x)|^2 \geq Cn^{-2\beta/(2\beta+2\alpha+1)}) = 0,$$

si g est smooth et $b \approx n^{-1/(2\alpha+2\beta+1)}$, avec β le réel défini ci-dessus, et α le réel strictement positif défini dans la définition d'une fonction smooth.

Il y a une définition analogue pour les fonctions supersmooth, mais ici on utilisera seulement la fonction smooth utilisée précédemment.

Cette propriété de convergence permet une convergence plus rapide de l'estimateur vers $m(x)$.

Nous allons vérifier cela.

Nous allons utiliser g la densité de l'exponentielle de paramètre 1.

On a

$$c_1 \cdot |t|^{-1} \leq |g^{ft}(t)| = \frac{1}{|1 - it|} = \frac{1}{\sqrt{1 + t^2}} \leq c_2 \cdot |t|^{-1},$$

avec $c_1, c_2 > 0$ On a donc $\alpha = 1$

Maintenant il reste à choisir le β qui vérifie les conditions de la propriété. On va comme au début choisir $K(x) = \frac{1 - \cos(x)}{\pi(1 + x^2)}$ qui vérifie toutes les propriétés du théorème.

$$\forall 0 < \beta < 1,$$

$$\begin{aligned} \int_{\mathbb{R}} \frac{1 - \cos(x)}{\pi(1 + x^2)} |x|^\beta dx &= 2 \int_0^\infty \frac{1 - \cos(x)}{\pi x^{2-\beta}} dx \\ &= 2 \left(\int_0^1 \frac{1 - \cos(x)}{\pi x^{2-\beta}} dx + \int_1^\infty \frac{1 - \cos(x)}{\pi x^{2-\beta}} dx \right) \end{aligned}$$

Or,

$$\frac{1 - \cos(x)}{\pi x^{2-\beta}} \leq \frac{x^2/2}{\pi x^{2-\beta}} = \frac{x^\beta}{2\pi}$$

Donc

$$\int_0^1 \frac{1 - \cos(x)}{\pi x^{2-\beta}} dx \leq \int_0^1 \frac{x^\beta}{2\pi} dx < +\infty$$

Et de plus,

$$\int_1^\infty \frac{1 - \cos(x)}{\pi x^{2-\beta}} dx \leq \int_1^\infty \frac{1}{\pi x^{2-\beta}} dx < +\infty,$$

car $2 - \beta > 1$ On a alors que,

$$\int_{\mathbb{R}} \frac{1 - \cos(x)}{\pi(1 + x^2)} |x|^\beta dx < +\infty, \forall 0 < \beta < 1$$

$$\forall 1 \leq \beta \leq 2,$$

$$\int_{\mathbb{R}} \frac{1 - \cos(x)}{\pi(1 + x^2)} |x|^\beta dx = 2 \left(\underbrace{\int_0^1 \frac{1 - \cos(x)}{\pi x^{2-\beta}} dx}_{:=A} + \int_1^\infty \frac{1 - \cos(x)}{\pi x^{2-\beta}} dx \right)$$

On a que,

$$A \leq \int_0^1 \frac{x^2/2}{\pi x^{2-\beta}} dx = \int_0^1 \frac{x^\beta}{2\pi} dx < +\infty$$

Maintenant, $\forall x \geq 1, x^{2-\beta} < x$, car $0 \leq 2 - \beta \leq 1$

Donc:

$$\int_1^\infty \frac{1 - \cos(x)}{\pi x^{2-\beta}} dx \geq \int_1^\infty \frac{1 - \cos(x)}{\pi x} dx = \underbrace{\int_1^\infty \frac{1}{\pi x} dx}_{=+\infty} - \underbrace{\int_1^\infty \frac{\cos(x)}{x} dx}_{<+\infty}$$

Donc,

$$\forall 1 < \beta < 2, \int_{\mathbb{R}} \frac{1 - \cos(x)}{\pi(1 + x^2)} |x|^\beta dx = +\infty$$

On peut montrer de la même manière que cette intégrale est divergente pour $\beta \geq 2$.
On en déduit que pour avoir la convergence faible définie ci-dessus, il faut prendre $0 < \beta < 1$
Pour vérifier cela numériquement, nous allons calculer la vitesse de l'algorithme qui calcule $\hat{m}(x)$,
pour des $0 < \beta < 1$ et des $\beta \geq 1$, et on va également mesurer la précision du modèle pour des b
optimaux et non optimaux. On devrait s'attendre à des d'excellente précision pour les b optimaux.
On crée une liste de $0 < \beta < 1$ de longueur 50, comme suit:

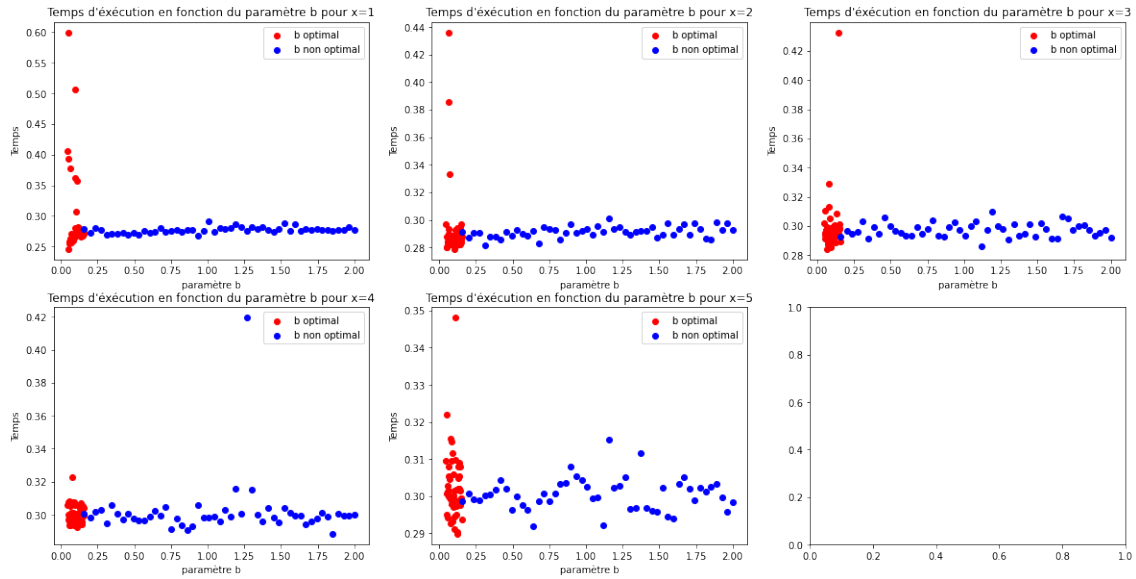
```
beta_lst_1=np.linspace(0.01,0.99,50)
```

Puis la liste b_lst_1 contient toutes les b optimaux calculés à partir des $0 < \beta < 1$. La liste b_bis
quant à elle, contient tous les b non optimaux, et ainsi la liste b_lst contient 100 paramètre b , avec
50 b optimaux, et 50 non-optimaux.

```
b_lst_1=10050**(-1/(2*beta_lst_1+2+1))
b_bis=np.linspace(0.2,2,50)
b_lst=np.concatenate((b_lst_1,b_bis), axis=0)
import time
```

Dans le code ci-dessous, on calcule $\hat{m}(x)$, pour $x = 1, \dots, 5$, pour tous les paramètres b de ma
liste b_lst , et pour chacun de ses calculs, on calcule à l'aide de la fonction `time.time()`, le temps
d'exécution.

```
time_lst=[]
for x in tqdm([1,2,3,4,5]):
    time_L=[]
    for b in b_lst:
        start=time.time()
        num=[]
        denom=[]
        for w,y in list(zip(W_train,y_train)):
            D=np.sin((w-x)/b)/(w-x)
            A=np.sin((w-x)/b)/((w-x)*b)+ (1/((w-x)**2))*(np.cos((w-x)/b)-1)
            C=-np.cos((w-x)/b)/(b**2 *(w-x))+2*np.sin((w-x)/b)/((w-x)**2 *b)
            + (2/((w-x)**3))*(np.cos((w-x)/b)-1)
            B=np.sin((w-x)/b)/((w-x)**2) - np.cos((w-x)/b)/(b*(w-x))
            num.append(y*(2*D-2*b*A-2*b*C+2*B))
            denom.append(2*D-2*b*A-2*b*C+2*B)
        sum(num)/sum(denom)
        time_L.append(time.time()-start)
    time_lst.append(time_L)
```

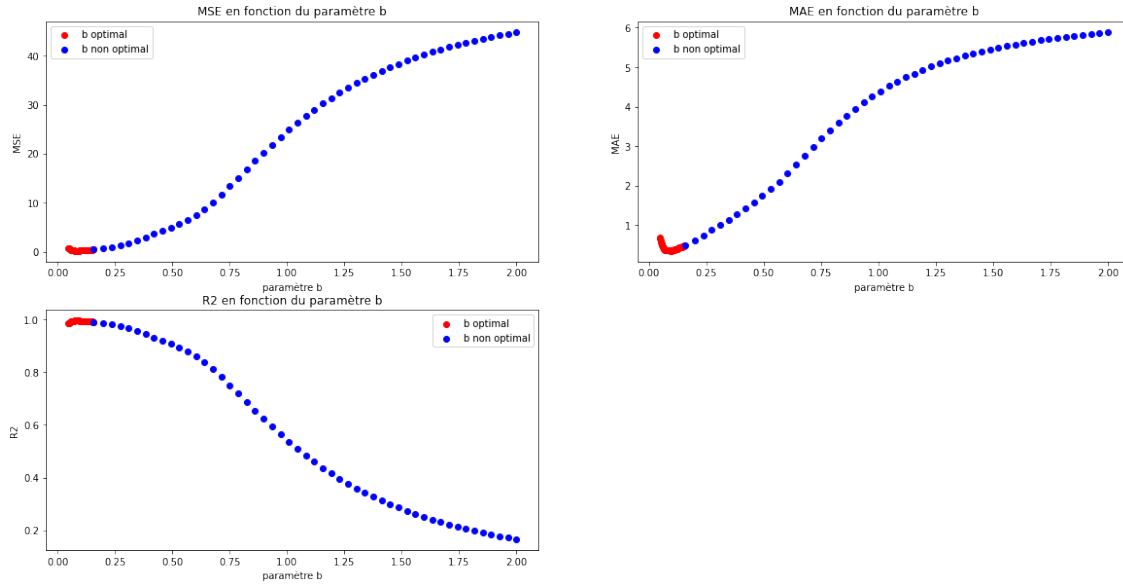


Les graphiques ci-dessus, quantifie le temps d'exécution du calcul de $\hat{m}(x)$, en fonction des b choisis. On ne remarque pas une réelle incidence des paramètres optimaux sur les temps d'exécution. Cependant, nous allons voir ci-dessous, que les paramètres optimaux, nous donne une bien meilleure précision que les autres paramètres.

On a calculé pour chaque b dans b_lst , le MAE, le MSE, et le R2, sur les données d'entraînement, mais seulement sur 100 données, car si le calcul avait été fait sur les 10050 données, on aurait eu un temps d'exécution de 10H et quelques. Et de plus le calcul de ces indicateurs, sur les 100 données sont largement suffisant pour avoir une idée sur l'efficacité des paramètres.

```
MSE_lst=[]
MAE_lst=[]
R2_lst=[]
for b in tqdm(b_lst):
    y_pred_train=[]
    for i in range(100):
        num=[]
        denom=[]
        for w,y in list(zip(W_train,y_train)):
            x=X_train[i]
            D=np.sin((w-x)/b)/(w-x)
            A=np.sin((w-x)/b)/((w-x)*b)+ (1/((w-x)**2))*(np.cos((w-x)/b)-1)
            C=-np.cos((w-x)/b)/(b**2 *(w-x))+2*np.sin((w-x)/b)/((w-x)**2 *b)
            + (2/((w-x)**3))*(np.cos((w-x)/b)-1)
            B=np.sin((w-x)/b)/((w-x)**2) - np.cos((w-x)/b)/(b*(w-x))
            num.append(y*(2*D-2*b*A-2*b*C+2*B))
            denom.append(2*D-2*b*A-2*b*C+2*B)
        y_pred_train.append(sum(num)/sum(denom))
    MSE_lst.append(mean_squared_error(y_train[0:100],y_pred_train))
```

```
MAE_lst.append(mean_absolute_error(y_train[0:100],y_pred_train))
R2_lst.append(r2_score(y_train[0:100],y_pred_train))
```



On remarque que les 50 b optimaux, nous donne une MAE, et un MSE très proche de 0, et un R^2 très proche de 1, et d'autre part, les b non-optimaux donnent des MSE, et des MAE assez grand, avec des MSE qui peuvent aller jusqu'à 40, donc ce n'est pas très intéressant de choisir ces b non-optimaux. Et de même, le R^2 pour les b non-optimaux, sont très bas.

5.6 Application à une base de données

On va désormais s'intéresser à un problème concret, on va s'intéresser à une vraie base de donnée d'un projet immobilier de Casablanca au Maroc. Pourquoi avoir choisi cette base de données en particulier? Pour rappel, notre modèle n'est applicable que si nos valeurs d'entrées et de sorties sont dans \mathbb{R} . Et tous les biens immobiliers de ce projet sont situés dans une même zone, donc la localisation des biens de ce projet n'a aucune incidence sur le prix du bien. Le seul facteur qui va jouer sur le prix est donc la superficie, on va alors obtenir une base de données avec une seule variable explicatives.

Voici un aperçu de la base de données :

	N° de Lot	Surface (en m²)	Nature	Prix de ventes(dhs)
0	1761	183	RDC Commercial + 4 étages	1006500
1	1762	202	RDC Commercial + 4 étages	1111000
2	1763	201	RDC Commercial + 4 étages	1105500
3	1764	199	RDC Commercial + 4 étages	1094500
4	1765	198	RDC Commercial + 4 étages	1089000

	Lot n°	Surface (m²)	Nature	Prix de ventes(dhs)
0	1	115	RDC Commercial + 3 étages	517500
1	2	99	RDC Commercial + 3 étages	445500
2	3	99	RDC Commercial + 3 étages	445500
3	4	99	RDC Commercial + 3 étages	445500
4	5	99	RDC Commercial + 3 étages	445500

Nous avons utilisé les méthodes à noyaux exactement comme dans les prototypes précédents et nous avons également utilisé le b optimal dont on connaît l'estimation. On obtient les métriques suivantes pour les données d'entraînement:

```
print("MAE:", mean_absolute_error(y_train, y_pred_train_B[20]))
print("MSE:", mean_squared_error(y_train,y_pred_train_B[20]))
print("R2:", r2_score(y_train,y_pred_train_B[20]))
```

Output :

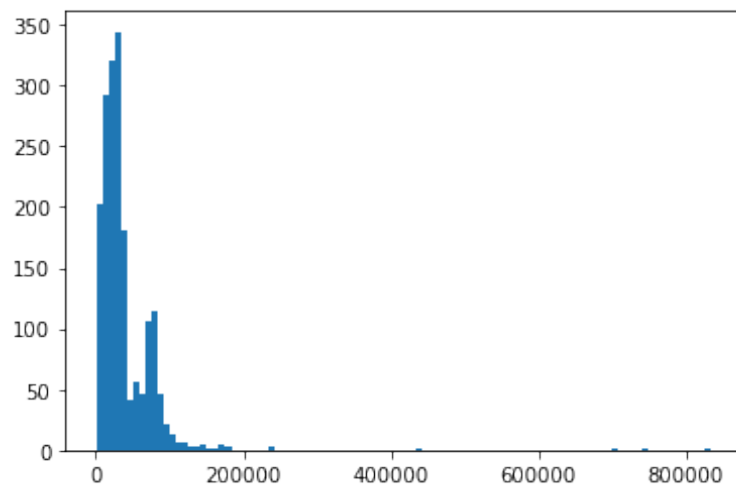
MAE: 38024.25164802749

MSE: 3918814639.717953

R2: 0.9191890207065867

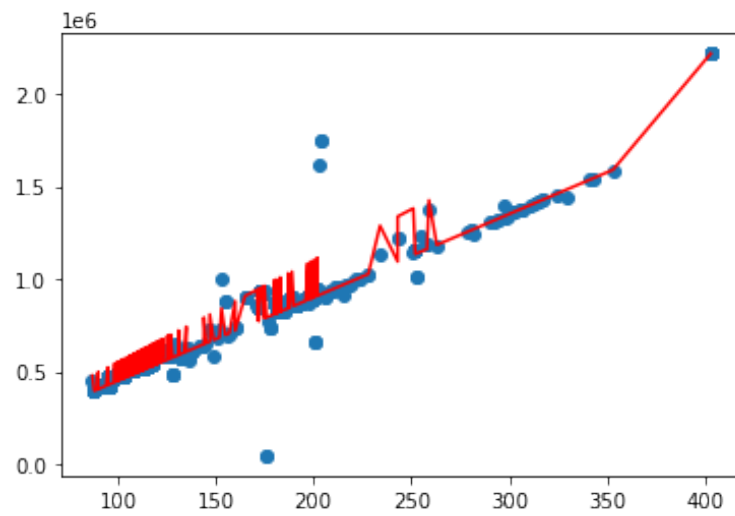
On remarque un R2 très proche de 1. Dans ce contexte, le MAE et MSE ne sont pas pertinents car certains prix de biens sont de l'ordre de 1 million.

Puis, comme précédemment, voici l'histogramme qui quantifie l'erreur pour chaque donnée du trainset :



Le modèle semble performant car la majorité des données du trainset sont proches de 0 en terme d'erreur.

Dans le graphe ci dessous, la fonction en rouge représente les prix des biens immobiliers en fonction de la superficie des biens du trainset et les points bleus représentent les prix des biens immobiliers prédits par le modèle en fonction de la superficie et on remarque visuellement une bonne prédiction du modèle.



On remarque aussi la présence de 5 outlier et ceci explique bien pourquoi le MSE et le MAE sont très grands. On fait de même pour le test set (mais on ne le détaillera pas ici).

6 Annexe

6.1 Preuve du Lemme 1

Lemme 1

$$E |Y_j - m(X_j)|^2 \leq E |Y_j - f(X_j)|^2$$

Pour toute fonction mesurable $f: \mathbb{R} \rightarrow \mathbb{R}$ avec $E(f^2(X_j)) < \infty$.

Preuve. Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ mesurable et telle que $E(f^2(X_j)) < \infty$

On a :

$$\begin{aligned} E(|Y_j - f(X_j)|^2) &= E(|Y_j - m(X_j) + m(X_j) - f(X_j)|^2) \\ &= E(|Y_j - m(X_j)|^2) + \underbrace{E(|m(X_j) - f(X_j)|^2)}_{\geq 0} \\ &\quad + 2E((Y_j - m(X_j))(m(X_j) - f(X_j))) \\ &\geq E(|Y_j - m(X_j)|^2) + 2E((Y_j - m(X_j))(m(X_j) - f(X_j))) \end{aligned}$$

Or:

$$\begin{aligned} E((Y_j - m(X_j))(m(X_j) - f(X_j))) &= E(E((Y_j - m(X_j))(m(X_j) - f(X_j)) | X_j)) \\ &= E((m(X_j) - f(X_j))E(Y_j - m(X_j) | X_j)) \\ &= E((m(X_j) - f(X_j))\underbrace{(E(Y_j | X_j) - m(X_j))}_{=0}) \\ &= 0 \end{aligned}$$

D'où,

$$E(|Y_j - m(X_j)|^2) \leq E(|Y_j - f(X_j)|^2)$$

□

6.2 Preuve du Lemme 2

Lemme 2

Si f_X est bornée et continue, $f_X^{ft} \in \mathbb{L}^1(\mathbb{R})$, $g^{ft}(t) \neq 0 \forall t \in \mathbb{R}$, $K \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ et K à support compact. On a $\forall x \in \mathbb{R}$:

- $\mathbb{E}(\hat{f}_X(x)) = (K_b * f_X)(x)$, où $K_b = \frac{1}{b} K(\frac{\cdot}{b})$
- $var(\hat{f}_X(x)) \leq \frac{1}{2\pi} \|f_X * g\|_\infty \frac{1}{n} \int \left| \frac{K^{ft}(tb)}{g^{ft}(t)} \right|^2 dt$

Preuve.

$$\begin{aligned} E(\widehat{f_X}(x)) &= E\left(\frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \frac{1}{n} \sum_{j=1}^n \frac{e^{itW_j}}{g^{ft}(t)} dt\right) \\ &= \frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \frac{1}{n} \sum_{j=1}^n \frac{E(e^{itW_j})}{g^{ft}(t)} dt \end{aligned}$$

↑ **Par Fubini**

$$\begin{aligned} &= \frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \frac{E(e^{itW_1})}{g^{ft}(t)} dt \\ &= \frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \frac{h^{ft}(t)}{g^{ft}(t)} dt \\ &= \frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) f^{ft}(t) dt \end{aligned}$$

Or $K^{ft}(tb) = K_b^{ft}(t)$, où on note $K_b(t) = \frac{K(\frac{t}{b})}{b}$
Ainsi,

$$\begin{aligned} E(\widehat{f_X}(x)) &= \frac{1}{2\pi} \int e^{-itx} K_b^{ft}(t) f^{ft}(t) dt \\ &= \frac{1}{2\pi} \int e^{-itx} (K_b * f)^{ft}(t) dt \\ &= (K_b * f)(x) \end{aligned}$$

Où l'on a appliqué l'inversion de Fourier dans la dernière égalité puisque $K \in \mathbb{L}^1(\mathbb{R})$ implique que $\|K_b^{ft}\|_\infty = \|K^{ft}\|_\infty \leq \|K\|_1$ et donc que $(K_b * f)^{ft} = K_b^{ft} f^{ft} \in \mathbb{L}^1(\mathbb{R})$

Pour la variance, comme les W_j sont iid, on a:

$$\begin{aligned} Var(\widehat{f_X}(x)) &= Var\left(\frac{1}{2\pi} \int e^{-itx} K^{ft}(tb) \frac{1}{n} \sum_{j=1}^n \frac{e^{itW_j}}{g^{ft}(t)} dt\right) \\ &= \frac{1}{(2\pi)^2 n} Var\left(\int e^{-it(x-W_1)} \frac{K^{ft}(tb)}{g^{ft}(t)} dt\right) \\ &\leq \frac{1}{(2\pi)^2 n} E(|\int e^{-it(x-W_1)} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2) \\ &\uparrow \text{car } Var(X) \leq E(|X|^2) \quad \forall X \in \mathbb{L}^2 \\ &= \frac{1}{(2\pi)^2 n} \int |\int e^{-it(x-y)} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2 h(y) dy \\ &= \frac{1}{(2\pi)^2 n} \int |\int e^{-itz} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2 (f_X * g)(x-z) dz \\ &\leq \|f * g\|_\infty \frac{1}{(2\pi)^2 n} \int |\int e^{-itz} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2 dz \\ &= \frac{1}{2\pi} \|f * g\|_\infty \frac{1}{n} \int |\frac{K^{ft}(tb)}{g^{ft}(t)}|^2 dt \end{aligned}$$

Où nous avons utilisé l'identité de Parseval dans la dernière égalité. Et notons que $\|f * g\|_\infty < \infty$ car pour tout x , on a

$$|(f * g)(x)| = \left| \int f(y)g(x-y)dy \right| \leq \|f\|_\infty \underbrace{\left| \int g(x-y)dy \right|}_{=1} = \|f\|_\infty < \infty$$

□

6.3 Preuve du Théorème 2

Théorème 2 On considère le modèle (4). On suppose que les conditions

$$|m(x)| \leq C_1 \text{ et } f_X(x) \geq C_2 \text{ pour des constantes } C_1, C_2 > 0$$

sont inclus dans la définition de \mathcal{P} .

On suppose également que $f_X(\cdot)Var(Y_1 | X_1 = \cdot)$ et $m(\cdot)^2 f_X(\cdot)$ sont uniformément bornés sur \mathbb{R} . On suppose aussi que p et f_X satisfont la condition de Hölder pour $\beta > 0$, que $p \in \mathbb{L}^1(\mathbb{R})$. En utilisant l'estimateur (11) avec K borné et continue et qui satisfait $K \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, $supp(K^{ft}) = [-1, 1]$, K est un noyau d'ordre $\lfloor \beta \rfloor$ tel que $\int |K(z)| |z|^\beta dz < \infty$

On a

- Pour g ordinary smooth, on prend $b \asymp n^{\frac{-1}{2\beta+2\alpha+1}}$ et on a alors :

$$\lim_{C \rightarrow \infty} \left(\limsup_{n \rightarrow \infty} \sup_{(m, f_X) \in \mathcal{P}} P \left[|\hat{m}(x) - m(x)|^2 \geq C \cdot n^{-\frac{2\beta}{2\beta+2\alpha+1}} \right] \right) = 0$$

- Pour g supersmooth, on prend $b \asymp c_b^{-\frac{1}{\gamma}} \log(n)^{-\frac{1}{\gamma}}$, avec $c_b \in]0, \frac{1}{1d_1}[$ et on a alors :

$$\lim_{C \rightarrow \infty} \left(\limsup_{n \rightarrow \infty} \sup_{(m, f_X) \in \mathcal{P}} P \left[|\hat{m}(x) - m(x)|^2 \geq C \cdot \log(n)^{-\frac{2\beta}{\gamma}} \right] \right) = 0$$

Preuve. Commençons par montrer l'inégalité suivante :

$$\begin{aligned} \mathbb{P}[|R(x)| > C\epsilon_n] &\leq \mathbb{P} \left[|\hat{p}(x) - p(x)| > \sqrt{C\epsilon_n} f_X(x)/3 \right] \\ &+ \mathbb{P} \left[\left| \hat{f}_X(x) - f_X(x) \right| > f_X(x) \min\{1/2, \sqrt{C\epsilon_n}/(6|m(x)|)\} \right] \end{aligned} \tag{12}$$

En effet, si $R(x) = |\hat{m}(x) - m(x)|^2 > C\epsilon_n$, alors

$$\begin{aligned}
\sqrt{C\epsilon_n} &< \left| \frac{\hat{p}(x)}{\hat{f}_X(x)} - \frac{p(x)}{f_X(x)} \right| \\
&\leq \left| \frac{\hat{p}(x)}{\hat{f}_X(x)} - \frac{p(x)}{\hat{f}_X(x)} \right| + \left| \frac{p(x)}{\hat{f}_X(x)} - \frac{p(x)}{f_X(x)} \right| \\
&\leq \left| \frac{\hat{p}(x) - p(x)}{\hat{f}_X(x)} - \frac{\hat{p}(x) - p(x)}{f_X(x)} + \frac{\hat{p}(x) - p(x)}{f_X(x)} \right| + |p(x)| \left| \frac{1}{\hat{f}_X(x)} - \frac{1}{f_X(x)} \right| \\
&\leq |\hat{p}(x) - p(x)| \left| \frac{1}{\hat{f}_X(x)} - \frac{1}{f_X(x)} \right| + \frac{|\hat{p}(x) - p(x)|}{|f_X(x)|} + |p(x)| \left| \frac{1}{\hat{f}_X(x)} - \frac{1}{f_X(x)} \right| \\
&= |\Delta p(x)| \left| \frac{f_X(x)}{\hat{f}_X(x)} - 1 \right| + |\Delta p(x)| + |m(x)| \left| \frac{f_X(x)}{\hat{f}_X(x)} - 1 \right| \\
&= \frac{|\Delta p(x) \Delta f_X(x)|}{|1 + \Delta f_X(x)|} + |\Delta p(x)| + \frac{|m(x) \Delta f_X(x)|}{|1 + \Delta f_X(x)|}
\end{aligned}$$

où $\Delta p(x) = \frac{\hat{p}(x) - p(x)}{f_X(x)}$ et $\Delta f_X(x) = \frac{\hat{f}_X(x) - f_X(x)}{f_X(x)}$

Ainsi, $|R(x) > C\epsilon_n|$ implique une des trois assertions suivantes:

1. $|\delta p(x)| > \frac{\sqrt{C\epsilon_n}}{3}$
2. $\frac{|\Delta p(x) \Delta f_X(x)|}{|1 + \Delta f_X(x)|} > \frac{\sqrt{C\epsilon_n}}{3}$
3. $\frac{|m(x) \Delta f_X(x)|}{|1 + \Delta f_X(x)|} > \frac{\sqrt{C\epsilon_n}}{3}$

Si on a $\frac{|m(x) \Delta f_X(x)|}{|1 + \Delta f_X(x)|} > \frac{\sqrt{C\epsilon_n}}{3}$ alors on a $|\Delta f_X(x)| > \frac{1}{2}$ ou $|\Delta f_X(x)| > \frac{\sqrt{C\epsilon_n}}{6|m(x)|}$.

En effet, cela se montre facilement par contraposée : si on suppose $|\Delta f_X(x)| \leq \frac{1}{2}$ et

$$|\Delta f_X(x)| \leq \frac{\sqrt{C\epsilon_n}}{6|m(x)|} \text{ alors } \frac{|m(x) \Delta f_X(x)|}{|1 + \Delta f_X(x)|} \leq \frac{\sqrt{C\epsilon_n}}{6|1 + \Delta f_X(x)|} \leq \frac{\sqrt{C\epsilon_n}}{6 \cdot \frac{1}{2}} = \frac{\sqrt{C\epsilon_n}}{3}$$

De la même manière, on montre que $\frac{|\Delta p(x) \Delta f_X(x)|}{|1 + \Delta f_X(x)|} > \frac{\sqrt{C\epsilon_n}}{3}$ implique que $|\Delta f_X(x)| > \frac{1}{2}$ ou $\frac{|\Delta p(x)|}{2} \geq \frac{\sqrt{C\epsilon_n}}{6}$

Ainsi,

$$|R(x) > C\epsilon_n| \implies |\Delta p(x)| > \frac{\sqrt{C\epsilon_n}}{3} \text{ ou } |\Delta f_X(x)| > \min\left(\frac{1}{2}, \frac{\sqrt{C\epsilon_n}}{6|m(x)|}\right)$$

D'où

$$\begin{aligned}
\mathbb{P}[|R(x)| > C\epsilon_n] &\leq \mathbb{P}\left[|\hat{p}(x) - p(x)| > \sqrt{C\epsilon_n} f_X(x)/3\right] \\
&+ \mathbb{P}\left[\left|\hat{f}_X(x) - f_X(x)\right| > f_X(x) \min\{1/2, \sqrt{C\epsilon_n}/(6|m(x)|)\}\right]
\end{aligned}$$

En appliquant l'inégalité de Markov, on obtient :

$$\begin{aligned}
&\sup_{(m, f_X) \in \mathcal{P}} P\left[|\hat{m}(x) - m(x)|^2 \geq C \cdot \epsilon_n\right] \\
&\leq \text{const} \cdot (C\epsilon_n)^{-1} \sup_{(m, f_X) \in \mathcal{P}} \left(E(|\hat{p}(x) - p(x)|^2) + E\left(\left|\hat{f}_X(x) - f_X(x)\right|^2\right) \right)
\end{aligned} \tag{13}$$

Étudions le terme $E(|\hat{p}(x) - p(x)|^2)$. Par la décomposition Biais-Variance, on a :

$$E(|\hat{p}(x) - p(x)|^2) = \text{var}(\hat{p}(x)) + (E(\hat{p}(x) - p(x)))^2$$

Étude de la Variance :

Avant de commencer, on rappelle l'inégalité de Parseval :

$$\|f\|_2^2 = \frac{1}{2\pi} \|f^{ft}\|_2^2 \quad \forall f \in \mathbb{L}^2(\mathbb{R})$$

$$\begin{aligned} \text{Var}(\hat{p}(x)) &= \frac{1}{n(2\pi)^2} \text{Var} \left(Y_1 \int e^{-it(x-W_1)} \frac{K^{ft}(tb)}{g^{ft}(t)} dt \right) \quad \text{car les } (W_i, Y_i) \text{ sont iid} \\ &\leq \frac{1}{n(2\pi)^2} E(|Y_1|^2 \mid \int e^{-itx} e^{itW_1} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2) \quad \text{car } \text{Var}(X) \leq E(X^2) \quad \forall X \in \mathbb{L}^2 \\ &= \frac{1}{n(2\pi)^2} E(|Y_1|^2 \mid \int e^{-itx} e^{it(X_1+\delta_1)} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2) \quad \text{car } W_1 = X_1 + \delta_1 \\ &= \frac{1}{n(2\pi)^2} E(E(|Y_1|^2 \mid \int e^{-itx} e^{itX_1} e^{it\delta_1} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2 \mid X_1 = \xi)) \\ &= \frac{1}{n(2\pi)^2} \int E(|Y_1|^2 \mid \int e^{-itx} e^{itX_1} e^{it\delta_1} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2 \mid X_1 = \xi) f_X(\xi) d\xi \\ &= \frac{1}{n(2\pi)^2} \int E(|Y_1|^2 \mid X_1 = \xi) E(\int e^{-itx} e^{it\xi} e^{it\delta_1} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2) f_X(\xi) d\xi \quad \text{car } \delta_1 \perp Y_1 \text{ et } \delta_1 \perp X_1 \\ &\leq \frac{1}{n(2\pi)^2} \|E(|Y_1|^2 \mid X_1 = \cdot) f_X(\cdot)\|_\infty \int E(\int e^{-itx} e^{it\xi} e^{it\delta_1} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2) d\xi \\ &\uparrow \text{Car } E(|Y_1|^2 \mid X_1 = \cdot) f_X(\cdot) \text{ est uniformément borné sur } \mathbb{R}. \\ &\text{Comme } E(|Y_1|^2 \mid X_1 = x) f_X(x) = f_X(x) \text{Var}(Y_1 \mid X_1 = x) + m(x)^2 f_X(x) \\ &\leq \frac{1}{2\pi n} \|E(|Y_1|^2 \mid X_1 = \cdot) f_X(\cdot)\|_\infty E(\frac{1}{2\pi} \int \int |e^{-itx} e^{it\xi} e^{it\delta_1} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2 d\xi) \\ &\uparrow \text{Par Fubini} \\ &\leq \frac{1}{2\pi n} \|E(|Y_1|^2 \mid X_1 = \cdot) f_X(\cdot)\|_\infty E(\frac{1}{2\pi} \int (\int |e^{it\xi} e^{it\delta_1} \frac{K^{ft}(tb)}{g^{ft}(t)} dt|^2) d\xi) \\ &\uparrow \text{car } |e^{-itx}| = 1 \\ &= \frac{1}{2\pi n} \|E(|Y_1|^2 \mid X_1 = \cdot) f_X(\cdot)\|_\infty E(\frac{1}{2\pi} \|l^{ft}\|_2^2) \\ &\uparrow \text{où } l(t) = e^{it\delta_1} \frac{K^{ft}(tb)}{g^{ft}(t)} \text{ et ainsi par Parseval :} \end{aligned}$$

$$\begin{aligned}
Var(\hat{p}(x)) &\leq \frac{1}{2\pi n} \underbrace{\| E(|Y_1|^2 | X_1 = \cdot) f_X(\cdot) \|_\infty}_{\leq Cste} \int \left| \frac{K^{ft}(tb)}{g^{ft}(t)} \right|^2 dt \\
&\leq Cste. \max_{|t| \leq \frac{1}{b}} |g^{ft}(t)|^{-2} \frac{2}{nb} \\
&\uparrow \text{ car } supp(K^{ft}) = [-1, 1] \implies supp(K^{ft}(b \cdot)) = \left[-\frac{1}{b}, \frac{1}{b}\right] \text{ et comme } g^{ft} \text{ est continue,} \\
&\text{non nulle sur un compact, par le th  or  me de Weierstrass, } |g^{ft}(t)|^{-2} \text{ est born  e} \\
&\text{et atteint ses bornes et car } K \in \mathbb{L}^2. \\
&\leq Cste \frac{1}{nb} \max_{|t| \leq \frac{1}{b}} |g^{ft}(t)|^{-2}
\end{aligned}$$

Ainsi,

$$Var(\hat{p}(x)) \leq Cste \frac{1}{nb} \max_{|t| \leq \frac{1}{b}} |g^{ft}(t)|^{-2} \quad (14)$$

  tude du Biais² : Nous avons montr   pr  c  demment que $\mathbb{E}[\hat{p}(x)] = (p * K_b)(x)$, o   $K_b = \frac{K(\frac{\cdot}{b})}{b}$.
Ainsi,

$$\begin{aligned}
E(\hat{p}(x) - p(x))^2 &= |p * K_b(x) - p(x)|^2 \\
&= \left| \int p(x-y) K_b(y) dy - p(x) \right|^2 \\
&= \left| \int p(x-y) \frac{1}{b} K\left(\frac{y}{b}\right) dy - p(x) \right|^2 \\
&\downarrow \text{ On effectue le changement de variable } z = \frac{y}{b} \\
&= \left| \int p(x-bz) K(z) dz - p(x) \right|^2 \\
&= \left| \int p(x-bz) K(z) dz - p(x) \int K(z) dz \right|^2 \\
&\uparrow \text{ car } \int K = 1 \\
&= \left| \int K(z) (p(x-bz) - p(x)) dz \right|^2
\end{aligned}$$

Alors :

$$\begin{aligned}
\sup_{(m, f_X) \in \mathcal{P}} |E(\hat{p}(x) - p(x))|^2 &= \sup_{(m, f_X) \in \mathcal{P}} \left| \int K(z)(p(x - bz) - p(x))dz \right|^2 \\
&\leq \underbrace{\sup_{(m, f_X) \in \mathcal{P}} \left| \int_{|z| \leq \frac{\delta}{b}} K(z)(p(x - bz) - p(x))dz \right|^2}_{A_1} \\
&\quad + \underbrace{\sup_{(m, f_X) \in \mathcal{P}} \left| \int_{|z| > \frac{\delta}{b}} K(z)(p(x - bz) - p(x))dz \right|^2}_{A_2}
\end{aligned}$$

En utilisant la formule de Taylor à l'ordre $\lfloor \beta \rfloor - 1$,

$$\begin{aligned}
A_1 &= \sup_{(m, f_X) \in \mathcal{P}} \left| \int_{|z| \leq \frac{\delta}{b}} K(z) \left(\sum_{j=1}^{\lfloor \beta \rfloor - 1} \frac{1}{j!} p^{(j)}(x)(-zb)^j + \frac{1}{\lfloor \beta \rfloor!} p^{(\lfloor \beta \rfloor)}(\xi_{b,z})(-zb)^{\lfloor \beta \rfloor} \right) dz \right|^2 \\
&\text{où } \xi_{b,z} \text{ est entre } x \text{ et } x - zb \\
&= \sup_{(m, f_X) \in \mathcal{P}} \left| \int_{|z| \leq \frac{\delta}{b}} K(z) \left(\sum_{j=1}^{\lfloor \beta \rfloor} \frac{1}{j!} p^{(j)}(x)(-zb)^j + \frac{1}{\lfloor \beta \rfloor!} (p^{(\lfloor \beta \rfloor)}(\xi_{b,z}) - p^{(\lfloor \beta \rfloor)}(x))(-zb)^{\lfloor \beta \rfloor} \right) dz \right|^2 \\
&\downarrow (a+b)^2 \leq 2a^2 + 2b^2 \\
&\leq 2 \underbrace{\sup_{(m, f_X) \in \mathcal{P}} \left| \sum_{j=1}^{\lfloor \beta \rfloor} (-1)^j \int_{|z| \leq \frac{\delta}{b}} K(z) z^j dz \frac{b^j}{j!} p^{(j)}(x) \right|^2}_{A_{1,1}} \\
&\quad + 2 \underbrace{\sup_{(m, f_X) \in \mathcal{P}} \left| \frac{1}{\lfloor \beta \rfloor!} (-b)^{\lfloor \beta \rfloor} \int_{|z| \leq \frac{\delta}{b}} K(z) |z|^{\lfloor \beta \rfloor} (p^{(\lfloor \beta \rfloor)}(\xi_{b,z}) - p^{(\lfloor \beta \rfloor)}(x)) dz \right|^2}_{A_{1,2}} \\
A_{1,2} &\leq 2 \sup_{(m, f_X) \in \mathcal{P}} \frac{b^{2\lfloor \beta \rfloor}}{\lfloor \beta \rfloor!^2} \left(\int_{|z| \leq \frac{\delta}{b}} |K(z)| |z|^{\lfloor \beta \rfloor} C |\xi_{b,z} - x|^{\beta - \lfloor \beta \rfloor} dz \right)^2 \\
&\leq \frac{2C^2}{\lfloor \beta \rfloor!^2} b^{2\lfloor \beta \rfloor} \left(\int |K(z)| |z|^{\lfloor \beta \rfloor} dz \right)^2 b^{2\beta - 2\lfloor \beta \rfloor} \\
&= O(b^{2\beta})
\end{aligned}$$

Pour le terme $A_{1,1}$, on a besoin du lemme suivant que l'on admettra :

Lemme 3. Soit une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ qui satisfait la condition de Hölder et est majorée par C_0 sur sa restriction sur l'intervalle $[x - \delta, x + \delta]$. Alors, on a

$$|f^{(j)}(y)| \leq C_j \quad \forall y \in [x - \delta, x + \delta], j = 0, \dots, \lfloor \beta \rfloor$$

où les C_j ne dépendent que de C_0, C, β, δ .

De plus, comme K est d'ordre $\lfloor \beta \rfloor$, on a :

$$\int_{\mathbb{R}} z^j K(z) dz = 0 \quad \forall j \in \{1, \dots, \lfloor \beta \rfloor\}$$

Ceci implique alors que

$$\int_{|z| \leq \frac{\delta}{b}} z^j K(z) dz = - \int_{|z| > \frac{\delta}{b}} z^j K(z) dz \quad \forall j \in \{1, \dots, \lfloor \beta \rfloor\}$$

Ainsi,

$$\begin{aligned} A_{1,1} &= 2 \sup_{(m, f_X) \in \mathcal{P}} \left| \sum_{j=1}^{\lfloor \beta \rfloor} (-1)^{j+1} \int_{|z| > \frac{\delta}{b}} K(z) z^j dz p^{(j)}(x) \frac{b^j}{j!} \right|^2 \\ A_{1,1} &\leq O(1) \sup_{(m, f_X) \in \mathcal{P}} \sum_{j=1}^{\lfloor \beta \rfloor} b^{2j} \left| \int_{|z| > \frac{\delta}{b}} K(z) z^j dz p^{(j)}(x) \right|^2 \\ &\leq O(1) \sum_{j=1}^{\lfloor \beta \rfloor} b^{2j} \left(\int_{|z| > \frac{\delta}{b}} |K(z)| |z|^\beta \underbrace{|z|^{j-\beta}}_{\leq (\frac{\delta}{b})^{j-\beta}} dz \right)^2 \end{aligned}$$

↑ **Lemme 2**

$$\begin{aligned} &\leq O(1) \sum_{j=1}^{\lfloor \beta \rfloor} b^{2j} \delta^{2j-2\beta} \left(\int |K(z)| |z|^\beta dz \right)^2 \\ &= O(b^{2\beta}) \end{aligned}$$

Ainsi, $A_1 = O(b^{2\beta})$

Pour A_2 , comme $\|p\|_\infty \leq C$,

$$\begin{aligned} A_2 &= \sup_{(m, f_X) \in \mathcal{P}} \left| \int_{|z| > \frac{\delta}{b}} K(z) (p(x - bz) - p(x)) dz \right|^2 \\ &\leq (2C)^2 \left(\int_{|z| > \frac{\delta}{b}} |K(z)| dz \right)^2 \\ &\leq O(b^{2\beta}) \left(\int_{|z| > \frac{\delta}{b}} |K(z)| |z|^\beta dz \right)^2 \\ &= O(b^{2\beta}) \end{aligned}$$

Enfin, $\sup_{(m, f_X) \in \mathcal{P}} |E(\hat{p}(x) - p(x))|^2 = O(b^{2\beta})$.

Retour Biais² + Variance :

$$\sup_{(m, f_X) \in \mathcal{P}} E(|\hat{p}(x) - p(x)|^2) \leq O\left(\frac{1}{bn}\right) \max_{|t| \leq \frac{1}{b}} |g^{ft}(t)|^{-2} + O(b^{2\beta}) \quad (15)$$

On peut obtenir un résultat analogue pour le terme en f_X en supposant les mêmes conditions de Hölder sur f_X .

- Ainsi, si g est "ordinary smooth", on a que $\max_{|t| \leq \frac{1}{b}} |g^{ft}(t)|^{-2} = O(b^{-2\alpha})$, c'est-à-dire que $\text{var}(\hat{p}(x)) = O(n^{-1}b^{-1-2\alpha})$. On obtient donc

$$\sup_{(m, f_X) \in \mathcal{P}} \left(E(|\hat{p}(x) - p(x)|^2) + E(|\hat{f}_X(x) - f_X(x)|^2) \right) = O(n^{-1}b^{-1-2\alpha} + b^{2\beta})$$

les taux de convergence optimaux sont atteints lorsque la variance et le terme de biais convergent vers zéro avec le même taux. Ceci est possible lorsque l'on choisit $b \asymp n^{\frac{1}{2\beta+2\alpha+1}}$. Dans ce cas,

$$\sup_{(m, f_X) \in \mathcal{P}} \left(E(|\hat{p}(x) - p(x)|^2) + E(|\hat{f}_X(x) - f_X(x)|^2) \right) = O(n^{-\frac{2\beta}{2\beta+2\alpha+1}})$$

Et si on prend $\epsilon_n = n^{-\frac{2\beta}{2\beta+2\alpha+1}}$, on obtient bien :

$$\begin{aligned} & \sup_{(m, f_X) \in \mathcal{P}} P[|\hat{m}(x) - m(x)|^2 \geq C \cdot \epsilon_n] \\ & \leq \text{const} \cdot (C\epsilon_n)^{-1} \sup_{(m, f_X) \in \mathcal{P}} \left(E(|\hat{p}(x) - p(x)|^2) + E(|\hat{f}_X(x) - f_X(x)|^2) \right) \\ & \leq \text{const} \cdot (C)^{-1} \xrightarrow{C \rightarrow \infty} 0 \end{aligned}$$

- Si g est "supersmooth", on a que $\max_{|t| \leq \frac{1}{b}} |g^{ft}(t)|^{-2} = O(\exp(2d_1 b^{-\gamma}))$, c'est-à-dire que $\text{var}(\hat{p}(x)) = O(n^{-1}b^{-1} \exp(2d_1 b^{-\gamma}))$. On obtient donc

$$\sup_{(m, f_X) \in \mathcal{P}} \left(E(|\hat{p}(x) - p(x)|^2) + E(|\hat{f}_X(x) - f_X(x)|^2) \right) = O(n^{-1}b^{-1} \exp(2d_1 b^{-\gamma}) + b^{2\beta})$$

On se propose de choisir $b \asymp c_b^{-\frac{1}{\gamma}} \log(n)^{-\frac{1}{\gamma}}$, avec $c_b \in]0, \frac{1}{1d_1}[$.

Dans ce cas,

$$\sup_{(m, f_X) \in \mathcal{P}} \left(E(|\hat{p}(x) - p(x)|^2) + E(|\hat{f}_X(x) - f_X(x)|^2) \right) = O(\log(n)^{-\frac{2\beta}{\gamma}})$$

Et si on prend $\epsilon_n = \log(n)^{-\frac{2\beta}{\gamma}}$, on obtient bien :

$$\begin{aligned} & \sup_{(m, f_X) \in \mathcal{P}} P[|\hat{m}(x) - m(x)|^2 \geq C \cdot \epsilon_n] \\ & \leq \text{const} \cdot (C\epsilon_n)^{-1} \sup_{(m, f_X) \in \mathcal{P}} \left(E(|\hat{p}(x) - p(x)|^2) + E(|\hat{f}_X(x) - f_X(x)|^2) \right) \\ & \leq \text{const} \cdot (C)^{-1} \xrightarrow{C \rightarrow \infty} 0 \end{aligned}$$

□

6.4 Preuve du Théorème 3

Théorème 3 : Dans le cadre de notre modèle de régression, si nos échantillons $(X_1, Y_1), \dots, (X_n, Y_n)$ suivent la loi du couple (X, Y) de densité:

$$f_{(X,Y),\theta}(x, y) = f_1(x)f_2(y) + \theta C \Delta_X(x) \Delta_Y(y),$$

avec $\theta \in \{0, 1\}$. Et,

$$f_0(x) = \frac{1 - \cos(x)}{\pi x^2}, \quad f_1(x) = \frac{1}{\pi(1 + x^2)}, \quad f_2(y) = \frac{1}{2} \exp(-|y|),$$

$$\Delta_X(x) = a_n^{-\beta} \cos(2a_n x) f_0(a_n x), \quad \Delta_Y(y) = \frac{1}{4} \text{sign}(y) \exp(-|y|)$$

avec $(a_n)_n \rightarrow +\infty$, $\beta \geq 1$, et que

$$\|\tau^2\| = \|\text{var}(Y_j^2 | X_j = \cdot)\|_\infty \leq \infty$$

alors il n'existe aucun estimateur de $m(x)$, basé sur les données $(W_1, Y_1), \dots, (W_n, Y_n)$, tel que:

$$\lim_{C \rightarrow +\infty} (\limsup_{n \rightarrow +\infty} \sup_{(m, f_X) \in P} P(|\hat{m}(x) - m(x)|^2 \geq C \epsilon_n) = 0$$

SI: Cas 1: g smooth, $|(g^{ft})'(t)| \leq \text{const.}|t|^{-\alpha}$, et $\epsilon_n = O(n^{\frac{-2\beta}{2\beta+2\alpha+1}})$, $a_n = n^{\frac{1}{(2\beta+2\alpha+1)}}$

OU

Cas 2: g supersmooth, $|(g^{ft})'(t)| \leq \exp(-d|t|^\gamma)$, $d > 0$, et $\epsilon_n = O((\log(n))^{-2\beta/\gamma})$, $a_n = c(\log(n))^{\frac{1}{\gamma}}$, avec c suffisamment grand.

Preuve. Dans cette partie, on va supposer que nos échantillons $(X_1, Y_1), \dots, (X_n, Y_n)$ suivent la loi du couple (X, Y) de densité:

$$f_{(X,Y),\theta}(x, y) = f_1(x)f_2(y) + \theta C \Delta_X(x) \Delta_Y(y),$$

avec $\theta \in \{0, 1\}$. On a donc effectivement deux densités proposées, une avec $\theta = 0$ et la deuxième avec $\theta = 1$. On pose,

$$f_0(x) = \frac{1 - \cos(x)}{\pi x^2}, \quad f_1(x) = \frac{1}{\pi(1 + x^2)}, \quad f_2(y) = \frac{1}{2} \exp(-|y|),$$

$$\Delta_X(x) = a_n^{-\beta} \cos(2a_n x) f_0(a_n x), \quad \Delta_Y(y) = \frac{1}{4} \text{sign}(y) \exp(-|y|)$$

avec $(a_n)_n \uparrow \infty$, $\beta \geq 1$ et,

$$\text{sign}(y) = \begin{cases} -1 & \text{si } y < 0 \\ 0 & \text{si } y = 0 \\ 1 & \text{si } y > 0 \end{cases}$$

La constante C et la suite $(a_n)_n$ seront choisies plus tard.
Maintenant nous allons vérifier que les $f_{(X,Y),\theta}$ sont bien des densités.
Pour $\theta = 0$:

$$f_{(X,Y),0}(x, y) = f_1(x)f_2(x)$$

et

$$\int f_1(x)f_2(y)dxdy = (\int f_1(x)dx)(\int f_2(y)dy) = 1$$

on arrive à cette égalité en utilisant Fubini Tonelli, et on obtient 1 car f_1 et f_2 sont des densités.
Pour $\theta = 1$:

$$f_{(X,Y),1}(x, y) = f_1(x)f_2(y) + C\Delta_X(x)\Delta_Y(y)$$

Or,

$$|\Delta_Y(y)| = \frac{1}{4}|sign(y)|\exp(-|y|)$$

donc

$$2|\Delta_Y(y)| = \frac{1}{2}|sign(y)|\exp(-|y|) \leq f_2(y)$$

Et

$$|\Delta_X(x)| = |a_n^{-\beta}||\cos(2a_n x)| \frac{(1 - \cos(a_n x))}{\pi a_n^2 x^2} \leq const. \frac{1 - \cos(a_n x)}{\pi a_n^2 x^2}$$

car comme $(a_n)_n$ tend vers $+\infty$, on peut prendre $|a_n^{-\beta}|$ aussi petit qu'on veut.
Donc,

$$|\Delta_X(x)| \leq const. \frac{1}{1 + a_n^2 x^2} = c.f_1(x)$$

On obtient alors,

$$f_{(X,Y),1}(x, y) \geq \frac{|\Delta_X(x)|}{c}f_2(y) + C\Delta_X(x)\Delta_Y(y) \geq 2\frac{|\Delta_X(x)|}{c}|\Delta_Y(y)| + C\Delta_X(x)\Delta_Y(y)$$

et donc, pour C assez petit, on a $f_{(X,Y),1}(x, y) \geq 0$.
De plus,

$$\int f_{(X,Y),1}(x, y)dxdy = \int f_1(x)f_2(y)dxdy + C(\int \Delta_X(x)dx)(\int \Delta_Y(y)dy)$$

en utilisant Fubini. Or,

$$\int \Delta_Y(y)dy = \int_{-\infty}^0 \frac{-1}{4}e^y dy + \int_0^{\infty} \frac{1}{4}e^{-y} dy = \int_{+\infty}^0 \frac{-1}{4}e^{-u}(-du) + \int_0^{\infty} \frac{1}{4}e^{-y} dy = 0,$$

effectuant le changement de variable $y = -u$.

Alors,

$$\int f_{(X,Y),1}(x, y)dxdy = \int f_1(x)dx. \int f_2(y)dy = 1$$

Donc les $f_{(X,Y),\theta}$ sont bien des densités.

Nous allons maintenant montrer un résultat qui sera utilisé plus tard, il s'agit de montrer que $(m, f_X) \in P$.

On rappelle que $(m, f_X) \in P$ si et seulement si:

- $\|m^2 f_X\|_\infty, \|f_X\|_\infty \leq C_1, m.f_X, (m.f_X)^{ft}, f_X^{ft} \in L_1(\mathbb{R})$
- $m.f_X, f_X$, vérifient la condition de Hölder pour $\beta \geq 1$.
- $f_X(x) \geq C_2 > 0$
- $\|\tau^2\|_\infty = \|var(Y_j^2 | X_j = .)\|_\infty \leq C_3$

Maintenant calculons $f_X(x)$:

$$f_X(x) = f_{X,\theta}(x) = \int f_{(X,Y),\theta}(x,y)dy = f_1(x) \underbrace{\int f_2(y)dy}_{=1} + \theta C \Delta_X(x) \underbrace{\int \Delta_Y(y)dy}_{=0} = f_1(x)$$

$$\begin{aligned} m(x) &= m_\theta(x) = E[Y_1 | X_1 = x] = \int y f_{Y_1 | X_1 = x}(y) dy = \int y \frac{f_{(X,Y)}(x,y)}{f_X(x)} dy \\ &= (f_X(x))^{-1} \left(\int y f_1(x) f_2(y) dy + \theta C \int y \Delta_X(x) \Delta_Y(y) dy \right) \\ &= (f_1(x))^{-1} \left(f_1(x) \int y f_2(y) dy + \theta C \Delta_X(x) \int y \Delta_Y(y) dy \right) \end{aligned}$$

Or,

$$\begin{aligned} \int y f_2(y) dy &= \int \frac{1}{2} y \exp(-|y|) dy = \int_{-\infty}^0 \frac{1}{2} y e^y dy + \int_0^{+\infty} \frac{1}{2} y e^{-y} dy \\ &= \int_{+\infty}^0 \frac{1}{2} (-u) e^{-u} (-du) + \int_0^{+\infty} \frac{1}{2} y e^{-y} dy \\ &= - \int_0^{+\infty} \frac{1}{2} u e^{-u} du + \int_0^{+\infty} \frac{1}{2} y e^{-y} dy = 0, \end{aligned}$$

en, effectuant le changement de variable $u = -y$. Donc,

$$m(x) = (f_1(x))^{-1} \theta C \Delta_X(x) \int y \Delta_Y(y) dy$$

Et,

$$\int y \Delta_Y(y) dy = 2 \int_0^{+\infty} \frac{1}{4} \text{sign}(y) y \exp(-|y|) dy = \frac{1}{2} \int_0^{+\infty} y e^{-y} dy = \frac{1}{2}$$

Donc,

$$m(x) = \theta C' \frac{\Delta_X(x)}{f_1(x)},$$

avec $C' > 0$.

Calculons maintenant $\text{Var}(Y_j^2 | X_j = x)$:

On a:

$$\begin{aligned} E[Y_j^2 | X_j = x] &= \int y^2 f_{(X,Y),\theta}(x, y) \frac{dy}{f_X(x)} \\ &= (f_1(x))^{-1} \left(\int y^2 f_1(x) f_2(y) dy + \theta C \int y^2 \Delta_X(x) \Delta_Y(y) dy \right) \end{aligned}$$

Et on peut vérifier par une intégration par parties que:

$$\int y^2 \Delta_Y(y) dy = 0$$

Donc

$$E(Y_1^2 | X_j = x) = \int y^2 f_2(y) dy$$

qui est une intégrale finie.

Et de plus

$$E(Y_j^4 | X_j = x) = \int y^4 \frac{f_{(X,Y)}(x, y)}{f_X(x)} dy$$

et

$$\int y^4 f_{(X,Y)}(x, y) dy = \underbrace{\int y^4 f_1(x) f_2(y) dy}_{=A} + \theta C \Delta_X(x) \underbrace{\int y^4 \Delta_Y(y) dy}_{=B}$$

$$A = 2f_1(x) \underbrace{\int_0^{+\infty} \frac{y^4}{2} e^{-y} dy}_{<+\infty} = \text{cste} f_1(x)$$

On peut vérifier par une intégration par partie que:

$$B = \int y^4 \text{sign}(y) e^{-|y|} dy = 0$$

Donc, $E(Y_1^4|X_1 = x) = cste(f_1(x))^{-1}f_1(x) = cste$, alors on vient de montrer que

$$\|Var(Y_j^2|X_j = .)\|_\infty < +\infty$$

On vérifie maintenant que $m^2 f_X$ est bornée.

$$m^2(x)f_X(x) = \theta^2 C''^2 \frac{\Delta_X(x)^2}{f_1^2(x)} f_1(x) = \theta^2 C''^2 \frac{\Delta_X(x)^2}{f_1(x)}$$

Or, on avait vu précédemment que $\Delta_X^2(x) \leq c^2 f_1^2(x)$.

Alors,

$$m^2(x)f_X(x) \leq \theta^2 C''^2 \frac{c^2 f_1^2(x)}{f_1(x)} = \theta^2 C''^2 C^2 f_1(x) \leq cste$$

Ensuite, on va montrer que $p^{ft}(\cdot) = (m(\cdot)f_X(\cdot))^{ft}$, et $f_X(\cdot)$, sont dans $\mathbb{L}^1(\mathbb{R})$

$$p(x) = p_\theta(x) = m(x)f_X(x) = \theta C' \frac{\Delta_X(x)}{f_1(x)} f_1(x) = \theta C' \Delta_X(x)$$

Donc,

$$p^{ft}(t) = \theta C' \Delta_X^{ft}(t)$$

.

$$\begin{aligned} \Delta_X^{ft}(t) &= a_n^{-\beta} (\cos(2a_n \cdot) f_0(a_n \cdot))^{ft}(t) = a_n^{-\beta} \int e^{itx} (\cos(2a_n x) f_0(a_n x)) dx \\ &= a_n^{-\beta} \int e^{it \frac{u}{a_n}} (\cos(2u) f_0(u)) \frac{1}{a_n} = a_n^{-\beta-1} (\cos(2 \cdot) f_0(\cdot))^{ft}\left(\frac{t}{a_n}\right) \end{aligned}$$

Avant de continuer le calcul, on rappelle que:

$$[f(\cdot) \cos(a \cdot)]^{ft}(t) = \frac{1}{2} f^{ft}(t+a) + \frac{1}{2} f^{ft}(t-a)$$

Alors,

$$\Delta_X^{ft}(t) = \frac{a_n^{-\beta-1}}{2} \left(f_0^{ft}\left(\frac{t+2a_n}{a_n}\right) + f_0^{ft}\left(\frac{t-2a_n}{a_n}\right) \right)$$

Alors, on obtient:

$$p^{ft}(t) = \theta C' \left(\frac{1}{2} a_n^{-\beta-1} f_0^{ft}\left(\frac{t+2a_n}{a_n}\right) + \frac{1}{2} a_n^{-\beta-1} f_0^{ft}\left(\frac{t-2a_n}{a_n}\right) \right)$$

Et de plus, on peut facilement vérifier que:

$$\frac{1}{2\pi} \int_{-1}^1 e^{-itx} (1 - |t|) dt = \frac{1 - \cos(x)}{\pi x^2} = f_0(x)$$

Par la formule d'inversion de Fourier, on en déduit que:

$$f_0^{ft} = (1 - |t|)_+$$

Donc f_0^{ft} est à support compact, et donc p^{ft} l'est aussi, alors p^{ft} est bien dans $\mathbb{L}^1(\mathbb{R})$.

Puis on vérifie de la même manière que $f_X^{ft}(t) = f_1^{ft}(t) = (\frac{1}{\pi(1+t^2)})^{ft}(t) = \exp(-|t|)$, donc $f_X^{ft} \in \mathbb{L}^1(\mathbb{R})$.

Maintenant, pour finalement montrer que $(m, f_X) \in P$, il reste à montrer que p vérifie la condition de Hölder.

On a que $p(x) = \theta C' \Delta_X(x) = \theta C' a_n^{-\beta} \cos(2a_n x) f_0(a_n x)$

Posons $H(x) = \cos(2x) f_0(x)$, donc $p(x) = \theta C' a_n^{-\beta} H(a_n x) \implies p'(x) = \theta C' a_n^{-\beta} a_n H'(a_n x)$

On remarque que $p^{\lfloor \beta \rfloor}(x) = \theta C' a_n^{-\beta} a_n^{\lfloor \beta \rfloor} H^{\lfloor \beta \rfloor}(a_n x) = \theta C' a_n^{-\beta} a_n^{\lfloor \beta \rfloor} V(a_n x)$

où on a posé que $V(x) = \frac{d^{\lfloor \beta \rfloor}}{dx^{\lfloor \beta \rfloor}}(\cos(2x) f_0(x))$.

On peut ensuite vérifier par récurrence que toutes les dérivées de f_0 sont bornées sur \mathbb{R} , ce qui implique que $V(x)$ et $V'(x)$ sont bornées.

On sait que

$$V(x) = \frac{d^{\lfloor \beta \rfloor}}{dx^{\lfloor \beta \rfloor}}(\cos(2x) f_0(x)) = \sum_{k=0}^{\lfloor \beta \rfloor} \binom{\lfloor \beta \rfloor}{k} \underbrace{g^{\lfloor \beta \rfloor - k}(x)}_{\leq \text{const}} \underbrace{f_0^{(k)}(x)}_{\leq \text{const}} \leq \text{const},$$

où on a posé que $g(x) = \cos(2x)$

Et,

$$V'(x) = \sum_{k=0}^{\lfloor \beta \rfloor} \binom{\lfloor \beta \rfloor}{k} \underbrace{g^{\lfloor \beta \rfloor - k + 1}(x)}_{\leq \text{const}} \underbrace{f_0^{(k)}(x)}_{\leq \text{const}} + \sum_{k=0}^{\lfloor \beta \rfloor} \binom{\lfloor \beta \rfloor}{k} \underbrace{g^{\lfloor \beta \rfloor - k}(x)}_{\leq \text{const}} \underbrace{f_0^{(k+1)}(x)}_{\leq \text{const}}$$

Et enfin, on peut vérifier les inégalités qui suivent:

$$\begin{aligned} |p^{\lfloor \beta \rfloor}(y) - p^{\lfloor \beta \rfloor}(\tilde{y})| &= |\theta C' a_n^{-\beta} a_n^{\lfloor \beta \rfloor} V(a_n y) - \theta C' a_n^{-\beta} a_n^{\lfloor \beta \rfloor} V(a_n \tilde{y})| \\ &= \theta C' a_n^{-\beta} a_n^{\lfloor \beta \rfloor} |V(a_n y) - V(a_n \tilde{y})| \\ &\leq O(a_n^{-\beta + \lfloor \beta \rfloor}) |a_n y - a_n \tilde{y}|^{\beta - \lfloor \beta \rfloor} \end{aligned}$$

$$\leq \text{const.} |y - \tilde{y}|^{\beta - \lfloor \beta \rfloor}$$

Alors on vient de montrer que $(m, f_X) \in P$.

Maintenant, nous allons introduire deux ensembles qui serviront plus tard. On peut écrire, en utilisant une notation alternative, que,

$$\hat{m}(x) = \hat{m}(x; (W_1, Y_1), \dots, (W_n, Y_n))$$

Et on introduit deux ensembles:

$$G(m, \epsilon) = \{Z \in (\mathbb{R}^2)^n, |\hat{m}(0; Z) - m(0)|^2 \geq \epsilon\}$$

$$D_{m, \tilde{m}} = \{Z \in (\mathbb{R}^2)^n, |\hat{m}(0; Z) - m(0)|^2 \leq |\hat{m}(0; Z) - \tilde{m}(0)|^2\},$$

où m et \tilde{m} sont deux fonctions de régression. On remarque que $G(m, \epsilon)$ est l'ensemble des données d'entraînement où l'écart entre la prédiction et la vraie valeur est plus grand que ϵ , et $D_{m, \tilde{m}}$, représente l'ensemble des données d'entraînement où l'estimateur \hat{m} est plus efficace pour estimer la fonction de régression m .

On va montrer le résultat suivant: $D_{m, \tilde{m}} \subset G(\tilde{m}, \epsilon)$, si $\epsilon < \frac{|m(0) - \tilde{m}(0)|^2}{4}$

Si $z \in D_{m, \tilde{m}}$, alors $|\hat{m}(0; z) - m(0)|^2 \leq |\hat{m}(0; z) - \tilde{m}(0)|^2$.

Supposons par l'absurde que $|\hat{m}(0; z) - \tilde{m}(0)|^2 \leq \epsilon$, donc $|\hat{m}(0; z) - \tilde{m}(0)| < \sqrt{\epsilon} < \frac{|m(0) - \tilde{m}(0)|}{2}$.

Or, $|\hat{m}(0; z) - m(0)| \leq |\hat{m}(0; z) - \tilde{m}(0)| < \sqrt{\epsilon}$.

Et $|m(0) - \tilde{m}(0)| \leq |\hat{m}(0; z) - \tilde{m}(0)| + |\hat{m}(0; z) - m(0)| < 2\sqrt{\epsilon} < |m(0) - \tilde{m}(0)|$.

Contradiction!

Donc $|\hat{m}(0; z) - \tilde{m}(0)|^2 \geq \epsilon$, d'où $z \in G(\tilde{m}, \epsilon)$.

On va maintenant montrer que $(\mathbb{R}^2)^n / D_{m, \tilde{m}} \subset G(m, \epsilon)$ si $\epsilon < |m(0) - \tilde{m}(0)|^2 / 4$.

Supposons par l'absurde que $|\hat{m}(0; z) - m(0)|^2 < \epsilon$, on a de plus que $\epsilon > |\hat{m}(0; z) - m(0)|^2 > |\hat{m}(0; z) - \tilde{m}(0)|^2$, car $z \in (\mathbb{R}^2)^n / D_{m, \tilde{m}}$.

Et puis, $|m(0) - \tilde{m}(0)| \leq |\hat{m}(0; z) - m(0)| + |\hat{m}(0; z) - \tilde{m}(0)| < 2\sqrt{\epsilon} < |m(0) - \tilde{m}(0)|$

Contradiction! Alors $|\hat{m}(0; z) - m(0)|^2 \geq \epsilon$ donc, $z \in G(m, \epsilon)$.

Maintenant que l'on a introduit ces propriétés indispensables, nous rentrons dans le vif du sujet. On remarque que la définition de la convergence faible de $\hat{m}(x)$ au taux $(\epsilon_n)_n$ peut s'écrire comme suit:

$$\lim_{C \rightarrow +\infty} (\limsup_{n \rightarrow +\infty} \sup_{(m, f_X) \in P} P(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m, C\epsilon_n))) = 0$$

Par définition,

$$P(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m, C\epsilon_n)) = P_0(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m_0, C\epsilon_n))$$

$$+P_1(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m_1, C\epsilon_n))$$

où $P_\theta, \theta \in \{0, 1\}$, est une loi de probabilité sous les hypothèses que $f_{(X,Y),\theta}$ représente la densité des (X_j, Y_j) . En effet la fonction de régression m , peut-être soit égale à m_0 si $\theta = 0$, soit à m_1 si $\theta = 1$.

-Maintenant, si on a $4c\epsilon_n < |m_0(0) - m_1(0)|^2$, une condition qu'on appellera (C1), $\forall n > 0$ suffisamment grand,

$$\begin{aligned} & P_0(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m_0, C\epsilon_n)) + P_1(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m_1, C\epsilon_n)) \\ & \geq P_0(((W_1, Y_1), \dots, (W_n, Y_n)) \in (\mathbb{R}^2)^n / D_{m_0, m_1}) + P_1(((W_1, Y_1), \dots, (W_n, Y_n)) \in D_{m_0, m_1}) \\ & = 1 - P_0(((W_1, Y_1), \dots, (W_n, Y_n)) \in D_{m_0, m_1}) + P_1(((W_1, Y_1), \dots, (W_n, Y_n)) \in D_{m_0, m_1}) \\ & \geq 1 - \sup_{D \in B((\mathbb{R}^2)^n)} (P_0(((W_1, Y_1), \dots, (W_n, Y_n)) \in D) - P_1(((W_1, Y_1), \dots, (W_n, Y_n)) \in D)) \end{aligned}$$

On va poser les mesures $P_\theta(D) = P_\theta(((W_1, Y_1), \dots, (W_n, Y_n)) \in D)$ On a donc que

$$\begin{aligned} & P_0(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m_0, C\epsilon_n)) + P_1(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m_1, C\epsilon_n)) \\ & \geq 1 - \sup_{D \in B((\mathbb{R}^2)^n)} |P_0(D) - P_1(D)| = 1 - TV(P_0, P_1), \end{aligned}$$

où

$$TV(P_0, P_1) = \sup_{D \in B((\mathbb{R}^2)^n)} |P_0(D) - P_1(D)|$$

Pour avancer dans nos calculs, on va montrer un résultat très utile, c'est le lemme de Scheffé's:

Lemme de Schéffé

Soient P, Q deux mesures de probabilité sur \mathbb{R}^d , $d \geq 1$, de densité respective p, q , alors on a $\|p - q\|_1 = 2TV(P, Q)$.

Preuve. Soit $B \subset \mathbb{R}^d$,

$$\begin{aligned} \|p - q\|_1 &= \int_{\mathbb{R}^d} |p(x) - q(x)| dx \\ &= \int_B |p(x) - q(x)| dx + \int_{B^c} |p(x) - q(x)| dx \\ &\geq \int_B (p(x) - q(x)) dx + \int_{B^c} (q(x) - p(x)) dx \\ &= \int_B p(x) dx - \int_B q(x) dx + 1 - \int_B q(x) dx - 1 + \int_B p(x) dx \\ &= 2 \left(\int_B p(x) dx - \int_B q(x) dx \right) \end{aligned}$$

En échangeant le rôle de p et q , on a:

$$\|p - q\|_1 \geq 2 \left(\int_B q(x) dx - \int_B p(x) dx \right)$$

Alors,

$$\|p - q\|_1 \geq 2 \left| \int_B p(x) dx - \int_B q(x) dx \right| = 2|P(B) - Q(B)|$$

Ceci étant vraie pour tout boréliens, on a alors $\|p - q\|_1 \geq 2TV(P, Q)$.

Pour l'inégalité inverse, on introduit:

$$A = \{x \in \mathbb{R}^d, p(x) > q(x)\}$$

On a :

$$\begin{aligned} \|p - q\|_1 &= \int_{\mathbb{R}^d} |p(x) - q(x)| dx \\ &= \int_A |p(x) - q(x)| dx + \int_{A^c} |p(x) - q(x)| dx \\ &= \int_A (p(x) - q(x)) dx + \int_{A^c} (q(x) - p(x)) dx \\ &= 2 \left(\int_A p(x) dx - \int_A q(x) dx \right) = 2(P(A) - Q(A)) \\ &\leq 2|P(A) - Q(A)| \leq 2 \sup_{B \in \mathcal{B}((\mathbb{R})^d)} |P(B) - Q(B)| = 2TV(P, Q) \end{aligned}$$

Alors on a bien $\|p - q\|_1 = 2TV(P, Q)$. □

Ce résultat étant maintenant démontré, remarquons que les densités p_θ des mesures P_θ définies avant sont la densité de l'échantillon i.i.d $((W_1, Y_1), \dots, (W_n, Y_n))$.

Donc,

$$p_\theta((w_1, y_1), \dots, (w_n, y_n)) = \prod_{j=1}^n f_{(W,Y);\theta}(w_j, y_j)$$

On a:

$$1 - TV(P_0, P_1) = 1 - \frac{1}{2} \int |p_0(x) - p_1(x)| dx$$

-si $p_0(x) \geq p_1(x)$: $\max\{p_0(x), p_1(x)\} - \min\{p_0(x), p_1(x)\} = p_0(x) - p_1(x) = |p_0(x) - p_1(x)|$

-sinon: $\max\{p_0(x), p_1(x)\} - \min\{p_0(x), p_1(x)\} = p_1(x) - p_0(x) = |p_1(x) - p_0(x)|$

On a alors: $\max\{p_0(x), p_1(x)\} - \min\{p_0(x), p_1(x)\} = |p_0(x) - p_1(x)|$

Donc,

$$\begin{aligned}
1 - TV(P_0, P_1) &= 1 - \frac{1}{2} \int \underbrace{\max \{p_0(x), p_1(x)\}}_{\leq p_0(x) + p_1(x)} dx + \frac{1}{2} \int \min \{p_0(x), p_1(x)\} dx \\
&\geq 1 - \frac{1}{2} \underbrace{\int (p_0(x) + p_1(x)) dx}_{=2} + \frac{1}{2} \int \min \{p_0(x), p_1(x)\} dx \\
&= \frac{1}{2} \int \min \{p_0(x), p_1(x)\} dx
\end{aligned}$$

car p_0 et p_1 sont des densités.

Pour continuer à minorer $1 - TV(P_0, P_1)$, on va introduire l'inégalité de LeCam.

Inégalité de LeCam:

Pour $d \geq 1$, et h, \tilde{h} des densités, on a:

$$\frac{1}{2} \left(\int \sqrt{h(x)} \sqrt{\tilde{h}(x)} dx \right)^2 \leq \int \min \{h(x), \tilde{h}(x)\} dx$$

Preuve. On peut vérifier facilement que,

$$\min \{h(x), \tilde{h}(x)\} \max \{h(x), \tilde{h}(x)\} = h(x) \tilde{h}(x)$$

Donc,

$$\begin{aligned}
\int \sqrt{h(x) \tilde{h}(x)} dx &= \int \sqrt{\min \{h(x), \tilde{h}(x)\}} \sqrt{\max \{h(x), \tilde{h}(x)\}} dx \\
&\stackrel{\substack{\leq \\ \text{Cauchy-Schwartz}}}{\leq} \left(\int \min \{h(x), \tilde{h}(x)\} dx \right)^{1/2} \left(\int \underbrace{\max \{h(x), \tilde{h}(x)\}}_{\leq h(x) + \tilde{h}(x)} dx \right)^{1/2} \\
&\leq \sqrt{2} \left(\int \min \{h(x), \tilde{h}(x)\} dx \right)^{1/2}
\end{aligned}$$

On retrouve bien l'inégalité. □

Retournons à notre problème, on a que:

$$1 - TV(P_0, P_1) \geq \frac{1}{2} \int \min \{p_0(x), p_1(x)\} dx$$

$$\begin{aligned}
& \underbrace{\geq}_{LeCam} \frac{1}{4} \left(\int \sqrt{p_0(x)} \sqrt{p_1(x)} dx \right)^2 \\
&= \frac{1}{4} \left(\int \dots \int \prod_{j=1}^n \sqrt{f_{(W,Y);0}(w_j, y_j)} \sqrt{f_{(W,Y);1}(w_j, y_j)} dw_1 dy_1 \dots dw_n dy_n \right)^2 \\
&= \frac{1}{4} \left(\prod_{j=1}^n \int \int \sqrt{f_{(W,Y);0}(w_j, y_j)} \sqrt{f_{(W,Y);1}(w_j, y_j)} dw_j dy_j \right)^2 \\
&= \frac{1}{4} \left(\int \int \sqrt{f_{(W,Y);0}(w, y)} \sqrt{f_{(W,Y);1}(w, y)} dw dy \right)^{2n}
\end{aligned}$$

Pour continuer à minorer on va introduire la distance de Hellinger.

Distance de Hellinger

$$H(h, \tilde{h}) = \left(\int \left(\sqrt{h(x)} - \sqrt{\tilde{h}(x)} \right)^2 dx \right)^{1/2}$$

où h et \tilde{h} sont des densités.
De plus,

$$\begin{aligned}
H^2(h, \tilde{h}) &= \int \left(\sqrt{h(x)} - \sqrt{\tilde{h}(x)} \right)^2 dx \\
&= \int \left(h(x) - 2\sqrt{h(x)\tilde{h}(x)} + \tilde{h}(x) \right) dx \\
&= \underbrace{\int h(x) dx}_{=1} - 2 \int \sqrt{h(x)} \sqrt{\tilde{h}(x)} dx + \underbrace{\int \tilde{h}(x) dx}_{=1} \\
&= 2 - 2 \int \sqrt{h(x)} \sqrt{\tilde{h}(x)} dx
\end{aligned}$$

Donc,

$$0 \leq 1 - \frac{1}{2} H^2(h, \tilde{h}) = \int \sqrt{h(x)} \sqrt{\tilde{h}(x)} dx \underbrace{\leq}_{CauchySchwartz} 1$$

De plus,

$$\sqrt{h(x)} - \sqrt{\tilde{h}(x)} = \frac{(\sqrt{h(x)} - \sqrt{\tilde{h}(x)})(\sqrt{h(x)} + \sqrt{\tilde{h}(x)})}{(\sqrt{h(x)} + \sqrt{\tilde{h}(x)})}$$

$$= \frac{h(x) - \tilde{h}(x)}{\sqrt{h(x)} + \sqrt{\tilde{h}(x)}}$$

Donc,

$$\begin{aligned} H^2(h, \tilde{h}) &= \int \frac{(h(x) - \tilde{h}(x))^2}{\left(\sqrt{h(x)} + \underbrace{\sqrt{\tilde{h}(x)}}_{\geq 0} \right)^2} dx \\ &\leq \int \frac{(h(x) - \tilde{h}(x))^2}{h(x)} dx \stackrel{def}{=} \chi^2(h, \tilde{h}) \end{aligned}$$

où χ^2 est une distance définie pour certaines densités.

Par définition, on a que:

$$\int \int \sqrt{f_{(W,Y);0}(w,y)} \sqrt{f_{(W,Y);1}(w,y)} dw dy = 1 - \frac{1}{2} H^2(f_{(W,Y);0}, f_{(W,Y);1})$$

Donc,

$$1 - TV(P_0, P_1) \geq \frac{1}{4} \left(1 - \frac{1}{2} H^2(f_{(W,Y);0}, f_{(W,Y);1}) \right)^{2n}$$

De plus, comme mentionné précédemment,

$$H^2(f_{(W,Y);0}, f_{(W,Y);1}) \leq \chi^2(f_{(W,Y);0}, f_{(W,Y);1}) \Leftrightarrow 1 - \frac{1}{2} H^2(f_{(X,Y);0}, f_{(X,Y);1}) \geq 1 - \frac{1}{2} \chi^2(f_{(X,Y);0}, f_{(X,Y);1})$$

Or, on a vu précédemment que

$$0 \leq 1 - \frac{1}{2} H^2(f_{(X,Y);0}, f_{(X,Y);1}) \leq 1$$

Donc

$$1 - \frac{1}{2} \chi^2(f_{(X,Y);0}, f_{(X,Y);1}) \leq 1$$

On va maintenant imposer une condition très importante, qu'on va appeler la condition (C2), on va supposer que $\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) = O(1/n)$ (C2).

Donc pour n suffisamment grand, on a que $1 - \frac{1}{2} \chi^2(f_{(X,Y);0}, f_{(X,Y);1}) > 0$

On peut donc minorer comme suit:

$$\begin{aligned} 1 - TV(P_0, P_1) &\geq \frac{1}{4} \left(1 - \frac{1}{2} H^2(f_{(W,Y);0}, f_{(W,Y);1}) \right)^{2n} \\ &\geq \frac{1}{4} \left(1 - \frac{1}{2} \chi^2(f_{(W,Y);0}, f_{(W,Y);1}) \right)^{2n} \geq \text{const} > 0 \end{aligned}$$

On vient alors de montrer que si la condition (C1), et (C2) sont vérifiées, alors

$$\begin{aligned} & P(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m, C\epsilon_n)) \\ &= P_0(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m_0, C\epsilon_n)) + P_1(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m_1, C\epsilon_n)) \\ &\geq 1 - TV(P_0, P_1) > 0 \end{aligned}$$

Or, on avait montré précédemment que à partir de $f_{(X,Y);\theta}$ qu'on avait construite, la fonction de régression m , et la fonction de densité f_X , vérifie $(m, f_X) \in P$. Donc

$$\sup_{(m, f_X) \in P} P(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m, C\epsilon_n)) > 0$$

Et finalement, on obtient,

$$\lim_{C \rightarrow +\infty} (\limsup_{n \rightarrow +\infty} \sup_{(m, f_X) \in P} P(((W_1, Y_1), \dots, (W_n, Y_n)) \in G(m, C\epsilon_n))) > 0$$

Autrement dit sous (C1) et (C2), \hat{m} ne converge pas faiblement au taux $(\epsilon_n)_n$.

Maintenant qu'on connaît les conditions (C1) et (C2), on va trouver une borne de la distance χ^2 .

Tout d'abord,

$$\begin{aligned} P(W_1 \in A, Y_1 \in B) &= P(X_1 + \delta_1 \in A, Y_1 \in B) \\ &= \int_{a \in A} \int_{b \in B} dP(X_1 + \delta_1 = a, Y_1 = b) \\ &= \int \int_{a \in A, b \in B} dP(X_1 + s = a, Y_1 = b) g(s) ds \\ &= \int_{a \in A, b \in B} \int f_{(X,Y);\theta}(a - s, b) g(s) ds da db \end{aligned}$$

La dernière égalité découle du théorème de Fubini. On a alors par le théorème de transfert que:

$$f_{(W,Y);\theta}(w, y) = \int f_{(X,Y);\theta}(w - s, y) g(s) ds$$

Puis, par définition,

$$\begin{aligned} \chi^2(f_{(W,Y);0}, f_{(W,Y);1}) &= \int \int f_{(W,Y);0}^{-1}(w, y) (f_{(W,Y);1}(w, y) - f_{(W,Y);0}(w, y))^2 dw dy \\ &= \int \int \left(\int f_{(X,Y);0}(w - s, y) g(s) ds \right)^{-1} (f_{(W,Y);1}(w, y) - f_{(W,Y);0}(w, y))^2 dw dy \\ &= \int \int \left(\int f_2(y) f_1(w - s) g(s) ds \right)^{-1} (f_{(W,Y);1}(w, y) - f_{(W,Y);0}(w, y))^2 dw dy \\ &= \int \int (f_1 * g)^{-1}(w) f_2^{-1}(y) (f_{(W,Y);1}(w, y) - f_{(W,Y);0}(w, y))^2 dw dy \end{aligned}$$

De plus,

$$\begin{aligned} f_{(W,Y);1}(w, y) &= \int (f_1(w-s)f_2(y) + C\Delta_X(w-s)\Delta_Y(y))g(s)ds \\ &= f_2(y)(f_1 * g)(w) + C\Delta_Y(y)(\Delta_X * g)(w) \end{aligned}$$

Et,

$$f_{(W,Y);0}(w, y) = \int f_1(w-s)f_2(y)g(s)ds = f_2(y)(f_1 * g)(w)$$

Donc,

$$f_{(W,Y);1}(w, y) - f_{(W,Y);0}(w, y) = C\Delta_Y(y)(\Delta_X * g)(w)$$

Alors,

$$\begin{aligned} \chi^2(f_{(W,Y);0}, f_{(W,Y);1}) &= \int \int (f_1 * g)^{-1}(w)f_2^{-1}(y)C^2\Delta_Y^2(y)(\Delta_X * g)^2(w)dwdy \\ &= C^2 \int f_2^{-1}(y)\Delta_Y^2(y)dy \cdot \int (f_1 * g)^{-1}(w)(\Delta_X * g)^2(w)dw \end{aligned}$$

Or,

$$\begin{aligned} (\Delta_X * g)^2(w) &= \left(\int \Delta_X(w-x)g(x)dx \right)^2 \\ &= \left(\int a_n^{-\beta} \cos(2a_n(w-x))f_0(a_n(w-x))g(x)dx \right)^2 \\ &= a_n^{-2\beta} (\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w) \end{aligned}$$

Alors,

$$\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) = C^2 a_n^{-2\beta} \int f_2^{-1}(y)\Delta_Y^2(y)dy \cdot \int (f_1 * g)^{-1}(w)(\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w)dw$$

Or,

$$\int f_2^{-1}(y)\Delta_Y^2(y)dy = \int 2e^{|y|} \frac{1}{16} \text{sign}^2(y) e^{-2|y|} dy = \frac{1}{8} \int e^{-|y|} dy < +\infty$$

On en d duit que,

$$\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) \leq O(a_n^{-2\beta}) \cdot \int (f_1 * g)^{-1}(w)(\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w)dw$$

Ensuite,

$$(f_1 * g)(x) = \int f_1(x-y)g(y)dy$$

et $f_1(w-y) = \frac{1}{\pi(1+(x-y)^2)}$.

De plus, $(x-y)^2 \leq 2x^2 + 2y^2 \Leftrightarrow \frac{1}{\pi(1+(x-y)^2)} \geq \frac{1}{\pi(1+2x^2+2y^2)}$.

Donc,

$$(f_1 * g)(x) \geq \frac{1}{\pi} \int \frac{1}{1+2x^2+2y^2} g(y)dy \underset{g \geq 0}{\geq} \frac{1}{\pi} \int_{|y| \leq \eta} \frac{1}{1+2x^2+2y^2} g(y)dy, \forall \eta > 0$$

$$\begin{aligned}
&\geq \frac{1}{\pi} \frac{1}{1+2x^2+2\eta^2} \int_{|y|\leq\eta} g(y)dy \\
&\geq C_\eta \cdot (1+x^2)^{-1} \int_{|y|\leq\eta} g(y)dy
\end{aligned}$$

Comme $\int g = 1$, on peut trouver un η suffisamment grand tel que $\int_{|y|\leq\eta} g(y)dy \geq 1/2$. Donc $(f_1 * g)(x) \geq \frac{C_\eta}{2} \cdot (1+x^2)^{-1} \geq \text{const.} \cdot (1+x^2)^{-1}$. On en déduit que:

$$\begin{aligned}
\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) &\leq O(a_n^{-2\beta}) \int (1+w^2)(\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w) dw \\
&= O(a_n^{-2\beta}) \left(\int (\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w) dw + \int (w(\cos(2a_n \cdot) f_0(a_n \cdot) * g)(w))^2(w) \right)
\end{aligned}$$

Or d'après la formule d'inversion de Fourier on a,

$$(\cos(2a_n \cdot) f_0(a_n \cdot) * g)(w) = \frac{1}{2\pi} ((\cos(2a_n \cdot) f_0(a_n \cdot) * g)^{ft})^{ft}(-w)$$

Alors,

$$\begin{aligned}
&\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) \\
&\leq O(a_n^{-2\beta}) \left(\int (\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w) dw + \int (w(\cos(2a_n \cdot) f_0(a_n \cdot) * g)(w))^2(w) \right) \\
&= O(a_n^{-2\beta}) \left(\int (\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w) dw + \int (u((\cos(2a_n \cdot) f_0(a_n \cdot) * g)^{ft})^{ft}(u))^2 du \right),
\end{aligned}$$

en effectuant le changement de variable $u = -w$.

Puis, on sait que $\forall t$

$$i(f')^{ft}(t) = t f^{ft}(t)$$

Donc,

$$\begin{aligned}
&\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) \\
&\leq O(a_n^{-2\beta}) \left(\int (\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w) dw - \int (((\cos(2a_n \cdot) f_0(a_n \cdot) * g)^{ft})')^{ft}(u))^2 du \right) \\
&\leq O(a_n^{-2\beta}) \left(\int (\cos(2a_n \cdot) f_0(a_n \cdot) * g)^2(w) dw + \int (((\cos(2a_n \cdot) f_0(a_n \cdot) * g)^{ft})')^{ft}(u))^2 du \right) \\
&\stackrel{\text{Parseval}}{=} O(a_n^{-2\beta}) \left(\int ((\cos(2a_n \cdot) f_0(a_n \cdot) * g)^{ft})^2(w) dw + \int ((\cos(2a_n \cdot) f_0(a_n \cdot) * g)^{ft})'(u))^2 du \right)
\end{aligned}$$

On rappelle l'égalité de Parseval pour la dernière égalité:

$$\|f\|_2^2 = \frac{1}{2\pi} \|f^{ft}\|_2^2$$

Or,

$$\begin{aligned}
& ((\cos(2a_n \cdot) f_0(a_n \cdot) * g)^{ft})'(u) = ((\cos(2a_n \cdot) f_0(a_n \cdot))^{ft}(u) g^{ft}(u))' \\
& = \frac{1}{2a_n} \left((f_0^{ft}(\frac{u}{a_n} + 2) + f_0^{ft}(\frac{u}{a_n} - 2)) g^{ft}(u) \right)' \\
& = \frac{1}{2a_n} \left((\frac{1}{a_n} (f_0^{ft})'(\frac{u}{a_n} + 2) + \frac{1}{a_n} (f_0^{ft})'(\frac{u}{a_n} - 2)) g^{ft}(u) + (f_0^{ft}(\frac{u}{a_n} + 2) + f_0^{ft}(\frac{u}{a_n} - 2)) (g^{ft})'(u) \right)
\end{aligned}$$

Donc,

$$\begin{aligned}
& \chi^2(f_{(W,Y);0}, f_{(W,Y);1}) \\
& \leq O(a_n^{-2\beta}) \left(\int \frac{1}{4a_n^2} (f_0^{ft}(\frac{u}{a_n} + 2) + f_0^{ft}(\frac{u}{a_n} - 2))^2 (g^{ft})^2(u) du \right. \\
& + \int \frac{1}{4a_n^2} \left((\frac{1}{a_n} (f_0^{ft})'(\frac{u}{a_n} + 2) + \frac{1}{a_n} (f_0^{ft})'(\frac{u}{a_n} - 2)) g^{ft}(u) + (f_0^{ft}(\frac{u}{a_n} + 2) + f_0^{ft}(\frac{u}{a_n} - 2)) (g^{ft})'(u) \right)^2 du \\
& \leq O(a_n^{-2\beta-2}) \int (f_0^{ft}(\frac{u}{a_n} + 2) + f_0^{ft}(\frac{u}{a_n} - 2))^2 (g^{ft})^2(u) \\
& + a_n^{-2} ((f_0^{ft})'(\frac{u}{a_n} + 2) + (f_0^{ft})'(\frac{u}{a_n} - 2)) g^{ft}(u)^2 + ((f_0^{ft}(\frac{u}{a_n} + 2) + f_0^{ft}(\frac{u}{a_n} - 2)) (g^{ft})'(u))^2 du
\end{aligned}$$

où on a utilisé que, $(x + y)^2 \leq 2x^2 + 2y^2$, pour la dernière inégalité.

Or, $\text{supp}(f_0^{ft}) = [-1, 1]$ donc, $\text{supp}(f_0^{ft}(\frac{\cdot}{a_n} + 2)) = [-3a_n, -a_n]$, et $\text{supp}(f_0^{ft}(\frac{\cdot}{a_n} - 2)) = [a_n, 3a_n]$. De même,

$$(f_0^{ft})'(t) = \begin{cases} -1 & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } -1 \leq t \leq 0 \\ 0 & \text{sinon} \end{cases}$$

On en déduit que,

$$\begin{aligned}
& \chi^2(f_{(W,Y);0}, f_{(W,Y);1}) \leq O(a_n^{-2\beta-2}) \int_{|t| \in [a_n, 3a_n]} \underbrace{(f_0^{ft}(\frac{u}{a_n} + 2) + f_0^{ft}(\frac{u}{a_n} - 2))^2}_{\leq 1} (g^{ft})^2(u) \\
& + a_n^{-2} \underbrace{((f_0^{ft})'(\frac{u}{a_n} + 2) + (f_0^{ft})'(\frac{u}{a_n} - 2))}_{\leq 1} \underbrace{g^{ft}(u)}_{\leq 1}^2 + \underbrace{((f_0^{ft}(\frac{u}{a_n} + 2) + f_0^{ft}(\frac{u}{a_n} - 2)) (g^{ft})'(u))}_{\leq 1}^2 du
\end{aligned}$$

Comme $a_n \rightarrow +\infty$, on peut prendre a_n^{-2} aussi petit qu'on veut dans le second membre de l'addition ci-dessus pour n suffisamment grand. Alors,

$$\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) = O(a_n^{-2\beta-1}). \max \left\{ |g^{ft}(t)|^2 + |(g^{ft})'(t)|^2 : |t| \in [a_n, 3a_n] \right\}$$

On va maintenant proposer deux cas pour $(g^{ft})'(t)$:

Cas 1: $|(g^{ft})'(t)| \leq \text{const.} |t|^{-\alpha}$, $\alpha > 0$, et g smooth.

Cas 2: $|(g^{ft})'(t)| \leq \exp(-d|t|^\gamma)$, $d > 0$, $\gamma > 0$ et g supersmooth.

Alors,

$$\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) = \begin{cases} O(a_n^{-2\beta-2\alpha-1}) & \text{pour le cas 1} \\ O(a_n^{-2\beta-1}) \exp(-da_n^\gamma) & \text{pour le cas 2} \end{cases}$$

Maintenant nous allons choisir convenablement notre suite a_n , la seule condition sur cette suite c'est qu'elle tend vers $+\infty$.

On va poser $a_n \approx n^{1/(2\beta+2\alpha+1)}$ pour le cas 1, et $a_n = c.(\log(n))^{1/\gamma}$, avec $c > 0$ pour le cas 2.

Pour le cas 1, $\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) \approx 1/n$, et pour le cas 2 on a,

$$a_n^{-2\beta-1} \exp(-da_n^\gamma) = cste.(\log(n))^{\frac{-2\beta-1}{\beta}} = cste \frac{1}{\log(n)^{\frac{2\beta+1}{\gamma}} n^{dc\gamma}},$$

avec c arbitraire.

Ainsi on peut trouver un c , tel que:

$$a_n^{-2\beta-1} \exp(-da_n^\gamma) \leq cste \frac{1}{\log(n)^{\frac{2\beta+1}{\gamma}} n} \leq cste \frac{1}{n}$$

Alors dans les deux cas, $\chi^2(f_{(W,Y);0}, f_{(W,Y);1}) = O(1/n)$ Donc la condition (C2) est vérifiée et de plus,

$$|m_0(0) - m_1(0)|^2 = (C')^2 \frac{\Delta_X^2(0)}{f_1(0)^2} = (C')^2 \frac{a_n^{-2\beta} f_0(0)^2}{\pi^2} \geq const. a_n^{-2\beta}$$

On remarque que si ϵ_n , converge plus rapidement que $a_n^{-2\beta}$, c'est à dire si $\epsilon_n = O(a_n^{-2\beta})$, alors

$$|m_0(0) - m_1(0)|^2 \geq const. \epsilon_n,$$

et donc la condition (C1) est validée, et donc il y a non convergence au taux ϵ_n de \hat{m} pour $x = 0$, mais en fait la preuve est identique pour n'importe quel x .

□

Ce manuscrit est basé sur le livre de Alexander Meister intitulé Deconvolution problems in nonparametric statistics (voir [1])

References

- [1] Alexander Meister. *Deconvolution Problems in Nonparametric Statistics*. Springer Berlin, Heidelberg, 2009. ISBN: 978-3-540-87556-7. DOI: 10.1007/978-3-540-87557-4.