N° d'ordre : 15/2014 -M/MT

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE
FACULTE DE MATHEMATIQUES

MÉMOIRE
Présenté pour l'obtention du diplôme de MAGISTER

EN : MATHEMATIQUES
SPÉCIALITÉ : PROBABILITÉS ET STATISTIQUE

Par : GHETTAB SARAH

Sujet

Estimation Non Paramétrique Par La Méthode Du Noyau : Applications à Des Données Censurées

Soutenu publiquement le 18/06/2014, devant le jury composé de :

Mme. H GUERBYENNE	Professeur	à l'USTHB	Présidente.
Mme. Z GUESSOUM	Maître de Conférence/A	à l'USTHB	Directrice de Mémoire.
Mme. D MERAD	Maître de Conférence/A	à l'USTHB	Examinatrice.
M. A TATACHAK	Maître de Conférence/A	à l'USTHB	Examinateur.

Remerciements

Tout d'abord, je tiens à remercier Dieu le tout puissant de m'avoir donné le courage, la force et la santé qui m'ont permis d'achever ce travail dans de bonnes conditions.

Je voudrais en premier lieu exprimer mes remerciements les plus sincères et les plus chaleureux à Madame Zohra GUESSOUM Maître de Conférence à l'USTHB, ma directrice du mémoire, pour m'avoir guidé et encouragé,de m'aoir consacré une partie de son temps precieux et d'avoir lu et relu chaque étape de ce travail : MERCI.

Je remercie vivement Madame Hafida GUERBYENNE Professeur à l'USTHB,de m'avoir fait l'honneur de présider ce jury,.

Mes remerciements sincères à ceux qui ont accepté sans hésitation d'examiner ce travail : Mme. Djenat MERAD Maître de Conférence à l'USTHB, et Mr. Abdelkader TATACHAK Maître de Conférence à l'USTHB.

Je remercier également Madame Ourida SADKI qui m'a initié dans ce travailet pour des raisons personnels n'a pas pu me suivre.

Je tiens à remercier également les membres du groupe du travail constitué par : Amina, Kenza, Nacera et moi meme, dirigé par Mme. GUESSOUM et Mr. TATACHAK avec lesquels j'avoue que j'ai appris beaucoup de choses et cela dans une ambiance conviviale et chaleureuse.

Mes remerciements vont aussi vers Nawel KHELLOUF et Arsalane GUIDOUM pour leur aide et à mes amies proches : Amina Mouna Fadila Yasmina et Hiba pour leur soutien moral.
Le plus fort de mes remerciements est pour mes chers parents et mes soeurs sans oublier les frères, qui ont toujours crû en moi, m'ont aidé et soutenu soit par leurs encouragements ou encore par leurs disponibilités.

Je remercie ma belle mère, mon beau père, mes belles soeurs et beaux frères qui m'ont toujours encouragés

J'adresse, évidement, de tendres remerciements à mon époux qui a été compréhensif et m'a supporté le long de la réalisation de ce travail. Et à ma petite princesse *Razane*.

Résumé

L'objectif de ce mémoire est d'étudier les propriétés de certains estimateurs non paramétriques dans le cas de données complètes et incomplètes. Une attention particulière est donnée au choix du paramètre de lissage (bandwidth) et on s'est intéressé principalement aux trois méthodes, le plug-in, la méthode de validation croisée non biaise ("UCV"), et la méthode du maximum de vraisemblance avec validation croisée ("MLCV"), ainsi qu'au choix du noyau . Une extension des résultats existants dans le cas de données censurées et où le noyau est symétrique est donnée en considérant un noyau non symétrique particulier : le noyau bêta. Des comparaisons par simulation ont été effectuées pour conforter nos résultats.

Mots clés : données censurées, estimation non paramétrique, paramètre de lissage, fonction de régression, fonction de densité, fonction de répartition, noyau, noyau béta, validation croisée.

Abstract

The aim of this thesis is to study the properties of some non-parametric estimators in the case of complete and incomplete data. Particular attention is given to the choice of the smoothing parameter (bandwidth), we are primarily interested in the three methods, the plug-in, Unbiaised Cross-Validation "UCV", and Maximum Likelihood Cross-Validation "MLCV", as well as the choice of kernel. An existing extension of results in the case of censored data and where the kernel is symmetric is finally given by considering a particular non-symmetric kernel : bêta kernel. Comparison by simulation are performed to consolidate our results.

Keywords : censored data, nonparametric estimation, smoothing parameter, density function, regression function, distribution function, kernel,bêta kernel regression, cross-validation.

Table des matières

Liste des tableaux

Table des figures

Introduction Générale

L'objectif principal de la statistique est de faire comprendre, à partir d'observations d'un phénomène aléatoire, la façon dont ces observations sont distribuées et d'étudier la loi de probabilité de la variable aléatoire sous-jacente en vue d'analyser le phénomène ou de prévoir un événement futur. Pour réduire la complexité du phénomène étudié, nous pouvons utiliser deux approches statistiques : non-paramétrique et paramétrique.

Dans le premier cas, le problème de l'estimation non-paramétrique consiste donc à estimer, à partir des observations, une fonction inconnue (par exemples : une fonction de densité f ou une fonction de régression r), élément d'une certaine classe fonctionnelle assez massive. En opposition, l'approche paramétrique cherche à représenter la distribution des observations par une fonction densité $f(x/\theta)$ où le paramètre θ est la seule inconnue.

Dans plusieurs cas, l'approche non paramétrique est préférable car les modèles statistiques qui expliquent plus précisément les données sont souvent plus complexes : Les inconnues de ces modèles sont, en général, des fonctions possédant certaines propriétés de régularité, et non plus les paramètres sous-jacents de certaines fonctions connues.

Historiquement, c'est en 1962 que *John Graunt* a utilisé pour la première fois l'histogramme, qui est l'estimateur non paramétrique de la densité le plus ancien et qui a été ensuite le sujet d'études de plusieurs chercheurs comme Scoutt D.W(1979) [43]. Un autre estimateur de la densité plus moderne est l'estimateur à noyau, qui a été proposé par *Rosenblatt*(1962) [41] et développé par *Parzen* (1962) [37], dont les travaux ont été étendus à la régression par la méthode du noyau par *Nadaraya*(1964) [36] et *Watson* (1964). D'autre auteurs ont étudié l'estimation non paramétrique dans le cas des données censurées à savoir : *Blum* et *Susarla* (1980) qui ont introduit l'estimateur de la densité, dont les propriétés asymptotiques ont été étudiés par *Diehl, S. and Stute, W. (1988)* [17], et aussi par *Zhang B* (1996) [46]. *Carbonez* et al (1985) [7], *Kohler* et *al* (2002) [31] se sont intéressés à l'estimateur de la régression ; ainsi que *Guessoum* et *Ould-Said*(2009) [25] qui ont étudié la convergence uniforme presque sur. Dans le cas où l'hypothèse de symétrie du noyau n'est pas remplie on trouve *Chen S,X (1999)* [8] qui a introduit l'estimateur à noyau bêta de la densité et *Bouezmarni T. and Rolin J-M. (2003)* [3] ont étudié la consistance de cet estimateur.

Dans toute la suite $(X_1, X_2, ..., X_n)$ sont n variables aléatoires indépendantes et identiquement distribuées (i.i.d) de fonction de répartition F et admettant une densité inconnue $f = F'$. Le but est d'estimer (à partir des observations) la densité f en faisant le moins d'hypothèses possibles sur cette densité. Typiquement, on supposera que $f \in \mathcal{F}$ espace

fonctionnel et on notera $\widehat{f_n}$ un estimateur de f.

Ce mémoire est composé de cinq chapitres :

Chapitre 1 : Dans ce chapitre nous présentons quelques rappels sur les divers outils mathématiques que nous serons amenés à utiliser dans les chapitres suivants. En particulier, nous rappellerons les différents types de convergence et nous exposerons quelques exemples de problèmes non-paramétrique.

Chapitre 2 : Ce chapitre est consacré à l'étude d'estimateurs non paramétriques pour des données complètes. Nous commençons par présenter un certain nombre d'estimateurs de la fonction densité dont l'Histogramme (voire Scoutt D.W(1979) [43]), l'Histogramme mobile, celui de *Parzen-Rosenbllat* pour lequel nous donnons la consistance au sens de la moyenne quadratique en calculant explicitement le biais et la variance asymptyotiques. Le choix du paramètre de lissage y est discuté : nous nous intéressons principalement à trois méthodes d'estimations : le plug-in , la méthode de validation croisée non biaisée (Unbiaised Cross-Validation "UCV"), et la méthode du maximum de vraisemblance avec validation croisée (Maximum Likelihood Cross-Validation "MLCV"). Nous clôturons ce chapitre par quelques résultats sur l'estimateur non paramétrique de la fonction de régression à savoir celui de *Nadaraya-Watson* dont on rappelle les propriétés asymptotiques.

Chapitre 3 : Nous faisons un rappel sur les modèles censurés et nous présentons les résultats sur les estimateurs non paramétriques de la densité (*Diehl* et *Stute* (1988)[17]) et de la régression (Guessoum, Z. et Ould-Said, E (2009)[25]) dans le cas censuré.

Chapitre 4 : Ce chapitre est consacré essentiellement à l'étude de l'estimateur de densité dans le cas de noyau bêta (proposé par *Chen* (2000)[8]), en montrant l'influence du choix de ce noyau sur l'estimateur de densité et de régression par une étude comparative avec un noyau standard en particulier le noyau gaussien. Notre contribution est donnée sous forme d'un résultat sur l'estimation à noyau non symétrique de la fonction de régression.

Chapitre 5 : Ce chapitre présente les simulations numérique et les résultats obtenues dans les chapitre précédents.

Nous cloturons notre travail par une conclusion qui recapitule ce que nous avons presenté.

Chapitre 1

Rappels et Définitions

Sommaire

1.1 Espace fonctionnel

Définition 1.1 (***Espace fonctionnel***)
On appelle espace fonctionnels, un ensemble $\mathcal{F}$ de fonctions ayant une structure d'espace vectoriel.
Les normes usuelles sur un espace fonctionnels sont définies par

$$\|u\|_{\mathbb{L}^p} = (\int |u|^p)^{1/p}, p \in [1, +\infty[\text{ et } \|u\|_{\mathbb{L}^\infty} = \sup|u|$$

Exemple 1.1 $C^p([a,b])$ désigne l'espace des fonctions définies sur l'intervalle $[a,b]$, dont toutes les dérivées jusqu'à l'ordre p existent et sont continue sur $[a,b]$.

Définition 1.2 ***(Distance)***
Une **distance** sur l'espace $\mathcal{F}$ est une application $d : \mathcal{F} \times \mathcal{F} \longrightarrow \mathbb{R}^+$ vérifiant les propriétés suivantes :

(1) $d(f,g) = d(g,f)$,pour tout $f,g \in \mathcal{F}$
(2) $d(f,g) = 0$ si et seulement si $f = g$
(3) $d(f,h) \leq d(f,g) + d(g,h)$,pour tout $f,g,h \in \mathcal{F}$

un espace **métrique**$(\mathcal{F},d)$ est un espace $\mathcal{F}$ muni d'une distance d

Exemple 1.2

- $d(f,g) = \| f-g \|_p = [\int | f-g |^p \, dx]^{1/p}, pour p \geq 1$ appelée distance $\mathbb{L}_p$
- $d(f,g) = \| f-g \|_\infty = \sup_x | f(x)-g(x) |$, appelée norme Sup

- $d(f,g) = \mid f(x_0) - g(x_0) \mid$ où x_0 fixé, appelée distance absolue

Définition 1.3 ***(La fonction de perte)***
On appelle fonction de **perte**, la fonction $w : \mathbb{R} \longrightarrow \mathbb{R}^+$ telle que

- w est convexe i.e $\forall (x_1,x_2) \in \mathbb{R}, \forall \lambda \in [0,1], w(\lambda x_1 + (1-\lambda)x_2) \leq \lambda w(x_1) + (1-\lambda)w(x_2)$
- $w(0) = 0$

Exemple 1.3 $w : u \to u^2$,fonction de perte quadratique

Définition 1.4 ***(La fonction de risque)***
On appelle fonction **risque** la fonction définie par

$$R(\widehat{f}_n, f) = \mathbb{E}_f \left(w \left(d(\widehat{f}_n, f) \right) \right)$$

où $\mathbb{E}_f$ désigne l'espérance quand la densité vaut f et w la fonction perte et d la distance

Exemple 1.4

- En prenant la distance ponctuelle en x_0 et la perte quadratique,on obtient le risque quadratique ponctuel (ou plus souvent **l'erreur quadratique moyenne** que nous noterons **MSE** (pour mean square error)
$$R(\widehat{f}_n, f) = MSE = \mathbb{E}_f \mid \widehat{f}_n(x_0) - f(x_0) \mid^2$$
- En prenant la distance $\mathbb{L}_2$ et la perte quadratique, on obtient le risque quadratique intégré ou plus souvent **l'erreur quadratique moyenne intégrée** que nous noterons **MISE** (pour Mean Integred Square Error)
$$R(\widehat{f}_n, f) = MISE = \mathbb{E}_f \parallel \widehat{f}_n - f \parallel_2^2 = \mathbb{E}_f \int \left(\widehat{f}_n(x) - f(x) \right)^2 \, dx$$

1.2 Les différents types de convergence

On considère une suite de variable aléatoire (v.a) $X_1, X_2, \ldots, X_n$ notée $(X_n)_n$. On peut définir plusieurs modes de convergence pour une telle suite. On notera F_{X_n} la fonction de répartition de X_n.

1.2.1 La convergence en loi

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

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

et on note

$$X_n \underset{n \to \infty}{\overset{\mathcal{L}}{\to}} X$$

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

1.2.2 La convergence en probabilité

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

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

et on note

$$X_n \underset{n\to\infty}{\overset{p}{\to}} X$$

1.2.3 La convergence presque sûre

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

$$\lim_{n\to\infty} \mathbb{P}(\sup_{m\geq n} \{|X_m - X|\} < \varepsilon) = 1$$

et on note

$$X_n \underset{n\to\infty}{\overset{p.s}{\to}} X$$

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

1.2.4 La convergence en moyenne quadratique

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

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

et on note

$$X_n \underset{n\to\infty}{\overset{m.q}{\to}} X$$

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

$$X_n \underset{n\to\infty}{\overset{p}{\to}} X \Rightarrow X_n \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} X$$

$$X_n \underset{n\to\infty}{\overset{m.q}{\to}} X \Rightarrow X_n \underset{n\to\infty}{\overset{p}{\to}} X$$

$$X_n \underset{n\to\infty}{\overset{p.s}{\to}} X \Rightarrow X_n \underset{n\to\infty}{\overset{p}{\to}} X$$

1.3 Quelques exemples de problèmes non-paramétriques

1.3.1 Estimation d'une fonction de répartition

L'estimateur naturel de la fonction de répartition F est la fonction de répartition empirique F_n. C'est un estimateur non paramétrique de F.

Définition 1.5 ***(fonction de répartition empirique)***

Soit $(X_1, X_2, \ldots, X_n)$ une suite de variables aléatoires i.i.d à valeurs réelles de densité de probabilité f et de fonction de répartition $F(x) = \int_{-\infty}^{x} f(t)dt$. Soient $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ les statistiques d'ordres associées. Soit p_n le nombre d'observation inférieures ou égales à x dans l'échantillon. Ainsi, la fonction de répartition empirique $F_n(x)$ basée sur l'échantillon est une fonction en escalier définie par :

$$\begin{aligned} F_n(x) &= \frac{p_n}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(X_i \leq x)} \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(X_{(i)} \leq x)} \end{aligned}$$

D'où

$$F_n(x) = \begin{cases} 0 & \text{si} \quad x < X_{(1)} \\ \dfrac{k}{n} & \text{si} \quad X_{(k)} \leq x < X_{(k+1)}, \quad k = 1, \ldots, n-1 \\ 1 & \text{si} \quad x \geq X_{(n)} \end{cases}$$

La figure 1.1 présentent une simulation sous langage Matlab de la fonction de répartition dans un intervalle $[-3,3]$ avec un pas $h = 0.01$ sur la base d'un échantillon simulé $X \sim \mathcal{N}(0,1)$.(voir le code du programme dans l'annexe)

1.3.1.1 Propriétés de la fonction de répartition empirique :

a) Propriétés ponctuelles de $F_n(x)$

Biais

$$\begin{aligned} \mathbb{E}[F_n(x)] &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i \leqslant x\}}\right) \\ &= \frac{1}{n} n \quad \mathbb{E}\left(\mathbb{I}_{\{X \leqslant x\}}\right) \\ &= \mathbb{P}(X \leqslant x) \\ &= F(x) \end{aligned}$$

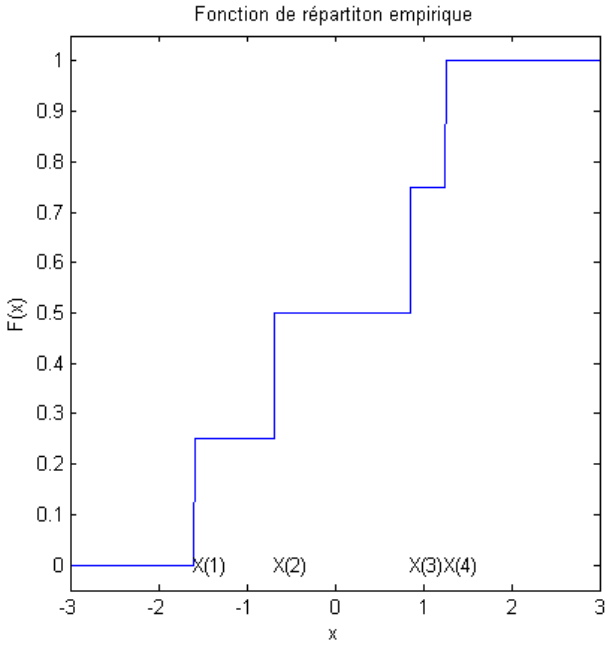

FIGURE 1.1 – Fonction de répartition empirique

donc F_n est un estimateur **sans biais** de F ie $\mathbb{B}\text{iais}(F_n(x)) = 0$

Variance

$$\begin{aligned}
\mathbb{V}ar[F_n(x)] &= \mathbb{E}\left(F_n^2(x)\right) - \mathbb{E}^2\left(F_n(x)\right) \\
&= \mathbb{E}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{\{X_i\leq x\}}\right)^2\right\} - F^2(x) \\
&= \mathbb{E}\left\{\frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{I}^2_{\{X_i\leq x\}} + \sum_{i\neq j}\mathbb{I}_{\{X_i\leq x\}}\mathbb{I}_{\{X_j\leq x\}}\right)\right\} - F^2(x) \\
&= \frac{1}{n^2}\left\{\sum_{i=1}^{n}\mathbb{E}(\mathbb{I}^2_{\{X_i\leq x\}}) + \sum_{i\neq j}\mathbb{E}(\mathbb{I}_{\{X_i\leq x\}}\mathbb{I}_{\{X_j\leq x\}})\right\} - F^2(x) \\
&= \frac{1}{n^2}\left\{\sum_{i=1}^{n}\mathbb{E}(\mathbb{I}^2_{\{X_i\leq x\}}) + \sum_{i\neq j}\mathbb{E}(\mathbb{I}_{\{X_i\leq x\}})\mathbb{E}(\mathbb{I}_{\{X_j\leq x\}})\right\} - F^2(x)
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}ar[F_n(x)] &= \frac{1}{n^2}\left\{n\mathbb{E}\left(\mathbb{I}_{\{X\leqslant x\}}\right)+n(n-1)\mathbb{E}\left(\mathbb{I}_{\{X\leqslant x\}}\right)\mathbb{E}\left(\mathbb{I}_{\{X\leqslant x\}}\right)\right\}-F^2(x)\\
&= \frac{1}{n}[\mathbb{P}(X\leqslant x)+\frac{n(n-1)}{n^2}\mathbb{P}^2(X\leqslant x)]-F^2(x)\\
&= \frac{1}{n}F(x)+F^2(x)-\frac{1}{n}F^2(x)-F^2(x)\\
&= \frac{1}{n}\left(1-F(x)\right)F(x)
\end{aligned}$$

donc $\mathbb{V}ar\left[F_n(x)\right] \underset{n\to\infty}{\longrightarrow} 0$

Erreur en moyenne quadratique(MSE)

$$\begin{aligned}
MSE(F_n) &= \mathbb{E}\left\{(F_n-F)^2\right\}\\
&= \mathbb{E}\left\{\left[F_n-\mathbb{E}(F_n)+\mathbb{E}(F_n)-F\right]^2\right\}\\
&= \mathbb{E}\left\{\left[F_n-\mathbb{E}(F_n)\right]^2+2\left[\left(F_n-\mathbb{E}(F_n)\right)\left(\mathbb{E}(F_n)-F\right)\right]+\left[\mathbb{E}(F_n)-F\right]^2\right\}\\
&= \mathbb{E}\left\{\left[F_n-\mathbb{E}(F_n)\right]^2\right\}+2\mathbb{E}\left\{\left[\left(F_n-\mathbb{E}(F_n)\right)\left(\mathbb{E}(F_n)-F\right)\right]\right\}+\mathbb{E}\left\{\left[\mathbb{E}(F_n)-F\right]^2\right\}\\
&= \mathbb{E}\left\{\left[F_n-\mathbb{E}(F_n)\right]^2\right\}+2\mathbb{E}\left\{F_n-\mathbb{E}(F_n)\right\}\left(\mathbb{E}(F_n)-F_n\right)+\mathbb{E}\left\{\left[\mathbb{E}(F_n)-F\right]^2\right\}\\
&= \mathbb{E}\left\{\left[F_n-\mathbb{E}(F_n)\right]^2\right\}+\mathbb{E}\left\{\left[\mathbb{E}(F_n)-F\right]^2\right\}\\
&= \mathbb{V}ar(F_n)+Biais(F_n)^2\\
&= \mathbb{V}ar(F_n)
\end{aligned}$$

d'où $MSE(F_n) \underset{n\to\infty}{\longrightarrow} 0$

Convergence en probabilité d'aprés l'inégalité de Chebyshev, la convergence en moyenne quadratique implique la convergence en probabilité,en effet

$$\forall\varepsilon>0, \qquad \mathbb{P}\left(|F_n(x)-F(x)|\geq\varepsilon\right)\leq\frac{\mathbb{V}ar\left(F_n(x)\right)}{\varepsilon^2}\underset{n\to\infty}{\longrightarrow}0$$

alors

$$F_n(x)\underset{n\to\infty}{\overset{p}{\to}}F(x)$$

La loi forte des grands nombres (LGN) Notons que $F_n(x)$ est la moyenne empirique d'une suite de variables aléatoire i.i.d $Z_i=\mathbb{I}_{(X_i\leqslant x)}$ (elles sont indépendantes comme fonctions respectives des X_i) toutes issues de la loi de Bernoulli de paramètre $p=F(x)$. Ainsi la loi des grands nombres s'applique et $F_n(x)$ converge aussi presque sûrement vers $F(x)$ c-à-d

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{X_i\leq x}\underset{n\to\infty}{\overset{p.s}{\to}}\mathbb{E}(\mathbb{I}_{X\leq x})=\mathbb{P}(X\leq x)=F(x)$$

$$F_n(x)\underset{n\to\infty}{\overset{p.s}{\to}}F(x)$$

Théorème Centrale Limite (TCL)

$$\sqrt{n}\,(F_n(x)-F(x)) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,F(x)(1-F(x)))$$

Loi du Logarithme itéré

Rappel si $\{Y_i\}_{i\geq 0}$ est une suite de v.a.i.i.d, centrées de variance $\sigma^2 < +\infty$ et $S_n = \sum_{i=1}^{n} Y_i$. Alors

$$\limsup_{n\to\infty} \frac{|S_n|}{\sigma\sqrt{2n\log\log n}} = 1 \qquad p.s$$

En particulier pour les variables $Y_i = \mathbb{I}_{\{X_i\leq x\}} - \mathbb{E}(\mathbb{I}_{\{X_i\leq x\}})$

$$\begin{aligned}
S_n &= \sum_{i=1}^{n} Y_i \\
&= \sum_{i=1}^{n} \mathbb{I}_{\{X_i\leq x\}} - \sum_{i=1}^{n} \mathbb{E}(\mathbb{I}_{\{X_i\leq x\}}) \\
&= \frac{n}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i\leq x\}} - \sum_{i=1}^{n} \mathbb{P}(X_i\leq x) \\
&= n\ F_n(x) - n\ \mathbb{P}(X\leq x) \\
&= n\ F_n(x) - n\ F(x) \\
&= n\ (F_n(x)-F(x))
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}ar(Y_i) &= \mathbb{E}(Y_i)^2 \\
&= \mathbb{E}\left\{(\mathbb{I}_{\{X_i\leq x\}} - \mathbb{E}(\mathbb{I}_{\{X_i\leq x\}}))^2\right\} \\
&= \mathbb{E}\left\{\mathbb{I}^2_{\{X_i\leq x\}} - 2\mathbb{I}_{\{X_i\leq x\}}\ \mathbb{E}(\mathbb{I}_{\{X_i\leq x\}}) + \mathbb{E}^2(\mathbb{I}_{\{X_i\leq x\}})\right\} \\
&= \mathbb{E}(\mathbb{I}_{\{X_i\leq x\}}) - 2\mathbb{E}(\mathbb{I}_{\{X_i\leq x\}})\mathbb{P}(X_i\leq x) + \mathbb{P}^2(X_i\leq x) \\
&= \mathbb{P}(X_i\leq x) - 2\mathbb{P}^2(X_i\leq x) + \mathbb{P}(X_i\leq x)^2 \\
&= \mathbb{P}(X_i\leq x) - \mathbb{P}(X_i\leq x)^2 \\
&= F(x) - F(x)^2 \\
&= F(x)(1-F(x))
\end{aligned}$$

Alors

$$\limsup_{n\to\infty} \frac{\sqrt{n}|F_n(x)-F(x)|}{\sqrt{F(x)(1-F(x))2\log\log n}} = 1 \qquad p.s$$

b)Propriétés uniformes

Théorème de Glivenko-Cantelli (G.C) Le théorème suivant est essentiel car il montre la convergence uniforme, sur $\mathbb{R}$, de F_n vers F

Théorème 1.1 Soit un échantillon aléatoire$(X_1, X_2, ..., X_n)$ issu de la loi de fonction de répartition F et F_n sa fonction de répartition empirique

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \overset{p.s}{\underset{n\to\infty}{\to}} 0$$

Remarque 1.1 Pour voir les choses concrètement, ce théoreme nous dit que l'on peut être assuré que l'écart maximal entre F_n et F va tendre vers 0 si l'on augmente la taille de l'échantillon à l'infini ou encore que partout, simultanément, la fonction de répartition empirique va se rapprocher de la vraie fonction de répartition.

Inégalité de Dvoretzky-Kiefer-Wolfowitz (DKW) En général on utilise l'inégalité de (DKW) pour construire des intervalles de confiances (IC) exacts pour $F(x)$

$$\forall n \in \mathbb{N}, \forall \varepsilon > 0, \qquad \mathbb{P}\left(\sup |F_n(x) - F(x)| > \varepsilon\right) \leq 2\exp^{-2n\varepsilon^2}$$

en effet $\forall\ x \in \mathbb{R}$, on a

$$\begin{aligned}\mathbb{P}\left(F(x) \in [F_n(x) - \varepsilon; F_n(x) + \varepsilon]\right) &= 1 - \mathbb{P}(|F_n(x) - F(x)| > \varepsilon) \\ &\geq 1 - \mathbb{P}(\sup|F_n(x) - F(x)| > \varepsilon) \\ &\geq 1 - 2e^{-2n\varepsilon^2}\end{aligned}$$

Soit $\alpha > 0$ le seuil de l'intervalle de confiance, on peut toujours trouver $\varepsilon > 0$ tel que $2e^{-2n\varepsilon^2} = \alpha$ i.e on prend $\varepsilon = \sqrt{\log(\frac{2}{\alpha})/(2n)}$ et on obtient

$$\mathbb{P}(F(x) \in \left[F_n(x) - \sqrt{\log(\tfrac{2}{\alpha})/(2n)}; F_n(x) + \sqrt{\log(\tfrac{2}{\alpha})/(2n)}\right]) \geq 1 - \alpha$$

donc $\left[F_n(x) - \sqrt{\log(\frac{2}{\alpha})/(2n)}; F_n(x) + \sqrt{\log(\frac{2}{\alpha})/(2n)}\right]$ est un intervalle de confiance au niveau $1 - \alpha$ pour $F(x)$

1.3.1.2 Estimation des fonctionnelles

Une fonctionnelle est une fonction régulière T définie sur l'ensemble des fonctions de répartition $\mathcal{F}$

Remarque 1.2 Une fonction régulière est une fonction dérivable, sans points de singularité (point où la dérivée n'a pas la même valeur, à gauche qu'à droite du point singulier, en question) Les fonctions suivantes sont régulières : Lipschitzienne, classe C^k, fonctions monotones, ... etc

Exemple 1.5 Quelques exemples de fonctionnelles régulières

- Moyenne : $F \longrightarrow T(F) = \mu(F) = \int x dF(x)$
- Variance : $F \longrightarrow T(F) = \sigma^2(F) = \int (x - \mu(F))^2 dF(x)$
- Médiane : $m(F) = F^{-1}(1/2)$ et plus généralement les quantiles :$F \longrightarrow q_\alpha(F) = F^{-1}(\alpha)$ (L'inverse de la fonction de répartition (*f.d.r*) est la fonction quantile)
- Skewness(ou coefficient d'asymétrie) :$F \longrightarrow \left\{ \int (x - \mu(F))^3 dF(x) \right\} / \sigma(F)^{3/2}$

Remarque 1.3 La densité $f = F^{'}$ n'est pas une fonctionnelle régulière de la densité

Méthode *plug-in* pour l'estimation de fonctionnelles Si $T : F \longrightarrow T(F)$ est une fonctionnelle alors un estimateur naturel de $T(F)$ est obtenu en **injectant** l'estimateur F_n de F dans l'expression de T, i.e. $\widehat{T} = T(F_n)$ est un estimateur naturel de $T(F)$.

Exemple 1.6 Les estimateurs suivants sont obtenus par la méthode plug-in

- Moyenne empirique : $\overline{X_n} = \int x \ dF_n(x) = \frac{1}{n} \sum_{i=1}^{n} X_i$ c'est un estimateur sans biais car

$$\begin{aligned} \mathbb{E}(\overline{X_n}) &= \mathbb{E}(\frac{1}{n} \sum_{i=1}^{n} X_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) \\ &= \frac{1}{n} n \, \mathbb{E}(X) \\ &= \int x \ dF(x) \\ &= \mu(F) \end{aligned}$$

- Variance empirique : $\widehat{\sigma}_n^2 = \int x^2 \ dF_n(x) - (\int x \ dF_n(x))^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - (\frac{1}{n} \sum_{i=1}^{n} X_i)^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X_n})$
  cet estimateur est biaisé, puisque

$$\begin{aligned} \mathbb{E}(\widehat{\sigma}_n^2) &= \mathbb{E}(X_1^2) - \frac{1}{n^2} \left(\sum_{i=1}^{n} \mathbb{E}(X_1^2) + \sum_{i \neq j} \mathbb{E}(X_i \, X_j) \right) \\ &= \mathbb{E}(X_1^2) - \frac{1}{n} \, \mathbb{E}(X_1^2) - \frac{n-1}{n} \, (\mathbb{E}(X_1))^2 \\ &= (1 - \frac{1}{n}) \left(\mathbb{E} \, (X_1^2) - (\mathbb{E} \, X_1)^2 \right) \\ &= (1 - \frac{1}{n}) \sigma^2 \end{aligned}$$

On lui préfère souvent $S_n^2 = (\frac{1}{n-1}) \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ qui est sans biais, appelée variance empirique corrigé

- Médiane empirique $\widehat{m}_n = F_n^{-1}(1/2)$ ou plus généralement le quantile empirique $\widehat{q}_\alpha = F_n^{-1}(\alpha)$

1.3.2 Estimation d'une densité de probabilité

Lorsque l'on cherche à étudier une suite de mesures provenant de la répétition d'une expérience, une méthode de modélisation consiste à supposer que ces mesures sont des réalisations de variables aléatoires indépendantes équi-distribuées. Comprendre ces mesures et la façon dont elles sont distribuées revient à étudier la loi de probabilité de la variable aléatoire sous-jacente. Lorsque l'on n'a pas d'idée a priori sur la forme particulière que peut prendre la densité f, construire un estimateur de f ne se résume pas à l'estimation d'une moyenne et d'une variance, comme c'est le cas pour des lois paramétrique. Il s'agit de reconstruire la loi dans sa globalité par une fonction. Le problème est alors dit **non-paramétrique**. Différentes méthodes existent pour estimer la densité f comme l'histogramme et l'estimateur à noyau que l'on va les étudier en détail au deuxième chapitre.

1.3.3 Régression non paramétrique

La méthode la plus communément utilisée pour étudier la relation entre deux variables est la régression linéaire simple, qui suppose un modèle de la forme

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

où les erreurs aléatoires ε_i sont non corrélées, de moyennes nulles et de variances σ^2 indépendantes de X. Cette méthode paramétrique possède l'avantage d'être facile à interpréter et, lorsque les hypothèses sur les résidus ε_i sont vérifiés, permet de faire des tests d'hypothèses statistique formels sur les paramètre. Par contre lorsque les hypothèses ne sont pas vérifiées ou lorsque la structure du modèle n'a pas essentiellement d'intérêt, on préfère choisir un modèle plus flexible qui reflète mieux la relation entre X et Y. On utilise le modèle de régression non paramétrique

$$Y_i = m(X_i) + \varepsilon_i$$

où ε_i représente la variation de Y_i autour de X_i et obéit aux mêmes hypothèses que dans le cas linéaire et à part certaines conditions de continuités et de lissages, il n'y a habituellement aucune contrainte associée à $m(X_i)$. On cherchera, dans une famille fixée de fonctions, quelle est celle pour la quelle les Y sont les plus proches de $m(X)$. Cette proximité se mesure en général par un risque utilisant l'erreur quadratique moyenne (MSE), et on essayera alors de déterminer la fonction $m^*(X_i)$ qui rendra cette erreur la plus petite possible, c'est à dire à trouver une fonction $m^*(X_i)$ qui minimise l'erreur quadratique moyenne :

$$\mathbb{E}|m^*(X_i) - Y|^2 = \min_m \mathbb{E}|m(X_i) - Y|^2$$

Il est connu que ce minimum est donné par l'espérance conditionnelle :

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

en effet, déterminer $m^*(X_i)$ revient à calculer $\arg\min_m \mathbb{E}((m(X)-Y)^2|X=x)$. En différenciant l'espérance $\mathbb{E}((m(X)-Y)^2|X=x)$ par rapport à $m(.)$, en égalant le résultat à 0, et finalement en isolant $m(.)$, on obtient :

$$\begin{aligned}
\frac{\partial}{\partial m}\mathbb{E}((m(X)-Y)^2|X=x) &= \frac{\partial}{\partial m}\mathbb{E}\left((m(X))^2+Y^2-2\,m(Y)\,Y|X=x\right)\\
&= \frac{\partial}{\partial m}\left\{\mathbb{E}((m(X))^2|X=x)+\mathbb{E}(Y^2|X=x)-2\,\mathbb{E}(m(X)\,Y|X=x)\right\}\\
&= \frac{\partial}{\partial m}\left\{(m(x))^2+\mathbb{E}(Y|X=x)-2\,m(x)\mathbb{E}(Y|X=x)\right\}\\
&= 2m(x)-2\mathbb{E}(Y|X=x)\\
&= 0
\end{aligned}$$

ce qui implique que

$$m^*(x) = \mathbb{E}(Y|X=x) := r(x)$$

le fait que la dérivée seconde, qui est égale à 2, soit positive nous permet de conclure que c'est bien un minimum.

Chapitre 2

Estimation non paramétrique sur des données completes

Sommaire

2.1 Introduction

Dans ce chapitre on s'intéresse principalement à l'estimation de la fonction de densité f. En effet, cette dernière est un concept fondamental dans la statistique. Spécifier la fonction densité f d'une variable aléatoire $X : (\Omega, \mathcal{A}, \mathbb{P} \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ donne une description de la distribution sous-jacente de X. Quand elle ne peut pas être spécifiée, une estimation de cette densité peut être effectuée en utilisant un échantillon de X.
Dans une seconde partie, l'estimation non paramétrique de la fonction de régression est abordée à travers l'estimateur de *Nadaraya-Watson*.

2.2 L'histogramme

L'histogramme est d'estimateur non paramétrique de densité le plus ancien, et qui probablement remonte aux études de mortalité de *John Graunt (1662)* d'après ([43], page 606). Aujourd'hui, l'histogramme reste un outil statistique important pour la présentation et la synthèse des données.

On considère que l'histogramme est défini sur une grille de découpage équidistants $\{\ldots, \alpha_{n(i)}; -\infty < i < \infty\}$ avec la largeur de chaque intervalle $h_n = \alpha_{n(i+1)} - \alpha_{n(i)}$.

Supposons que $X_1, X_2, \ldots, X_n$, soient des v.a indépendantes et identiquement distribuées de densité de probabilité f, où f est deux fois dérivable et à dérivées bornées. On a besoin d'identifier l'intervalle qui contient un point fixé x si n varie. Soit $I_n(x) =]\alpha_n(x), \alpha_n(x) + h_n]$ cet intervalle. On définit la probabilité d'appartenance à la classe $I_n(x)$ par :

$$\begin{aligned} p_n(x) &= \mathbb{P}(X \in I_n(x)) \\ &= \mathbb{E}(\mathbb{I}_{X \in I_n(x)}) \\ &= \int_{\alpha_n(x)}^{\alpha_n(x)+h_n} f(y)dy \end{aligned}$$

Soit $V_n(x)$ le nombre d'observation appartenant à $I_n(x)$. Ainsi $V_n(x)$ suit une loi binomiale $\mathcal{B}(n, p_n(x))$. l'estimateur par histogramme de la densité est donné par :

$$\widehat{f_n}(x) = \frac{V_n(x)}{(nh_n)} = \frac{I}{(nh_n)} \sum_{i=1}^{n} \mathbb{I}_{\{\alpha_n(x) < X_i < \alpha_n(x)+h_n\}} \tag{2.1}$$

2.2.1 Propriétés asymptotiques de l'estimateur par histogramme

Si on pense à la consistance de cet estimateur, il est clair que $\widehat{f_n}$ sera plus proche de la vraie densité f si la largeur de l'intervalle tend vers 0 quand n tend vers l'infini, mais il ne faut pas que h_n tende vers zéro trop vite pour que l'effectif par classe puisse quand même tendre vers l'infini et assurer la convergence au point x. Il faut donc assurer les conditions suivantes en même temps

$$n \to \infty, \qquad h_n \to 0, \qquad nh_n \to \infty \tag{2.2}$$

2.2.1.1 calcul du biais

On a

$$\begin{aligned}
\mathbb{E}\left(\widehat{f_n}(x)\right) &= \mathbb{E}\left(\frac{V_n(x)}{nh_n}\right) \\
&= \frac{1}{nh_n}\,\mathbb{E}\left(V_n(x)\right) \\
&= \frac{1}{nh_n}\; np_n(x) \\
&= \frac{1}{h_n}\int_{\alpha_n(x)}^{\alpha_n(x)+h_n} f(y)dy
\end{aligned}$$

le biais peut s'écrire :

$$\mathbb{E}\left(\widehat{f_n}(x)\right) - f(x) = \frac{1}{h_n}\int_{\alpha_n(x)}^{\alpha_n(x)+h_n} f(y)dy - f(x)$$

le biais ne dépend pas de la taille de l'échantillon et ne peut être réduit à zéro qu'en faissant tendre h_n vers 0, ce qui implique que

$$\int_{\alpha_n(x)}^{\alpha_n(x)+h_n} f(y)dy \underset{h_n\to 0}{\sim} \delta_x \qquad \text{Loi de dirac en } x$$

En faisant un développement de *Taylor* d'ordre 2 de f au voisinage de x

$$f(y) = f(x) + f^{'}(x)(y-x) + f^{''}(x)(y-x)^2 + o[(y-x)^2], \text{pour } \; y \in I_n(x)$$

comme $(y-x)^2 \leq h_n^2$ et $f^{''}(x)$ bornée alors on peut écrire

$$f(y) = f(x) + f^{'}(x)(y-x) + O(h_n^2) \tag{2.3}$$

le biais s'écrit alors :

$$\begin{aligned}
\mathbb{E}\left(\widehat{f_n}(x)\right) - f(x) &= \frac{1}{h_n}\int_{\alpha_{n(x)}}^{\alpha_{n(x)}+h_n} \left\{f(x) + f^{'}(x)(y-x) + O(h_n^2)\right\}dy - f(x) \\
&= \left(\frac{1}{h_n}\int_{\alpha_{n(x)}}^{\alpha_{n(x)}+h_n} f(x)\,dy + \frac{1}{h_n}\int_{\alpha_{n(x)}}^{\alpha_{n(x)}+h_n} f^{'}(x)\,ydy - \frac{1}{h_n}\int_{\alpha_{n(x)}}^{\alpha_{n(x)}+h_n} f^{'}(x)\,xdy\right) \\
&\quad + O(h_n^2) - f(x) \\
&= f(x) + \frac{1}{h_n} f^{'}(x)\left[\frac{y^2}{2}\right]_{\alpha_{n(x)}}^{\alpha_{n(x)}+h_n} - f^{'}(x)\,x + O(h_n^2) - f(x)
\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(\widehat{f_n}(x)\right) - f(x) &= \frac{1}{2h_n}\left[\left(\alpha_{n(x)}+h_n\right)^2 - \alpha_{n(x)}^2\right] f^{'}(x) - f^{'}(x)\, x + O(h_n^2) \\ &= \frac{1}{2h_n}\left[2\alpha_n(x)h_n + h_n^2\right] f^{'}(x) - f^{'}(x)\, x + O(h_n^2) \\ &= \frac{1}{2}h_n f^{'}(x) + \alpha_{n(x)} f^{'}(x) - f^{'}(x)\, x + O(h_n^2) \\ &= \frac{1}{2}h_n f^{'}(x) + f^{'}(x)\left(\alpha_{n(x)} - x\right) + O(h_n^2)\end{aligned}$$

d'où

$$\mathbb{E}\left(\widehat{f_n}(x)\right) - f(x) = \frac{1}{2}h_n f^{'}(x) - f^{'}(x)\left(x - \alpha_{n(x)}\right) + O(h_n^2) \tag{2.4}$$

L'estimateur par histogramme est un estimateur biaisé de la densité

2.2.1.2 calcul de la variance

On a

$$\begin{aligned}\mathbb{V}ar\left(\widehat{f_n}(x)\right) &= \mathbb{V}ar\left(\frac{V_n(x)}{nh_n}\right) \\ &= \frac{1}{n^2h_n^2}\,\mathbb{V}ar\left(V_n(x)\right) \\ &= \frac{1}{n^2h_n^2}\, np_n(x)(1-p_n(x)) \\ &= \frac{1}{(nh_n^2)}\, p_n(x)(1-p_n(x))\end{aligned}$$

comme $p_n(x) = \int_{\alpha_{n(x)}}^{\alpha_{n(x)}+h_n} f(y)dy$, et en utilisant l'expression (2.3) on a :

$$\begin{aligned}p_n(x) &= \int_{\alpha_{n(x)}}^{\alpha_{n(x)}+h_n} \left\{f(x) + f^{'}(x)(y-x) + O(h_n^2)\right\} \\ &= h_n f(x) + \frac{1}{2}f^{'}(x)\left[h_n^2 - 2h_n(x-\alpha_{n(x)})\right] + O(h_n^3)\end{aligned}$$

d'où on a d'une part

$$p_n(x) = O(h_n) \tag{2.5}$$

d'autre part

$$\alpha_{n(x)} < x < \alpha_{n(x)} + h_n \Longleftrightarrow 0 < x - \alpha_{n(x)} < h_n$$

donc

$$x - \alpha_{n(x)} = O(h_n) \Longleftrightarrow 2\ h_n(x - \alpha_{n(x)}) = 2\ h_n\ O(h_n) = O(h_n^2)$$

ainsi

$$p_n(x) = h_n\ f(x) + O(h_n^2)$$

d'où la variance peut s'écrire

$$\mathbb{V}ar\left(\widehat{f_n}(x)\right) = \frac{1}{(n\ h_n^2)}\left\{h_n\ f(x) + O(h_n^2)\right\}\ \{1 - O(h_n)\}$$

or $1 - O(h_n) \approx 1$ car $O(h_n) \longrightarrow 0$ quand $n \to \infty$ ce qui implique

$$\mathbb{V}ar\left(\widehat{f_n}(x)\right) = \frac{f(x)}{n\ h_n} + O(\frac{1}{n}) \tag{2.6}$$

2.2.2 Erreur quadratique moyenne de l'estimateur par histogramme (MSE)

Le comportement asymptotique du risque quadratique de $\widehat{f_n}$ au point x est donnée par la proposition suivante

Proposition 2.1 sous les hypothèse (2.2) nous obtenons

$$MSE(x) = \frac{f(x)}{nh_n} + \frac{1}{4}h_n^2 f^{'}(x)^2 + f^{'}(x)^2(x - \alpha_{n(x)})^2 - h_n f^{'}(x)^2(x - \alpha_{n(x)}) + O(\frac{1}{n} + h_n^3) \tag{2.7}$$

Démonstration 2.1 La preuve résulte des expressions (2.4) et (2.6) ainsi que de la décomposition en biais-variance

$$MSE(x) = \mathbb{V}ar(\widehat{f_n}) + Biais(\widehat{f_n})^2$$

on a

$$\mathbb{V}ar(\widehat{f_n}) = \frac{f(x)}{nh_n} + O(\frac{1}{n})$$

et d'aprés (2.4)

$$\begin{aligned}\mathbb{V}ar\left(\widehat{f_n}(x)\right) &= \mathbb{V}ar\left(\frac{V_n(x)}{nh_n}\right)\\ &= \frac{1}{n^2h_n^2}\,\mathbb{V}ar\left(V_n(x)\right)\\ &= \frac{1}{n^2h_n^2}\,np_n(x)(1-p_n(x))\\ &= \frac{1}{(nh_n^2)}\,p_n(x)(1-p_n(x))\end{aligned}$$

$$\begin{aligned}B(\widehat{f_n})^2 &= \left(\frac{1}{2}h_nf^{'}(x)-f^{'}(x)(x-\alpha_{n(x)})+O(h_n^2)\right)^2\\ &= \frac{1}{4}h_n^2f^{'}(x)^2+[f^{'}(x)(x-\alpha_{n(x)})]^2+O(h_n^4)-h_nf^{'}(x)^2(x-\alpha_{n(x)})+h_nf^{'}(x)O(h_n^2)\\ &\quad -2f^{'}(x)(x-\alpha_{n(x)})O(h^2)\\ &= \frac{1}{4}h_n^2f^{'}(x)^2+f^{'}(x)^2(x-\alpha_{n(x)})^2+O(h_n^4)-h_nf^{'}(x)^2(x-\alpha_{n(x)})+f^{'}(x)O(h_n^3)\\ &\quad -2f^{'}(x)O(h_n).O(h_n^2)\\ &= \frac{1}{4}h_n^2f^{'}(x)^2+f^{'}(x)^2(x-\alpha_{n(x)})^2+h_nO(h_n^3)-h_nf^{'}(x)^2(x-\alpha_{n(x)})+f^{'}(x)O(h_n^3)\\ &\quad -2f^{'}(x)O(h_n^3)\\ &= \frac{1}{4}h_n^2f^{'}(x)^2+f^{'}(x)^2(x-\alpha_{n(x)})^2+o(h_n^3)-h_nf^{'}(x)^2(x-\alpha_{n(x)})+f^{'}(x)O(h_n^3)\\ &\quad -2f^{'}(x)O(h_n^3)\\ &= \frac{1}{4}h_n^2f^{'}(x)^2+f^{'}(x)^2(x-\alpha_{n(x)})^2+O(h_n^3)-h_nf^{'}(x)^2(x-\alpha_{n(x)})+f^{'}(x)O(h_n^3)\\ &\quad -2f^{'}(x)O(h_n^3)\\ &= \frac{1}{4}h_n^2f^{'}(x)^2+f^{'}(x)^2(x-\alpha_{n(x)})^2-h_nf^{'}(x)^2(x-\alpha_{n(x)})+O(h_n^3)\end{aligned}$$

d'où l'approximation du critère MSE en un point x fixé est

$$MSE(x)=\frac{f(x)}{nh_n}+\frac{1}{4}h_n^2(f^{'}(x))^2+(f^{'}(x))^2(x-\alpha_{n(x)})^2-h_n(f^{'}(x))^2(x-\alpha_{n(x)})+o(\frac{1}{n}+h_n^3) \tag{2.8}$$

2.2.3 Erreur quadratique intégrée moyenne (MISE) de l'estimateur par histogramme

On s'intéresse dans cette partie au risque global de l'estimateur $\widehat{f_n}$ sur toute la droite réelle. On introduit pour cela l'erreur quadratique intégrée moyenne (MISE)

$$\begin{aligned} MISE(x) &:= \mathbb{E}[\int_{\mathbb{R}}(\widehat{f_n}-f(x))^2]dx \\ &= \int_{\mathbb{R}} MSE(x)dx \end{aligned}$$

alors en intégrant l'équation (2.8) sur tout $\mathbb{R}$ on obtient :

$$\begin{aligned} MISE(x) &= \int_{\mathbb{R}}\left[\frac{f(x)}{nh_n}+\frac{1}{4}h_n^2(f^{'}(x))^2+(f^{'}(x))^2(x-\alpha_{n(x)})^2-h_n(f^{'}(x))^2(x-\alpha_{n(x)})\right]dx \\ &= \frac{1}{nh_n}+\frac{1}{4}h_n^2\int_{\mathbb{R}}(f^{'}(x))^2dx+\int_{\mathbb{R}}(f^{'}(x))^2(x-\alpha_{n(x)})^2dx-h_n\int_{\mathbb{R}}(f^{'}(x))^2 \\ &\quad (x-\alpha_{n(x)})dx+O(\frac{1}{n}+h_n^3) \end{aligned}$$

Rappelons que $\alpha_{n(x)}$ désigne la grille de découpage équidistant. Ainsi on peut écrire le troisième terme de l'expression précédent comme suit

$$\begin{aligned} \int_{\mathbb{R}} f^{'}(x)^2(x-\alpha_{n(x)})^2dx &= \sum_{-\infty}^{+\infty}\int_{\alpha_{n(x)}}^{\alpha_{n(x)}+h_n} f^{'}(x)^2(x-\alpha_{n(x)})^2dx \\ &= \sum_{-\infty}^{+\infty}\int_0^{h_n}(f^{'}(\alpha_{n(x)}+y))^2y^2dy \end{aligned} \tag{2.9}$$

par le changement de variable $y = x-\alpha_{n(x)} \Longrightarrow x = y+\alpha_{n(x)}$
en utilisant un développement de *Taylor* à l'ordre 1 au voisinage de $\alpha_{n(x)}$ on a :

$$f^{'}(\alpha_{n(x)}+y) = f^{'}(\alpha_{n(x)})+yf^{''}(\alpha_{n(x)})+o(y)$$

or $0 \leq y \leq h_n \Rightarrow y = O(h_n)$ d'où on obtient

$$f^{'}(\alpha_{n(x)}+y) = f^{'}(\alpha_{n(x)})+O(h_n)$$

alors (2.9) devient

$$\begin{aligned}
\sum_{-\infty}^{+\infty}\int_0^{h_n}(f'(\alpha_{n(x)}+y))^2y^2dy &= \sum_{-\infty}^{+\infty}\int_0^{h_n}\left[f'(\alpha_{n(x)})^2+O(h_n)\right]y^2dy \\
&= \sum_{-\infty}^{+\infty}\left(\frac{1}{3}h_n^3f'(\alpha_{n(x)})^2+\frac{1}{3}h_n^3\,O(h_n)\right) \\
&= \frac{1}{3}h_n^3\int_{-\infty}^{+\infty}f'(x)^2dx+h_nO(h_n^3) \\
&= \frac{1}{3}h_n^3\int_{-\infty}^{+\infty}f'(x)^2dx+o(h_n^3) \\
&= \frac{1}{3}h_n^3\int_{-\infty}^{+\infty}f'(x)^2dx+O(h_n^3) \qquad (2.10)
\end{aligned}$$

de la même manière, on montre que :

$$h_n\int_{\mathbb{R}}f'(x)^2(x-\alpha_{n(x)})dx+O(\frac{1}{n})=\left(-\frac{1}{2}\int_{-\infty}^{+\infty}f'(x)^2dx\right)+O(h_n^3)$$

par conséquent :

$$MISE(x)=\frac{1}{nh_n}+\frac{1}{12}h_n^2\int_{-\infty}^{+\infty}f'(x)^2dx+O(\frac{1}{n}+h_n^3) \qquad (2.11)$$

L'interêt du calcul du MISE est de trouver sa vitesse de convergence vers zéro, dans le cas où h_n est choisi de façon optimale. En effet, déterminer la fenêtre optimale h_n^* revient à calculer

$$argmin\left(\tfrac{1}{nh_n}+\tfrac{1}{12}h_n^2\int_{-\infty}^{+\infty}f'(x)^2dx\right)$$

c-à-d :

$$\begin{aligned}
\frac{\partial}{\partial h_n}MISE &= 0 \\
\frac{-n}{n^2h_n^2}+\frac{1}{6}h_n\int_{-\infty}^{+\infty}f'(x)^2dx &= 0
\end{aligned}$$

$$\begin{aligned}
\frac{-1}{nh_n^2}+\frac{1}{6}h_n\int_{-\infty}^{+\infty}f^{'}(x)^2dx &= 0\\
\frac{nh_n^3\int_{-\infty}^{+\infty}f^{'}(x)^2dx-6}{6nh_n^3}dx &= 0\\
nh_n^3\int_{-\infty}^{+\infty}f^{'}(x)^2dx-6 &= 0\\
h_n^3=\frac{6}{n\int_{-\infty}^{+\infty}f^{'}(x)^2dx}&
\end{aligned}$$

Le choix optimale de h_n est donc

$$h^*=\left(\frac{6}{\int_{-\infty}^{+\infty}f^{'}(x)^2dx}\right)^{1/3}n^{-1/3}$$

et $MISE$ asymptotique avec cet h_n optimale est égal a :

$$\begin{aligned}
MISE(h_n^*) &= n^{-1}(h_n^*)^{-1}+\frac{1}{12}(h_n^*)^2\int_{-\infty}^{+\infty}f^{'}(x)^2dx\\
&= n^{-1}\left(\frac{6}{\int_{-\infty}^{+\infty}f^{'}(x)^2dx}\right)^{-1/3}n^{1/3}+\frac{1}{12}\left(\frac{6}{\int_{-\infty}^{+\infty}f^{'}(x)^2dx}\right)^{2/3}n^{-2/3}\int_{-\infty}^{+\infty}f^{'}(x)^2dx\\
&= \left(\frac{6}{\int_{-\infty}^{+\infty}f^{'}(x)^2dx}\right)^{-1/3}n^{-2/3}+\frac{6^{2/3}}{12}\left(\int_{-\infty}^{+\infty}f^{'}(x)^2dx\right)^{1/3}n^{-2/3}\\
&= \left[\left(\frac{6}{\int_{-\infty}^{+\infty}f^{'}(x)^2dx}\right)^{-1/3}+\frac{6^{2/3}}{12}\left(\int_{-\infty}^{+\infty}f^{'}(x)^2dx\right)^{1/3}\right]n^{-2/3}\\
&=: C(x)n^{-2/3}
\end{aligned}$$

le MISE asymptotique avec le h_n^* optimale, est equivalent à $C(x)n^{-2/3}=O(n^{-2/3})$, où $C(x)$ dépend de $f^{'}(x)$. On diras que la vitesse de convergence de MISE de l'estimateur de

densité par histogramme est de l'ordre de $n^{-2/3}$.

Remarque 2.1 L'estimateur par histogramme présente des avantages et des inconvénients, parmi lesquels :

1. **Avantages**
   - L'histogramme est l'estimateur de densité le plus communément utilisé à cause de sa simplicité, il ne nécessite pas de compétences particulières pour le manipuler, de plus la quantité d'information fournie par un histogramme est une représentation globale de la densité sous-jacente des données. En fait, un histogramme affiche le nombre de données (ou observations) d'un ensemble fini de données qui appartiennent à une classe donnée.
2. **Inconvénients**
   - le nombre de classes croit exponentiellement avec la dimentionnalité : avec m classes par variable, on obtient m^d classes pour d variables donc à n'utiliser qu'à des fins exploratoires en 1 ou 2 dimensions.
   - L'estimateur histogramme de la densité donné par l'expression (2.1) est une fonction étagée, et donc discontinue . En raison de cette discontinuité , l'histogramme ne peut pas être ajusté dans le cas où nous disposons d'une information à priori sur la régularité de la densité à estimer. Par exemple, si on pose que la densité à estimer doit être deux fois continûment différentiable alors l'histogramme qui est discontinu ne répond pas au problème

2.3 Estimateur simple de densité (histogramme mobile)

L'estimateur simple de la densité, appelé aussi méthode d'estimation par histogramme à fenêtre mobile, a été proposé par *Rosenbllat* (1956) [41]. L'estimation de la densité en un point x, par cette méthode, consiste à construire autour de chaque x une classe de longueur $2h$ centrée sur x : $[x-h;x+h]$, on fait ensuite varier x et on compte à chaque fois le nombre d'observations dans cette classe.
Partons du lien existant entre la densité de probabilité f et la fonction de répartition F :

$$F(x) = \int_{-\infty}^{x} f(u)du, \quad \forall x \in \mathbb{R}$$

On peut écrire

$$\begin{aligned} f(x) &= \lim_{h_n \to 0} \frac{\mathbb{P}(x - h_n \leq X_i \leq x + h_n)}{2h_n} \\ &= \lim_{h_n \to 0} \frac{F(x + h_n) - F(x - h_n)}{2h_n} \end{aligned}$$

En remplaçant F par la fonction de répartition empirique F_n, on obtient l'estimateur simple de f, noté $\widehat{f_n}$, défini par :

$$\begin{aligned} \widehat{f_n}(x) &= \lim_{h_n \to 0} \frac{F_n(x + h_n) - F_n(x - h_n)}{2h_n} \\ &= \frac{1}{2h_n} \frac{card\{x - h_n \leq X_i \leq x + h_n\}}{n} \\ &= \frac{1}{n2h_n} \sum_{i=1}^{n} \mathbb{I}_{\{x - h_n \leq Xi \leq x + h_n\}} \\ &= \frac{1}{n2h_n} \sum_{i=1}^{n} \mathbb{I}_{[-1,1]} \left(\frac{x - X_i}{h_n} \right) \end{aligned}$$

où *card* est le cardinal d'un ensemble. Cet estimateur peut aussi s'écrire

$$\widehat{f_n}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_0 \left(\frac{x - X_i}{h_n} \right) \tag{2.12}$$

où K_0 est une fonction de poids définie par :

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

cette fonction de poids n'est rien d'autre que la densité de probabilité uniforme sur $[-1, 1]$

Remarque 2.2 L'inconvénient substantiel de l'estimateur simple défini par l'expression (2.12) est comme l'estimation par histogramme de fournir un estimateur discontinu. Il est cependant discontinu uniquement aux points $(X_i - h_n, X_i + h_n)_{i \in 1,\ldots,n}$ contrairement à l'estimateur par histogramme qui est discontinu aux bornes de chaque classe. Cette discontinuité est une conséquence de la discontinuité de la fonction de poids K_0 (2.3).

2.4 Estimateur à noyau pour la densité

L'estimateur simple de la densité défini précédement par l'expression (2.12) peut être généralisé en remplaçant la fonction de poids (2.3) (qui est une densité de probabilité uniforme) par une fonction de poids plus générale, notée K, (qui est une densité de probabilité quelconque). Cette nouvelle fonction de poids est appelée noyau (Kernel en anglais). On obtient donc l'estimateur à noyau , appelé aussi estimateur de densité de *Parzen-Rosenblatt* qui a été introduit par *Rosenblatt*[41] et développé par *Parzen*[37].

Définition 2.1 ***(noyau)***

La fonction noyau K est une densité de probabilité définie de $\mathbb{R}$ dans $\mathbb{R}^+$ et qui peut-être symétrique par rapport à 0 :

$$K(-u) = K(u) \tag{2.13}$$

ce qui implique l'égalité suivante

$$\int_{\mathbb{R}} tK(t)dt = 0 \tag{2.14}$$

de plus, elle est de carré integrable

$$\int_{\mathbb{R}} [K(t)]^2 \, dt < +\infty \tag{2.15}$$

et on a aussi la variance de K finie

$$\int_{\mathbb{R}} t^2 K(t) \, dt < +\infty \tag{2.16}$$

Les noyaux les plus utilisés sont des noyaux symétriques, données dans le tableau suivant :

Remarque 2.3 les programmes matlab données en annexe, permettent de simuler chacun des noyaux précédents.

Noyau	$K(u)$	Domaine de définition
Rectangulaire	$K(u) = 1/2$	$[-1,1]$
triangulaire	$K(u) = 1 - \mid u \mid$	$[-1,1]$
d'Epanechnikov	$K(u) = (3/4)(1-u^2)$	$[-1,1]$
Tukey ou Biweight	$K(u) = (15/16)(1-u^2)^2$	$[-1,1]$
Gaussien	$K(u) = (1/\sqrt{2\pi})\exp\left(-u^2/2\right)$	$\mathbb{R}$

TABLE 2.1 – Exemple de noyau symétrique

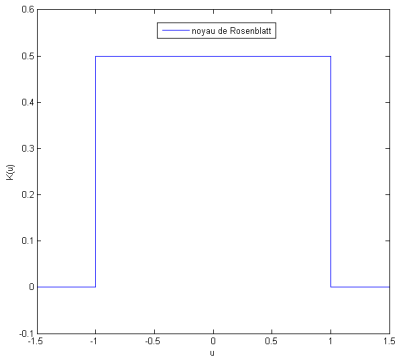

FIGURE 2.1 – Noyau Rectangulaire.

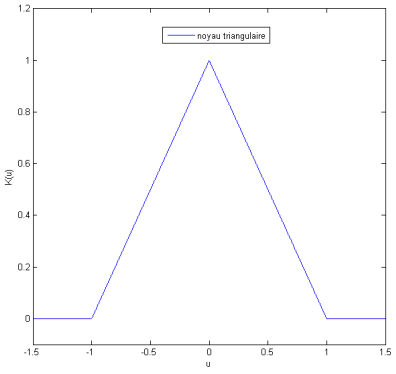

FIGURE 2.2 – Noyau triangulaire.

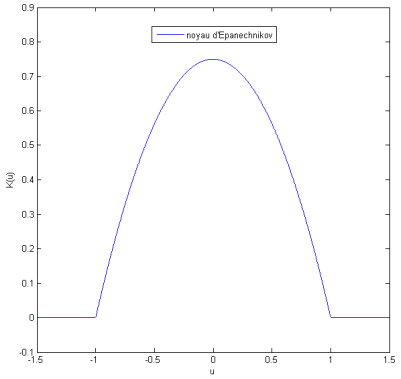

FIGURE 2.3 – Noyau d'Epanechnikov.

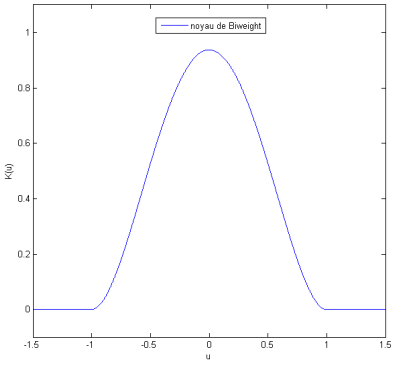

FIGURE 2.4 – Noyau de Biweight.

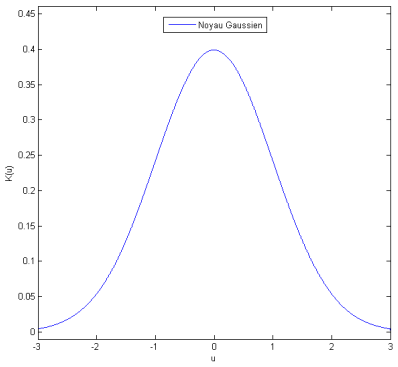

FIGURE 2.5 – Noyau Gaussien.

Définition 2.2 (Estimateur de densité de Parzen-Rosenblatt)

Considérons un échantillon de variables aléatoires $X_1, X_2, \ldots, X_n$, indépendant et identiquement distribué, de densité de probabilité f. L'estimateur de densité à noyau de *Parzen-Rosenblatt* , noté $\widehat{f}_n(x)$ est défini par

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \tag{2.17}$$

où K est un noyau, et h_n un réelle positif appelé fenêtre ou paramètre de lissage

2.4.1 Quelques propriétés de l'estimateur à noyau

Il est facile de voir que l'estimateur à noyau possède les propriétés suivantes :

1) Si K est une densité de probabilité, alors $\widehat{f}_n(x)$ est aussi une densité de probabilité. En effet

$$\begin{aligned}
\int_{-\infty}^{+\infty} \widehat{f}_n(x)\ dx &= \int_{-\infty}^{+\infty} \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) dx \\
&= \frac{1}{nh_n} \sum_{i=1}^{n} \int_{-\infty}^{+\infty} K\left(\frac{x - X_i}{h_n}\right) dx \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{+\infty} K_u(u)\ du = 1 \qquad (u = (x - X_i)/h_n)
\end{aligned}$$

2) $\widehat{f_n}(x)$ a les même propriétés de continuité et de différentiabilité que K :
- Si K est continue, $\widehat{f_n}(x)$ sera une fonction continue.
- Si K est différentiable, $\widehat{f_n}(x)$ sera une fonction différentiable.
- Si K peut prendre des valeurs négatives, alors $\widehat{f_n}(x)$ pourra aussi prendre des valeurs négatives

2.4.2 Calcul du Biais

Comme les variables aléatoires $X_1, X_2, \ldots, X_n$ sont i.i.d, nous avons successivement :

$$\begin{aligned}
\mathbb{E}\left(\widehat{f_n}(x)(X)\right) &= \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\left(\frac{x-X_i}{h_n}\right)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\{\frac{1}{h_n}K\left(\frac{x-X_i}{h_n}\right)\right\} \\
&= \mathbb{E}\left\{\frac{1}{h_n}K\left(\frac{x-X}{h_n}\right)\right\} \\
&= \frac{1}{h_n}\int_{\mathbb{R}}K\left(\frac{x-t}{h}\right)f(t)\,dt
\end{aligned}$$

Nous effectuons le changement de variables suivant : $-u = \dfrac{x-t}{h_n}$, d'où $du = \dfrac{1}{h_n}dt$

$$\mathbb{E}\left(\widehat{f_n}(x)\right) = \frac{h_n}{h_n}\int_{\mathbb{R}}K(-u)\,f(x+h_nu)\,du$$

comme $\int_{\mathbb{R}}K(u)\,du = 1$ et $K(-u) = K(u)$ le biais peut s'écrire :

$$\begin{aligned}
\mathbb{E}\left(\widehat{f_n}(x)\right) - f(x) &= \int_{\mathbb{R}}K(u)\;\; f(x+h_nu)\;\; du - f(x) \\
&= \int_{\mathbb{R}}K(u)\left\{f(x+h_nu)-f(x)\right\}\;\; du
\end{aligned}$$

Dans le but d'avoir une forme plus simple, qui ne dépend que du paramètre h_n, nous approximons la formule du biais en utilisant le développement de *Taylor* à l'ordre 2 de f au voisinage de x

$$f(x+uh_n) = f(x) + uh_nf'(x) + \frac{u^2h_n^2}{2}f''(x) + o(h_n^2)$$

De là, en utilisant l'hypothèse (2.14), le biais de $\widehat{f_n}(x)$ s'exprime ainsi par

$$\begin{aligned}\mathbb{E}\left(\widehat{f_n}(x)\right) - f(x) &= h_n f^{'}(x)\int_{\mathbb{R}} u\,K(u)\;du + \frac{h_n^2}{2}f^{''}(x)\int_{\mathbb{R}} u^2\,K(u)du + o(h_n^2)\\ &= \frac{h_n^2}{2}f^{''}(x)\int_{\mathbb{R}} u^2\,K(u)\;du + o(h_n^2) \qquad (2.18)\end{aligned}$$

2.4.3 Calcul de la variance

Comme les variables aléatoires $X_1, X_2, \ldots, X_n$ sont i.i.d, nous avons successivement

$$\begin{aligned}\mathbb{V}ar\left\{\widehat{f_n}(x)\right\} &= \mathbb{V}ar\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}K\left(\frac{x-X_i}{h_n}\right)\right\}\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}ar\left\{\frac{1}{h_n}K\left(\frac{x-X_i}{h_n}\right)\right\}\\ &= \frac{1}{n}\mathbb{V}ar\left\{\frac{1}{h_n}K\left(\frac{x-X}{h_n}\right)\right\}\\ &= \frac{1}{n}\mathbb{E}\left[\left\{\frac{1}{h_n}K\left(\frac{x-X}{h_n}\right)\right\}^2\right] - \frac{1}{n}\left[\mathbb{E}\left\{\frac{1}{h_n}K\left(\frac{x-X}{h_n}\right)\right\}\right]^2\\ &= \frac{1}{n}\int_{\mathbb{R}}\frac{1}{h_n^2}K^2\left(\frac{x-t}{h_n}\right)f(t)\;dt - \frac{1}{n}\left\{\int_{\mathbb{R}}\frac{1}{h_n}K\left(\frac{x-t}{h_n}\right)f(t)\;dt\right\}^2\end{aligned}$$

en faisons le changement de variable $-u = \dfrac{x-t}{h}$, d'où $du = \dfrac{1}{h_n}dt$, et comme $K(-u) = K(u)$

$$\begin{aligned}\mathbb{V}ar\left\{\widehat{f_n}(x)\right\} &= \frac{1}{nh_n^2}\int_{\mathbb{R}}K^2(-u)f(x+h_nu)\;h_n\;\;du - \frac{1}{n}\left\{\int_{\mathbb{R}}\frac{1}{h_n}K(-u)f(x+h_nu)\;h\;\;du\right\}^2\\ &= \frac{1}{nh_n}\int_{\mathbb{R}}[K(u)]^2f(x+h_nu)du - \frac{1}{n}\left\{\int_{\mathbb{R}}K(u)f(x+h_nu)du\right\}^2\end{aligned}$$

On voit bien que le deuxième terme de la variance est d'ordre $\dfrac{1}{n}$, ce que l'on note par

$$\tfrac{1}{n}\left\{\textstyle\int_{\mathbb{R}}K(u)f(x+h_nu)\;\;du\right\}^2 = O(\frac{1}{n})$$

et en utilisant le développement de *Taylor* de f au voisinage de x

$$f(x+uh_n) = f(x) + uh_nf^{'}(x) + o(h_n)$$

donc la variance est telle que

$$\mathbb{V}ar\left\{\widehat{f_n}(x)\right\} = \frac{1}{nh_n}\int_{\mathbb{R}}[K(u)]^2f(x)\;\;du + O(\frac{1}{n}) \qquad (2.19)$$

2.4.4 Erreur quadratique moyenne (MSE)

On voit que le premier terme de la variance ne tend vers zéro que si $nh_n \to \infty$ et le deuxième tend bien vers zéro quand $n \to \infty$. Par conséquent, pour que $\widehat{f_n}(x)$ converge vers $f(x)$ en moyenne quadratique les même conditions sont nécessaires que pour l'histogramme

$$n \to \infty \qquad h_n \to \infty \qquad nh_n \to \infty$$

Le comportement asymptotique de l'erreur quadratique moyenne de $\widehat{f_n}$ au point x, est donnée par la proposition suivante

Proposition 2.2 Sous les hypothèse précédentes nous obtenons

$$MSE(x) = \frac{h_n^4}{4}\left[f''(x)\right]^2\left[\int_{\mathbb{R}} u^2K(u)\ du\right]^2 + \frac{f(x)}{nh_n}\int_{\mathbb{R}}[K(u)]^2\ du + o(h_n^4) + O(\frac{1}{n}) \quad (2.20)$$

Démonstration 2.2 La preuve résulte des expression (2.19) et (2.18) ainsi que la décomposition en biais-variance

$$MSE(x) = \mathbb{V}ar(\widehat{f_n}) + Biais(\widehat{f_n})^2$$

on a

$$\begin{aligned} Biais(\widehat{f_n})^2 &= \left[\frac{h_n^2}{2}f''(x)\int_{\mathbb{R}} u^2K(u)\ du + o(h_n^2)\right]^2 \\ &= \frac{h_n^4}{4}[f''(x)]^2\left[\int_{\mathbb{R}} u^2K(u)\ du\right]^2 + [o(h_n^2)]^2 + \left(2\,\frac{h_n^2}{2}f''(x)\int_{\mathbb{R}} u^2K(u)\ du\right)\times\left(o(h^2)\right) \\ &= \frac{h_n^4}{4}[f''(x)]^2\left[\int_{\mathbb{R}} u^2K(u)\ du\right]^2 + o(h_n^4) + o(h_n^2)\ o(h_n^2) \\ &= \frac{h_n^4}{4}[f''(x)]^2\left[\int_{\mathbb{R}} u^2K(u)\ du\right]^2 + o(h_n^4) \end{aligned}$$

et

$$\mathbb{V}ar\left\{\widehat{f_n}(x)\right\} = \frac{1}{nh_n}\int_{\mathbb{R}}[K(u)]^2 f(x)\ du + O(\frac{1}{n}) \quad (2.21)$$

d'où l'approximation du critère MSE en un point x fixé est

$$MSE(x) = \frac{h_n^4}{4}\left[f''(x)\right]^2\left[\int_{\mathbb{R}} u^2K(u)\ du\right]^2 + \frac{f(x)}{nh_n}\int_{\mathbb{R}}[K(u)]^2\ du + o(h_n^4) + O(\frac{1}{n}) \quad (2.22)$$

2.4.5 Erreur quadratique moyenne intégrée (MISE)

L'erreur quadratique moyenne intégrée *MISE* est la moyenne théorique commune la plus utilisée pour évaluer l'erreur entre la fonction f et $\widehat{f_n}(x)$ sur tout la droite réelle $\mathbb{R}$

$$\begin{aligned} MISE(n,h,K,f) &= \int_{\mathbb{R}} MSE(x)\ dx \\ &= \int_{\mathbb{R}} \mathbb{V}ar(\widehat{f_n}(x))\ dx + \int_{\mathbb{R}} \mathbb{B}iais^2(\widehat{f_n}(x))\ dx \end{aligned}$$

En utilisant l'expression approximée du critère $MSE(x)$ on trouve successivement :

$$\begin{aligned} AMISE(n,h,K,f) &= \frac{1}{nh_n}\int_{\mathbb{R}}[K(u)]^2\ du \int_{\mathbb{R}} f(x)\ dx + \frac{h_n^4}{4}\left[\int_{\mathbb{R}} u^2K(u)\ du\right]^2 \int_{\mathbb{R}}\left[f''(x)\right]^2\ dx \\ &= \frac{1}{nh_n}\int_{\mathbb{R}}[K(u)]^2\ du + \frac{h_n^4}{4}\left[\int_{\mathbb{R}} u^2K(u)\ du\right]^2 \int_{\mathbb{R}}\left[f''(x)\right]^2\ dx \\ &= \frac{1}{nh_n}\int_{\mathbb{R}}[K(u)]^2\ du + \frac{h_n^4}{4}[V(K)]^2 \int_{\mathbb{R}}\left[f''(x)\right]^2\ dx \end{aligned} \tag{2.23}$$

avec $V(K) = \int_{\mathbb{R}} u^2K(u)\ du = \mathbb{V}ar(K)$

Remarque 2.4 L'estimateur à noyau présente des avantages et des inconvénients, parmi lesquels :

1. **Avantages**
   - aucune hypothèse n'est faite à l'avance quant à la distribution des données. La densité est estimée entièrement à partir des données.
   - Il peut être utilisé pour des densités multidimensionnelles.
2. **Inconvénients**
   - Il produit les effets aux bords pour des densités à supports compacts.

2.5 choix du noyau

Le problème du choix optimal de K, consiste à chercher un noyau optimal sous la contrainte de posivité, $K \geq 0$. On fait le rappel de l'expression asymptotique de l'erreur quadratique intégrée *AMISE*

$$AMISE(n,h,K,f) = \frac{1}{n\,h}\int_{\mathbb{R}}[K(t)]^2\,dt + \frac{1}{4}h^4\,[V(K)]^2\int_{\mathbb{R}}[f''(x)]^2\,dx$$

on remarque que la dépendance du *AMISE* par rapport au noyau K s'exprime par l'intervention de sa variance $V(K)$. Un noyau optimal K^* est donc un noyau qui minimise la fonctionnelle $V(K)$, soit

$$V(K^*) = \min_{K \in \mathcal{K}} V(K) \tag{2.24}$$

où $\mathcal{K}$ désigne l'ensemble des noyaux positifs d'ordre 1 satisfaisant aux conditions (2.15 2.16). La solution du problème est donnée par la proposition suivante :

Proposition 2.3 (Tsybakov (2004)[45])

une solution du problème de minimisation (2.24) est donnée par le noyau d'Epanechnikov

$$K^*(u) = \tfrac{3}{4}(1-u^2)$$

qui fournit la valeur minimale $V(K^*) = 3^{4/5}\,5^{-6/5}$

On peut considérer l'eficacité de chacun des noyaux symétrique présenté dans le tableau (2.1), en comparant avec le noyau d'epanechnikov. On défini l'efficacité (voir *Silverman*,[44]) par :

$$\begin{aligned} \mathit{eff}(K) &= \left\{ \frac{C(K_e)}{C(K)} \right\}^{5/4} \\ &= \frac{3}{5\sqrt{5}} \; \frac{1}{\sqrt{\int u^2 K(u)\; du} \; \int K(u)^2\, du} \end{aligned} \tag{2.25}$$

avec $C(K) = (V(K))^{2/5} \;\; \left\{ \int (K(u))^2 \;\; du \right\}^{4/5}$
la raison de la puissance $5/4$ dans (2.25) est que pour n grand l'erreur quadratique moyenne intégrée sera la même si on utilise n observations et le noyau K ou si on utise n $\mathit{eff}(K)$ observations et le noyau d'epanechnikov K_e ([44]).
le tableau (2.2) présente les valeur d'efficacité de quelques noyaux continues symétriques.

Noyau	Efficacité
d'Epanechnikov	≈ 1.000
Biweight	≈ 0.9939
triangulaire	≈ 0.9859
Gaussien	≈ 0.9512
Rectangulaire	≈ 0.9295

TABLE 2.2 – Efficacité des noyaux continus symétriques

On remarque que les valeurs d'efficacité obtenus sont très proches de 1 et qu'il ya très peu de différence entre les différents noyaux sur la base de l'erreur quadratique moyenne intégrée.

2.6 Choix théorique optimal du paramètre de lissage

Le paramètre de lissage h_n est un réel positif dont le choix est dominant par rapport au choix du noyau K. On voit que plus le paramètre h_n est faible plus le biais diminue mais plus la variance augmente et, de façon inverse l'élargissement de h_n augmente le biais et diminue la variance. Il existe un optimum qui est la valeur de h_n qui minimise l'erreur quadratique moyenne intégrée *MISE*

$$h_n^{opt} = \underset{h}{argmin}\, MISE$$

Comme pour l'histogramme pour une taille d'échantillon n donnée et un noyau K fixé, en dérivant par rapport à h_n, on obtient

$$\frac{\partial}{\partial h_n} AMISE(h_n) = 0$$

ce qui est équivalent à

$$h_n^3 [V(K)]^2 \int_{\mathbb{R}} [f''(x)]^2 \, dx - \frac{1}{n h_n^2} \int_{\mathbb{R}} [K(u)]^2 \, du = 0$$

Ainsi, en obtient successivement

$$\begin{aligned}
h_n^3 [V(K)]^2 \int_{\mathbb{R}} [f''(x)]^2 \, dx &= \frac{1}{n h_n^2} \int_{\mathbb{R}} [K(u)]^2 \, du \\
n h_n^5 [V(K)]^2 \int_{\mathbb{R}} [f''(x)]^2 \, dx &= \int_{\mathbb{R}} [K(u)]^2 \, du \\
h_n^5 &= \frac{\int_{\mathbb{R}} [K(u)]^2 \, du}{n [V(K)]^2 \int_{\mathbb{R}} [f''(x)]^2 \, dx} \\
h_n^{opt} &= \frac{1}{n^{1/5}} \left\{ \frac{\int_{\mathbb{R}} [K(u)]^2 \, du}{[V(K)]^2 \int_{\mathbb{R}} [f''(x)]^2 \, dx} \right\}^{1/5} \\
h_n^{opt} &= \left\{ \frac{\int_{\mathbb{R}} [K(u)]^2 \, du}{[V(K)]^2 \int_{\mathbb{R}} [f''(x)]^2 \, dx} \right\}^{1/5} n^{-1/5}
\end{aligned}$$

D'où

$$h_n^{opt} = \{V(K)\}^{-2/5} \left\{ \int_{\mathbb{R}} [K(u)]^2 \, du \right\}^{1/5} \left\{ \int_{\mathbb{R}} [f''(x)]^2 \, dx \right\}^{-1/5} n^{-1/5} \tag{2.26}$$

finalement, à partir de (2.23), nous obtenons

$$\begin{aligned}
AMISE(h_n^{opt}) &= n^{-1} \; n^{1/5} \Big(\int_{\mathbb{R}} [K(u)]^2 \, du \Big)^{-1/5} (V(K))^{2/5} \Big(\int_{\mathbb{R}} [f''(x)]^2 \, dx \Big)^{1/5} \; \Big(\int_{\mathbb{R}} [K(u)]^2 \, du \Big) + \frac{1}{4} \\
&\quad n^{-4/5} \Big(\int_{\mathbb{R}} [K(u)]^2 \, du \Big)^{4/5} (V(K))^{8/5} \Big(\int_{\mathbb{R}} [f''(x)]^2 \, dx \Big)^{-4/5} (V(K))^2 \Big(\int_{\mathbb{R}} [f''(x)]^2 \, dx \Big) \\
&= n^{-4/5} \Big(\int_{\mathbb{R}} [K(u)]^2 \, du \Big)^{4/5} (V(K))^{2/5} \Big(\int_{\mathbb{R}} [f''(x)]^2 \, dx \Big)^{1/5} \\
&\quad + \frac{1}{4} \, n^{-4/5} \Big(\int_{\mathbb{R}} [K(u)]^2 \, du \Big)^{4/5} (V(K))^{2/5} \Big(\int_{\mathbb{R}} [f''(x)]^2 \, dx \Big)^{1/5} \\
&= \frac{5}{4} \; n^{-4/5} \Big(\int_{\mathbb{R}} [K(u)]^2 \, du \Big)^{4/5} (V(K))^{2/5} \Big(\int_{\mathbb{R}} [f''(x)]^2 \, dx \Big)^{1/5} \\
&= \frac{5}{4} \; I(K) \; g(f'') \; n^{-4/5}
\end{aligned}$$

avec

$$I(K) = \Big(\int_{\mathbb{R}} [K(u)]^2 \, du \Big)^{4/5} (V(K))^{2/5}$$

et

$$g(f'') = \Big(\int_{\mathbb{R}} [f''(x)]^2 \, dx \Big)^{1/5}$$

Remarque 2.5 La convergence de l'estimateur à noyau $\widehat{f_n}$ est plus rapide que pour l'histogramme, étant d'ordre $n^{-4/5}$ au lieu de $n^{-2/3}$, mais ce résultat basé sur un choix optimal théorique n'est pas utilisable en pratique car il dépend de la quantité inconnue f'', c'est pourquoi on considère des méthode pratique pour le choix du h_n dans le paragraphe suivant.

2.7 Choix pratique du paramètre de lissage(h)

2.7.1 Méthode *Plug-in*

L'ideé de base de la procédure de *Plug-in* pour le choix du paramètre h, est d'estimer dans l'expression de h_{opt} théorique (2.26), la quantité inconnue : $\int_{\mathbb{R}}[f''(x)]^2\,dx$. En effet on suppose que $f(x)$ appartient à une famille de distributions normales $N(\mu,\sigma^2)$, de moyenne μ et variance σ^2 inconnues. Sous cette hypothèse

$$f(x)=\frac{1}{\sigma}\,\phi\left(\frac{x-\mu}{\sigma}\right),\qquad avec\qquad \phi(x)=\frac{1}{\sqrt{2\pi}}\,e^{\frac{x^2}{2}}$$

la densité de probabilité normale centré réduite et

$$f'(x)=\frac{1}{\sigma^2}\,\phi'\left(\frac{x-\mu}{\sigma}\right)\quad\Rightarrow\quad f''(x)=\frac{1}{\sigma^3}\,\phi''\left(\frac{x-\mu}{\sigma}\right)$$

la quantité inconnue $\int_{\mathbb{R}}[f''(x)]^2\,dx$, s'écrit alors

$$\int_{\mathbb{R}}[f''(x)]^2\,dx=\frac{1}{\sigma^6}\int_{\mathbb{R}}\left\{\phi''\left(\frac{x-\mu}{\sigma}\right)\right\}^2 dx$$

faisons le changement de variable $v=\frac{x-\mu}{\sigma}$, d'où $dv=\frac{1}{\sigma}\,dx$

$$\int_{\mathbb{R}}[f''(x)]^2\,dx=\frac{1}{\sigma^5}\int_{\mathbb{R}}\left\{\phi''(v)\right\}^2 dv$$

mais

$$\phi(v)=\frac{1}{\sqrt{2\pi}}\,e^{\frac{-v^2}{2}}\qquad\Rightarrow\qquad \phi'(v)=\frac{-v}{\sqrt{2\pi}}\,e^{\frac{-v^2}{2}}\qquad\Rightarrow\qquad \phi''(v)=\frac{1}{\sqrt{2\pi}}\,(v^2-1)\,e^{\frac{-v^2}{2}}$$

$$\begin{aligned}\int_{\mathbb{R}}[f''(x)]^2\,dx &= \frac{1}{\sigma^5}\,\frac{1}{2\pi}\int_{\mathbb{R}}(v^2-1)^2\,e^{-v^2}\,dv\\ &= \frac{1}{\sigma^5}\,\frac{1}{2\pi}\left\{\int_{\mathbb{R}}v^4\,e^{-v^2}\,dv-2\int_{\mathbb{R}}v^2\,e^{-v^2}\,dv+\int_{\mathbb{R}}e^{-v^2}\,dv\right\}\\ &= \frac{1}{\sigma^5}\,\frac{1}{2\pi}\left\{-\frac{1}{2}\int_{\mathbb{R}}v^2\,e^{-v^2}\,dv+\int_{\mathbb{R}}e^{-v^2}\,dv\right\}\end{aligned}$$

posons $\quad u=\sqrt{2}\,v \quad \Rightarrow \quad du=\sqrt{2}\,dv$

$$\begin{aligned}\int_{\mathbb{R}}[f''(x)]^2\,dx &= \frac{1}{\sigma^5}\,\frac{1}{2\pi}\left\{-\frac{1}{2}\int_{\mathbb{R}}\frac{u^2}{2}\,e^{\frac{-u^2}{2}}\,\frac{du}{\sqrt{2}}+\frac{1}{\sqrt{2}}\int_{\mathbb{R}}e^{\frac{-u^2}{2}}\,du\right\}\\ &= \frac{1}{\sigma^5}\,\frac{1}{2\pi}\left\{-\frac{1}{4}\;\sqrt{\pi}+\sqrt{\pi}\;1\right\}\\ &= \frac{1}{\sigma^5}\;\;\frac{1}{2\pi}\;\;\frac{3}{4}\sqrt{\pi}\\ &= \frac{1}{\sigma^5}\;\;\frac{3}{8\sqrt{\pi}}\\ &\backsimeq\; 0.212\,\sigma^5\end{aligned}$$

Il reste alors à remplacer l'ecart type σ par la valeur estimée. En choisissant l'ecart type empirique comme valeur optimale

$$\widehat{\sigma}=\sqrt{\tfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

tel que $\quad \overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

Donc, on remplace le résultat obtenue dans la formule de h_n^{opt} on obtient :

$$h_n^{opt}=\{V(K)\}^{-2/5}\;\left\{\int_{\mathbb{R}}[K(u)]^2\,du\right\}^{1/5}\;\left\{\frac{3}{8\sqrt{\pi}\,\widehat{\sigma}^{\,-5}}\right\}^{-1/5}\;n^{-1/5}$$

si on utilise le noyau gaussien i,e $\qquad K\sim N(0,1)$

$$\begin{aligned}\int_{\mathbb{R}}[K(u)]^2\,du &= \int_{\mathbb{R}}\left(\frac{1}{\sqrt{2\pi}}\;\;e^{u^2/2}]^2\right)du\\ &= \int_{\mathbb{R}}\frac{1}{2\pi}\;\;e^{-u^2}\,du\\ &= \frac{\sqrt{\pi}}{2\pi}\\ &= \frac{1}{2\sqrt{\pi}}\end{aligned}$$

ce qui implique

$$\begin{aligned} h_n^{opt} &= 1 \left\{(4\pi)^{-1/2}\right\}^{1/5} \left\{\frac{3}{8}\ \pi^{-1/2}\widehat{\sigma}^{-5}\right\}^{-1/5} n^{-1/5} \\ &= 4^{-1/10}\ \pi{-}1/10\ \left(\frac{3}{8}\right)^{-1/5}\ \pi^{1/10}\ \widehat{\sigma}\ n^{-1/5} \\ &= 2^{-2/5}\ 3^{-1/5}\ 2^{4/5}\ \widehat{\sigma}\ n^{-1/5} \\ &= 2^{2/5}\ 3^{-1/5}\ \widehat{\sigma}\ n^{-1/5} \\ &= 4^{1/5}\ 3^{-1/5}\ \widehat{\sigma}\ n^{-1/5} \\ &= \left(\frac{4}{3}\right)^{1/5}\ \widehat{\sigma}\ n^{-1/5} \\ &= 1.06\,\widehat{\sigma}\ n^{-1/5} \end{aligned}$$

Alors le paramètre de lissage donneé par la méthodes de plug-in pour un noyau gaussien est défini par

$$h_n^{opt} = 1.06\,\widehat{\sigma}\ n^{-1/5}$$

2.7.2 Méthode de validation croisée par moindre carrés

Cette méthode appellée aussi méthode de validation croisée non biaisée (Unbiaised Cross-Validation "UCV") proposée par *Rudemo* (1982) [42] et *Bowman* (1984) [4]. Le principe de cette méthode est la minimisation d'un estimateur de l'erreur quadratique moyenne intégrée *MISE* par rapport à h_n. En effet, le *MISE* dépend de la fonction inconnue f. On va remplacer $MISE(h_n)$ par une fonction de h_n, mesurable par rapport à l'échantillon et dont la valeur , pour chaque $h_n > 0$, est un estimateur sans biais de $MISE(h_n)$, pour cela, on a

$$\begin{aligned} MISE(h_n) &= \mathbb{E}\int_{\mathbb{R}} \left\{\widehat{f_n}(x) - f(x)\right\}^2 dx \\ &= \mathbb{E}\int_{\mathbb{R}} [\widehat{f_n}(x)]^2\,dx - 2\mathbb{E}\int_{\mathbb{R}} \widehat{f_n}(x)\,f(x)\,dx + \mathbb{E}\int_{\mathbb{R}} [f(x)]^2\,dx \end{aligned}$$

Le dernier terme ne dépend pas de h_n, pour minimiser $MISE(h_n)$ il suffit de minimiser l'expression

$$J(h_n) = \mathbb{E}\int_{\mathbb{R}} [\widehat{f_n}(x)]^2\,dx - 2\ \mathbb{E}\int_{\mathbb{R}} \widehat{f_n}(x)\,f(x)\,dx$$

Puisque J dépend de f inconnue donc on suppose de l'estimer et de, choisir h_n qui minimise son estimateur.

Le premier terme admet comme estimateur trivial l'estimateur $\int_{\mathbb{R}}[\widehat{f_n}(x)]^2\,dx$ (en effet d'après la propriété des estimateurs sans biais : $\mathbb{E}(\widehat{\beta}) = \beta$)

Pour le second terme on peut monter qu'il admet comme estimateur sans biais la quantité :

$$\widehat{G} = \frac{1}{n}\sum_{i=1}^{n} \widehat{f}_{n,-\,i}(X_i)$$

avec

$$\widehat{f}_{n,-\,i}(X_i) = \frac{1}{(n-1)\,h_n}\sum_{j=1,j\neq i}^{n} K\left(\frac{X_i - X_j}{h_n}\right)$$

en effet comme les X_i sont i.i.d, d'une part on a

$$\begin{aligned}
\mathbb{E}\left\{\widehat{G}\right\} &= \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n} \widehat{f}_{n,-\,i}(X_i)\right\} \\
&= \mathbb{E}\left\{\widehat{f}_{n,-\,1}(X_1)\right\} \\
&= \mathbb{E}\left\{\frac{1}{(n-1)\,h_n}\sum_{j=1,j\neq 1}^{n} K\left(\frac{X_1 - X_j}{h_n}\right)\right\} \\
&= \frac{1}{(n-1)}\sum_{j=1,j\neq 1}^{n} \mathbb{E}\left\{\frac{1}{h_n}K\left(\frac{X_1 - X_j}{h_n}\right)\right\} \\
&= \mathbb{E}\left\{\frac{1}{h_n}K\left(\frac{X_1 - X}{h_n}\right)\right\} \\
&= \frac{1}{h_n}\int\int K\left(\frac{x-z}{h_n}\right) f(z)\,f(x)\,dz\,dx \\
&= \frac{1}{h_n}\int_{\mathbb{R}} f(x)\int_{\mathbb{R}} K\left(\frac{x-z}{h_n}\right) \quad f(z)\,dz\,dx
\end{aligned}$$

d'autre part on a

$$\begin{aligned}
\mathbb{E}\left\{\int_{\mathbb{R}} \widehat{f_n}(x)\,f(x)\,dx\right\} &= \mathbb{E}\left\{\int_{\mathbb{R}} \frac{1}{n\,h_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \quad f(x)\,dx\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left\{\int_{\mathbb{R}} \frac{1}{h_n}K\left(\frac{x-X_i}{h_n}\right) f(x)\,dx\right\} \\
&= \frac{1}{h_n}\mathbb{E}\left\{\int_{\mathbb{R}} K\left(\frac{x-X}{h_n}\right) \quad f(x)\,dx\right\} \\
&= \frac{1}{h_n}\int_{\mathbb{R}} f(x)\int_{\mathbb{R}} K\left(\frac{x-z}{h_n}\right) \quad f(z)\,dz\,dx
\end{aligned}$$

Ce qui implique que $\mathbb{E}(\widehat{G}) = \mathbb{E}\int_{\mathbb{R}} \widehat{f}_n(x)\, f(x)\, dx$.
en définitif, l'estimateur sans biais de $J(h_n)$ est donnée par

$$UCV(h_n) = \int_{\mathbb{R}} [\widehat{f}_n(x)]^2\, dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{n,-i}(X_i) \tag{2.27}$$

et le paramètre de lissage du type " Unbiaised Cross-Validation " est la valeur de h_n qui minimise la quantité $UCV(h_n)$, c'est-à-dire

$$h_{UCV} = \underset{h>0}{argmin}\, UCV(h_n)$$

2.7.3 Méthode du maximum de vraisemblance par validation croisée

La méthode du maximum de vraisemblance avec validation croisée (Maximum Likelihood Cross-Validation "MLCV") est une méthode proposé par Habbema, Hermans and Van den Broek (1974) [27] et Duin (1976) [18]. Ils ont proposé de choisir h_n de sorte que la pseudo-vraisemblance $\prod_{i=1}^{n} \widehat{f}_n(X_i)$ soit maximisée. Cependant cela a un maximum trivial à $h_n = 0$, donc le principe de validation croisée est introduit par le remplacement de $\widehat{f}_n(x)$ par $\widehat{f}_{n,-i}(x)$, où

$$\widehat{f}_{n,-i}(X_i) = \frac{1}{h_n\,(n-1)} \sum_{j\neq i} K\left(\frac{X_i - X_j}{h_n}\right)$$

Le paramètre de lissage donné par la méthode du maximum de vraisemblance avec validation croisée est le paramètre qui maximise l'expression :

$$MLCV(h_n) = \left(\frac{1}{n}\sum_{i=1}^{n} \log\left[\sum_{j\leq i} K\left(\frac{X_i - X_j}{h_n}\right)\right] - \log\left[(n-1)/h_n\right]\right)$$

C'est-à-dire

$$h_{MLCV} = \underset{h>0}{argmax}\, MLCV(h_n)$$

2.8 Estimateur de la fonction de régression de Nadaraya-Watson :

On a montré dans (1.3.3) que la solution du problème de la regression non paramétrique

$$Y_i = m(X_i) + \varepsilon_i$$

est donné par l'espérance conditionnelle $r(x) := \mathbb{E}(Y|X = x)$.

Supposons que l'on dispose d'un n-échantillon $(X_1,Y_1),\ldots,(X_n,Y_n)$ de v.a de même loi que (X,Y). On se propose de construire un estimateur $\widehat{r}_n(x,(X_1,Y_1),\ldots,(X_n,Y_n))$ de la fonction r. Il existe plusieurs types d'estimateur à noyau pour la régression dont le plus célèbre est celui de *Nadaraya-Watson* (*Nadaraya*(1964) et *Watson* (1964).

Supposons que (X,Y) ait une densité $f : (x,y) \longrightarrow f_{X,Y}(x,y)$ sur $\mathbb{R}^2$ et que $f_X : x \longrightarrow f_X(x) = \int f_{X,Y}(x,y)dy > 0$ (densité de X). Alors, on peut écrire

$$\forall x \in \mathbb{R}, \qquad r(x) = \mathbb{E}(Y|X = x) = \frac{\int y\, f_{X,Y}(x,y)dy}{f_X(x)} \tag{2.28}$$

comme les densités $f_{X,Y}$ et f_X sont inconnus, On peut les estimer en suivant les même étapes que pour l'estimateur à noyau. On considère donc

$$\forall (x,y) \in \mathbb{R}^2, \qquad \widehat{f}_{X,Y}(x,y) = \frac{1}{n\,h_n^2} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) K\left(\frac{y-Y_i}{h_n}\right) \tag{2.29}$$

$$\widehat{f}_X(x) = \frac{1}{n\,h_n} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)$$

et on obtient l'estimateur de *Nadaraya-Watson* (1964)(NW)

$$\widehat{r}_{NW;n}(x) := \frac{\frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)} := \frac{\widehat{r}_{1;n}(x)}{\widehat{f}_n(x)} \tag{2.30}$$

en effet si on remplace la densité conjointe $f_{X,Y}$ par son estimateur défini dans (2.29) on a :

$$\begin{aligned} \int y\,\widehat{f}_{X,Y}(x,y)dy &= \int \frac{1}{n\,h_n^2} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) y\, K\left(\frac{y-Y_i}{h_n}\right) dy \\ &= \frac{1}{n\,h_n} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \int \frac{1}{h_n}\, y\, K\left(\frac{y-Y_i}{h_n}\right) dy \end{aligned}$$

si on suppose de plus que le noyau K est symétrique, et en faisant le changement de variables $u = \frac{y-Y_i}{h_n}$, on montre que

$$\begin{aligned}\int \frac{1}{h_n} y\, K\left(\frac{y-Y_i}{h_n}\right) dy &= \int (u\,h_n + Y_i)\, K(u) du \\ &= h_n \int u\, K(u)\, du + \int Y_i\, K(u)\, du \\ &= Y_i\end{aligned}$$

d'où l'expression (2.30) :

$$\widehat{r}_{NW;n}(x) = \frac{\frac{1}{n\,h_n} \sum_{i=1}^{n} Y_i\, K\left(\frac{x-X_i}{h_n}\right)}{\frac{1}{n\,h_n} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)}$$

L'estimateur de *Nadaraya-Watson* est une moyenne pondérée des observations Y_i. On a

$$\forall x \in \mathbb{R}, \qquad \widehat{r}_{NW,n}(x) = \sum_{i=1}^{n} w_{n,i}(x)\, Y_i$$

où les poids $w_{n,i}(x)$ vérifient

$$w_{n,i}(x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} \mathbb{I}_{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) \neq 0}$$

ces poids ne dépendent pas des Y_i. En particulier $\widehat{r}_{NW;\,n}$ est un estimateur linéaire de la régression non paramétrique de Y_i sur les X_i

2.8.1 Propriété asymptotique de l'estimateur de Nadaraya-Watson

Comme pour l'estimateur à noyau de la densité l'estimateur à noyau de la régression est dépendant du choix de la fenêtre h_n et du noyau K. Dans ce qui va suivre nous allons déterminer les conditions sur la fenêtre et le noyau, nécessaire à la convergence de l'estimateur $\widehat{r}_{NW;\,n}$. Nous utiliserons pour cela la décopmposition biais-variance suivante :

$$\mathbb{E}\left[\left(\widehat{r}_{NW;\,n}(x) - r(x)\right)^2\right] = Var\left[\widehat{r}_{NW;\,n}(x)\right] + \left[\mathbb{E}\left(\widehat{r}_{NW;\,n}(x)\right) - r(x)\right]^2 \tag{2.31}$$

2.8.1.1 Variance

Pour calculer la variance de l'estimateur $\widehat{r}_{NW;\,n}(x)$ ainsi que son expression asymptotique, nous supposerons que le noyau K vérifie les hypothèses suivantes :

(K1) K est bornée

(K2) $\lim_{n\to+\infty} |t|\,K(t) = 0$

(K3) $K(.) \in L_1(\mathbb{R})$

(K4) $\int K(t)dt = 1$

et lorsque les deux expressions sont bien définies

$$\bar{\sigma}^2(x) := Var(Y/X = x) = \frac{1}{f(x)} \int y^2 f_{X,Y}(x,y)dy \;\; - \;\; [r(x)]^2$$

et $\kappa = \int K^2(t)dt$. Alors la variance de l'estimateur de $\widehat{r}_{NW;\,n}(x)$ est donnée par la proposition suivante (pour plus de détail voir Guessoum (2009), [23])

Proposition 2.4 On suppose que $\mathbb{E}(Y^2) < +\infty$. A chaque point de continuité des fonctions $r(x), f(x)$ et $\bar{\sigma}^2(x)$ tels que $f(x) > 0$, on a

$$Var\,[\widehat{r}_{NW;\,n}(x)] = \frac{1}{nh_n} \left(\frac{\bar{\sigma}^2(x)}{f(x)} \kappa \right) (1 + o(1)) \tag{2.32}$$

où le terme $o(1)$ tend vers 0 lorsque h_n tend vers 0.

2.8.1.2 Biais

Le calcul du biais est basé principalement sur des développements de Taylors, ce qui nous a amené à poser certaines conditions de régularités sur les fonctions $r(.)$ et $f(.)$ qui déterminerons l'ordre du biais asymptotique en fonction du paramètre de lissage h_n. Pour faciliter les calculs et du fait que l'estimateur $\widehat{r}_{NW;\,n}(.)$ est lui même sous forme d'un rapport aléatoire, nous allons introduire un terme sous forme de rapport, défini par $\tilde{\mathbb{E}}\left(\widehat{r}_{NW;\,n}(x)\right) := \frac{\mathbb{E}(\widehat{r}_{1;\,n}(x))}{\mathbb{E}\left(\widehat{f}_n\,(x)\right)}$. Le biais est donné par la proposition suivante (voir Guessoum 2009, p 22) :

Proposition 2.5 On suppose que $r(x)$ et $f(x)$sont de classe $\mathcal{C}^2$sur $\mathbb{R}$ et que le noyau K est d'ordre 2 c-à-d

$$\int K(t)dt = 1 \;\; , \int tK(t)dt = 0 \;\; , \int t^2 K(t)dt < +\infty$$

alors lorsque h tend vers 0 et nh tend vers $+\infty$ on a

$$\mathbb{E}\left(\widehat{r}_{NW;\,n}(x)\right) - r(x) = \frac{h^2}{2}\left\{\left(r''(x) + 2r'(x)\frac{f'(x)}{f(x)}\right)\int t^2 K(t)dt\right\} + o(1)$$

Chapitre 3

Estimation non paramétrique sur des données censurées

Sommaire

3.1 Introduction

Dans cette partie nous introduisons la notion de censure dans les données, nous faisons quelques rappels basiques pour ce type de modèle, et nous donnons quelques résultats principaux concernant le comportement asymptotique de l'estimateurs de la fonction de densité et de la fonction de régression. Nous nous intéressons en particulier, à un estimateur à noyau non symétrique de la fonction de régression pour lequel nous établissons la vitesse de convergence.

Dans cette section, on va établir quelques résultats sur les estimateurs de la densité et de la fonction de régression pour un modèle censuré, mais avant nous rappelons ce qu'est un modèle de survie, un modèle de censure.

3.2 Modèle de survie

Un modèle de survie s'appuie sur des durées de vie. Le terme de durée de vie est utilisé pour indiquer le temps qui passe jusqu'à la survenue d'un évènement particulier qui n'est pas forcément la mort. Elle s'applique cependant à d'autres sortes d'évènements, par exemple, l'apparition d'une maladie, la guérison d'une maladie, la panne d'une machine, etc ...
Les modèles de survie sont souvent caractérisés par la présence de censure : une censure se produit lorsque l'évènement étudié n'intervient pas pendant la période d'observation pour une raison ou une autre. Cette censure est dite "censure à droite" et est la plus courante mais n'est pas la seule censure que l'on peut rencontrer sur des données de survie.

3.3 Différents types de censure

3.3.1 Censure de type I : (fixé)

Soit C un nombre positif fixé. Au lieu d'observer $X_1, X_2, \ldots, X_n$ qui nous intéressent, on observe X_i que lorsque $X_i \leq C$, sinon on sait seulement que $X_i > C$.
L'observation est alors $Y_i = \min(X_i, C) = X_i \wedge C$. C'est le cas lorsqu'on décide à l'avance que le nombre C est la durée de l'étude.

3.3.2 Censure de type II : (attente)

On décide d'observer les durées de survie de n patients jusqu'à ce que r d'entre eux soient décédés et d'arrêter l'étude à ce moment là. Si l'on ordonne les durées de survie $X_1, X_2, \ldots, X_n$, soit $X_{(1)}$ la plus petite, $X_{(i)}$ la ième et ainsi de suite

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(i)} \leq \cdots \leq X_{(n)}$$

On dit que les $X_{(i)}$ sont les statistiques d'ordre des X_i. La date de censure est alors $X_{(r)}$ et on observe : $Y_i = X_i \wedge X_{(r)}$

$$\begin{cases} Y_1 = X_{(1)} \\ Y_2 = X_{(2)} \\ \cdots \\ Y_r = X_{(r)} \\ Y_{r+1} = X_{(r)} \cdots \\ Y_n = X_{(r)} \end{cases}$$

Ce cas est fréquemment utilisé en fiabilité lorsqu'on observe jusqu'à la première panne.

3.3.3 Censure de type III : (aléatoire)

A chaque individu i, est associé un couple de v.a (X_i, C_i) positives où X_i est son temps de survie et C_i son temps de censure, Tel que seule la plus petite est observée , c'est-à-dire $Y_i = X_i \wedge C_i$

$$\delta_i = \mathbb{I}_{(Y_i = X_i)} = \mathbb{I}_{(X_i \leq C_i)} = \begin{cases} 1 & \text{si} \quad \text{non censure} \\ 0 & \text{si} \quad \text{censure} \end{cases}$$

En pratique la censure aléatoire peut avoir plusieurs causes : par exemple perte de vue, arrêt du traitement ou bien fin de l'étude. Alors ce qu'on observe c'est le couple (Y_i, δ_i) et $\delta_i = \mathbb{I}_{\{X_i \leq C_i\}}$ (l'indicatrice de non censure).

3.4 Détermination de la loi d'une durée de survie

Supposons que la durée de survie X soit une variable positive ou nulle, et absolument continue, alors sa loi de probabilité peut être définie par l'une des fonctions suivantes :

Définition 3.1 [Fonction de survie]

La fonction de survie noteé $S_X(t)$, est la probabilité pour un individu de vivre au moins jusqu'au temps t.

$$\begin{aligned} S_X(t) &= \mathbb{P}(X > t) \\ &= 1 - \mathbb{P}(X \leq t) \\ &= 1 - F_X(t) \end{aligned}$$

Si la fonction de répartition a une dérivée au point t alors la densité de X est donnée par :

$$f(t) = \lim_{h \to 0} \frac{F_X(t+h) - F_X(t)}{h} = F_X'(t) = -S_X'(t)$$

Définition 3.2 [Taux de hasard]

Le taux de hasard ou la fonction de risque λ est définie comme la probabilité qu'un individu fasse l'évènement considéré durant un intervalle de temps très court sachant qu'il a survécu jusqu'au début de l'intervalle

$$\begin{aligned}\lambda(t) &= \lim_{h\to 0} \frac{\mathbb{P}(X < t+h / X \geq t)}{h} \\ &= \lim_{h\to 0} \frac{\mathbb{P}(X < t+h, X \geq t)/\mathbb{P}(X \geq t)}{h} \\ &= \frac{1}{\mathbb{P}(X \geq t)} \lim_{h\to 0} \frac{\mathbb{P}(t \leq X < t+h)}{h} \\ &= \frac{f_X(t)}{S_X(t)} \\ &= \frac{f_X(t)}{1-F_X(t)}\end{aligned}$$

La fonction de risque mesure le risque instantané de survenue de l'évènement.

Définition 3.3 [Taux de hasard cumulé]

Le taux de hasard cumulé ou la fonction de risque cumulative évaluée au temps t est l'intégrale de la fonction de risque entre 0 et t

$$\Lambda(t) = \int_0^t \lambda(u) \quad du = -\log S(t)$$

On peut déduire la fonction de survie à partir du taux de hasard cumulé par la relation

$$S(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(u) \quad du\right)$$

3.5 Estimateur de la fonction de survie

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X_1, X_2, \ldots, X_n$ une suite de variables aléatoires (v.a) positives indépendantes et identiquement distribuées (i.i.d) désignant des durés de survie d'un évenement donné de fonction de répartition (f.d.r) F. Soit $C_1, C_2, \ldots, C_n$ une suite de (v.a) de censures, positives (i.i.d) de (f.d.r) G. Généralement, les (v.a) C_i sont supposées être indépendantes des X_i. Soit $(T_i, \delta_i)_{i=1,\ldots,n}$ l'échantillon réellement observé. où

$$T_i = \min(X_i, C_i) \quad \text{et} \quad \delta_i = \mathbb{I}_{\{X_i \leq C_i\}}$$

3.5.1 Estimateur de Kaplan-Meier

L'estimateur de la fonction de survie le plus utilisé lorsqu'aucune hypothèse ne veut être faite sur la distribution des temps de survie est l'estimateur de *Kaplan-Meier*. Cet estima-

teur (que l'on notera *EKM*) est aussi appelé estimateur *Product Limit(PL)* car il s'obtient comme limite d'un produit. L'ideé de la construction de *l'EKM* est la suivante, pour $t^{'} < t$

$$\begin{aligned} S(t) &= \mathbb{P}(X > t, X > t^{'}) \\ &= \mathbb{P}(X > t / X > t^{'}) S(t^{'}) \end{aligned}$$

On renouvelle l'opération en choisissant $t^{''} < t^{'}$ on obtient :

$$S(t^{'}) = \mathbb{P}(X > t^{'} / X > t^{''}) S(t^{''})$$

D'où

$$S(t) = \mathbb{P}(X > t / X > t^{'}) \quad \mathbb{P}(X > t^{'} + / X > t^{''}) S(t^{''})$$

Si on choisit pour dates où l'on conditionne celles où il s'est produit un évènement (décès ou censure) i.e $T_{(i)}$ on estime seulement des quantité de la forme

$$p_i := \mathbb{P}(X > T_{(i)} / X > T_{(i-1)})$$

p_i est la probabilité de survivre pendant l'intervalle $I_i =]T_{(i-1)}, T_{(i)}]$ quand on est vivant au début de cet intervalle. soit R_i le nombre de sujets à risque à l'instant $T_{(i)}$. M_i le nombre de décès observées à l'instant $T_{(i)}$ et $q_i = 1 - p_i =$ probabilité de mourir pendant l'intervalle I_i sachant qu'on était vivant au début de l'intervalle. Alors un estimateur naturel de q_i est

$$\widehat{q_i} = \frac{M_i}{R_i} = \frac{\textit{nombre de mort à l'instant } T_{(i)}}{\textit{nombre de sujets à risque}}$$

Supposons qu'il n'y ait pas d'exequo (c-à-d tous les $T_{(i)}$ sont différents). Si $\delta_{(i)} = 1$ il n'ya pas de censure à l'instant $T_{(i)}$, implique $M_i = 1$. Et si $\delta_{(i)} = 0$ il ya censure à l'instant $T_{(i)}$, implique $M_i = 0$. On a alors

$$\widehat{q_i} = \begin{cases} \frac{1}{R_i} & \text{si} \quad \delta_{(i)} = 1 \\ 0 & \text{si} \quad \delta_{(i)} = 0 \end{cases}$$

$\Longrightarrow$

$$\widehat{p_i} = 1 - \widehat{q_i} = \begin{cases} 1 - \frac{1}{R_t} & \text{si} \quad \delta_{(i)} = 1 \\ 1 & \text{si} \quad \delta_{(i)} = 0 \end{cases}$$

$\Longrightarrow$

$$\widehat{p_i} = \left(1 - \tfrac{1}{R_i}\right)^{\delta_{(i)}}$$

il est claire que $R_i = n - i + 1$. On obtient finalement l'*EKM* pour la fonction de survie de la variable durée de vie X :

$$\widehat{S}_{KM}(t) = 1 - \widehat{F}_{KM}(t) = \begin{cases} \prod_{T_{(i)} \leq t} \left(1 - \dfrac{1}{n-i+1}\right)^{\delta_{(i)}} & \text{si} \quad t < T_{(n)} \\ 0 & \text{si} \quad t \geq T_{(n)} \end{cases} \tag{3.1}$$

et donc on a aussi l'EKM pour la fonction de survie de la variable de censure C

$$\bar{G}_n(t) := 1 - \widehat{G}_{KM}(t) = \begin{cases} \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{n-i+1}\right)^{1-\delta_{(i)}} & \text{si} \quad t < T_{(n)} \\ 0 & \text{si} \quad t \geq T_{(n)} \end{cases} \tag{3.2}$$

où $(T_{(i)}, \delta_{(i)})_{i=1,\ldots,n}$ sont tels que $T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(n)}$ et les $\delta_{(i)}$ sont les indicatrice correspondantes.

Remarque 3.1 l'estimateur de *Kaplan Meier* peut aussi se mettre sous la forme suivante

$$\widehat{S}_{KM}(t) = \begin{cases} \prod_{i=1}^{n} \left(1 - \frac{1}{n-i+1}\right)^{\mathbb{I}\{T_{(i)} \leq t\}} & \text{si} \quad t < T_{(n)} \\ 0 & \text{si} \quad t \geq T_{(n)} \end{cases}$$

et

$$\bar{G}_n(t) = \begin{cases} \prod_{i=1}^{n} \left(1 - \frac{1}{n-i+1}\right)^{\mathbb{I}\{T_{(i)} \leq t\}} & \text{si} \quad t < T_{(n)} \\ 0 & \text{si} \quad t \geq T_{(n)} \end{cases} \tag{3.3}$$

3.6 Estimateur de la densité pour les données censureés

On considère le modèle de la censure aléatoire avec deux suites de *v.a* positives et *i.i.d* $X_1, X_2, \ldots, X_n$ représentant les durées de survie et $C_1, C_2, \ldots, C_n$ représentant les censures telle que $(X_i), (C_i)$ sont indépendentes $(i = 1, \ldots, n)$. Soient Fet G les fonctions de répartition de X et C, respectivement. On suppose que les (X_i) et (C_i) possèdent les densité f et g. on veut estimer f en utilisant les données observées suivantes :

$$T_i = min(X_i, C_i), \qquad \delta_{(i)} = \mathbb{I}\{X_i \leq C_i\}$$

Soit H la fonction de répartition de T et $T_{(1)}, T_{(2)}, \ldots, T_{(n)}$ désigne les statistiques d'ordre associées.
Blum et *Susarla* (1980) ont introduit l'estimateur à noyau de f dans ce cas , considéré ensuite par *Földes*, *Rejtö*, et *Winter* (1981). L'estimateur est basé sur l'estimateur de *Kaplan Meier* (3.1)

$$\widehat{f}_n(x) = \frac{1}{h_n} \int_0^{\infty} K\left(\frac{x-t}{b_n}\right) \widehat{F}_{KM}(dt) \tag{3.4}$$

où $(h_n)_n$ est une suite de nombres positif telle que $b_n \to 0$, $n\, b_n \to \infty$, K est un noyau.

3.7 Estimateur à noyau de la fonction régression pour des données censurées :

Dans cette partie on défini l'estimateur à noyau de régression conditionnelle dans le cas d'un modèle censurée. Ensuite on donne les hypothèses utilisée pour montré la convergence uniforme presque sûr de cet estimateur. Considérant une variable aléatoire réelle (v.r) Y et une suite de variables aléatoires réelles $(Y_i)_{i\geq 1}$ de même fonction de répartition absolument continue et inconnue (f.r) F et soit $(C_i)_{i\geq 1}$, une suite de variables aléatoire censurées de même (f.r) inconnue G. Soit X un vecteur aléatoire dans $\mathbb{R}^d$. Et soit $(X_i)_{i\geq 1}$ une suite de copies de vecteur aléatoire X et indiqué par $X_{i,1},\cdots,X_{i,d}$, cordonnées de X_i. Contrairement au modèle avec données completes, le modèle censuré utilise la suite des observations $(T_i,\delta_i,X_i)_{i\geq 1}$, où $T_i = Y_i \wedge C_i$ et $\delta_i = \mathbb{I}_{\{Y_i\leq C_i\}}$ sont observées.

3.7.1 Définitions

Supposons que $(Y_i)_{i\geq 1}$ et $(C_i)_{i\geq 1}$ soient deux suites de variables aléatoires stationnaires indépendantes. Posons

$$m(x) := \mathbb{E}(Y/X) = \frac{\int_{\mathbb{R}} y f_{X,Y}(x,y)dy}{\ell(x)} =: \frac{r_1(x)}{\ell(x)} \tag{3.5}$$

où $f_{X,Y}(\cdot,\cdot)$ est la densité conjointe de (X,Y) et $\ell(\cdot)$, la fonction de densité de X.
Il est bien connu que l'estimateur à noyau de la fonction de régression $m(.)$ dans le cas censurée (voir, par exemple *Carbonez* et *al* [7]) est donné par :

$$\tilde{m}_n(x) = \sum_{i=1}^{n} W_{in}(x)\frac{\delta_i T_i}{\overline{G}(\,T_i)} \tag{3.6}$$

où $\overline{G}$ est la fonction de survie de la v.a C,

$$W_{in}(x) = \frac{K_d\left(\frac{x-X_i}{h_n}\right)}{\sum\limits_{j=1}^{n} K_d\left(\frac{x-X_j}{h_n}\right)} \quad , \text{ sont les poids de Watson-Nadaraya}$$

et K_d est la fonction de densité de probabilité défini sur $\mathbb{R}^d$ et h_n une suite de nombre positif converge vers 0 quand n tend vers l'infini∞. Ainsi (3.6) peut s'écrire comme :

$$\tilde{m}_n(x) =: \frac{\tilde{r}_{1,n}(x)}{\ell_n(x)} \tag{3.7}$$

avec

$$\tilde{r}_{1,n}(x) = \frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\delta_i T_i}{\overline{G}(T_i)}K_d\left(\frac{x-X_i}{h_n}\right) \quad \text{et} \quad \ell_n(x) = \frac{1}{nh_n^d}\sum_{i=1}^{n}K_d\left(\frac{x-X_i}{h_n}\right) \tag{3.8}$$

En pratique, puisque $\overline{G}$ est généralement inconnue, on le remplace par l'estimateur de Kaplan-Meier [30] *EKM* $\overline{G}_n$ correspondant, alors un estimateur possible de $m(x)$ est donné par :

$$m_n(x) = \frac{\frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\delta_i T_i}{\overline{G}_n(T_i)}K_d\left(\frac{x-X_i}{h_n}\right)}{\frac{1}{nh_n^d}\sum_{i=1}^{n}K_d\left(\frac{x-X_i}{h_n}\right)} =: \frac{r_{1,n}(x)}{\ell_n(x)} \tag{3.9}$$

Soit $\tau_F = \sup\{y, \bar{F}(y) > 0\}$ et $\tau_G = \sup\{y, \bar{G}(y) > 0\}$ la borne supérieur de $\bar{F}$ et $\bar{G}$, respectivement. Supposons que $\tau_F < \infty$, $\bar{G}(\tau_F) > 0$ (ceci implique que $\tau_F \leq \tau_G$) et que C et (X,T) soient indépendantes. Soit $\mathcal{C}$ un ensemble de $\mathbb{R}^d$ qui est inclus dans $\mathcal{C}_0 = \{x \in \mathbb{R}^d \,/\, \ell(x)) > 0\}$.

Nous commençons avec un noyau standard, étudié par Guessoum et Ould-Said (2009) [23], et nous rappelons leur premier théorème (pour $d = 1$).

Hypothèses

A1. La fenêtre h_n satisfait : $\lim_{n\to+\infty} h_n = 0$, $\lim_{n\to+\infty}\frac{nh_n}{logn} = +\infty$ et $\log\log n = o(\frac{1}{h_n^\mu})$ where $0 < \mu < 1$

A2. Le noyau $K := K_1$ est bornée, symmetrique a support compact. Il est aussi Hölderien d'ordre $\gamma > 0$. En outre $\int_{\mathbb{R}} t\,|K(t)|\,dt < +\infty$ et $\int_{\mathbb{R}} t^2 K(t)dt < +\infty$.

A3. La fonction $r_1(x)$ défini dans (3.5) est deux fois différentiable et $\sup_{x\in\mathcal{C}}|r_1''(x)| < +\infty$.

A4. L'intégrale défini par $\int_{\mathbb{R}}\frac{y^2}{\bar{G}(y)}f_{X,Y}(x,y)dy =: r_2(x)$ est deux fois différentiable et $\sup_{x\in\mathcal{C}}\left|r_2''(x)\right| < +\infty$.

A5. La densité marginale $\ell(.)$ est deux fois différentiable et satisfait la condition de Lipschitz. En outre $\ell(x) > \Gamma$ quelque soit $x \in C$ et $\Gamma > 0$

Le théorème suivant donne le taux de convergence uniforme presque sûre de l'estimateur à noyau de la régression.

Théorème 3.1 (Guessoum, and Ould-Said [25])

Sous les hypothèses **A1-A5**, on a pour $n \to \infty$:

$$\sup_{x\in\mathcal{C}}|m_n(x) - m(x)| = O\left(\max\left\{\sqrt{\frac{\log n}{nh_n}}, h_n^2\right\}\right) \quad p.s.$$

Remarque 3.2 La preuve du théorème (3.1), est faite pour $d = 1$, on écrivant premièrement l'expression $|m_n(x) - m(x)|$ sous la forme :

$$\begin{aligned} |m_n(x) - m(x)| &= \left| \frac{r_{1,n}(x)}{\ell_n(x)} - \frac{r_1(x)}{\ell(x)} \right| \\ &= \left| \left(\frac{r_{1,n}(x)}{\ell_n(x)} - \frac{\tilde{r}_{1,n}(x)}{\ell_n(x)} \right) + \left(\frac{\tilde{r}_{1,n}(x)}{\ell_n(x)} - \frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{\ell_n(x)} \right) \right. \\ &\quad \left. + \left(\frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{\ell_n(x)} - \frac{r_1(x)}{\ell_n(x)} \right) + \left(\frac{r_1(x)}{\ell_n(x)} - \frac{r_1(x)}{\ell(x)} \right) \right| \end{aligned}$$

on contrôle alors chacun des quatre termes de la somme précédents dont le biais en utilisant l'inégalité de Bernstein qui est une inégalité exponentielle qui permet de quantifier la vitesse de convergence.

Chapitre 4

Estimation non paramétrique à noyau non symétrique

Sommaire

4.1 Introduction

Dans ce chapitre, nous présentons l'estimateur à noyau non symétrique de la densité, en particulier l'estimateur à noyau bâta proposé par *Chen*(1999) [8] dans le cas complet. Nous détaillons les propriétés élémentaires de cet estimateur telles que biais, variance. Ensuite nous étudions le comportement asymptotique de l'estimateur à noyau non symétrique de la régression dans le cas censurée en donnant le taux de convergence uniforme presque sûre.

4.2 Estimateur à noyau Bêta de la densité

Soient $X_1,\ldots,X_n$ des v.a.i.i.d de densité de probabilité inconnue f deux fois dérivables a support compact. On suppose le support compact connu, et égale a $[0,1]$. Soit $K_{p,q}$ la fonction densité d'une v.a Béta(p,q).
On considère l'estimateur à noyau béta pour f de paramètres $p=\frac{x}{h_n}+1,\ q=\frac{(1-x)}{h_n}+1$ avec $x\in[0,1]$, où h_n est le paramètre de lissage satisfaisant la condition $h_n \underset{n\to\infty}{\to} 0$, défini comme suit

$$\widehat{f_n}(x)=\frac{1}{n}\sum_{i=1}^{n}K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X_i) \tag{4.1}$$

il est similaire à l'estimateur à noyau standard (2.17) seulement on remplace le noyau fixé par un noyau bêta de paramètre $\frac{x}{h_n}+1$ et $\frac{(1-x)}{h_n}+1$ donné par

$$K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(t)=\frac{1}{\beta(\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1)}\,t^{\frac{x}{h_n}}(1-t)^{\frac{(1-x)}{h_n}}\mathbb{I}_{[0,1]}(t) \tag{4.2}$$

où $\beta(.,.)$ est la fonction bêta standard définie par

$$\beta(n,p)=\int_0^1 t^{n-1}(1-t)^{p-1}dt$$

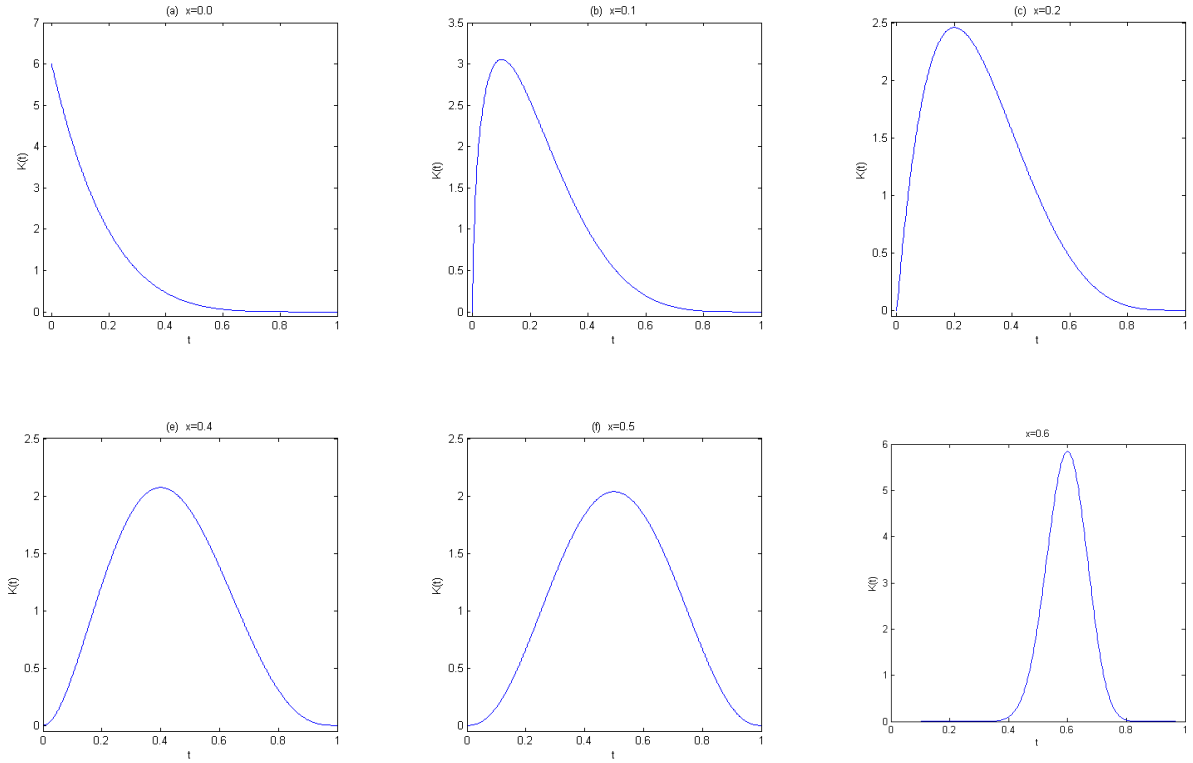

FIGURE 4.1 – Allure général d'un noyau bêta pour $h_n=0.2$

La figure (4.1) donne l'allure du noyau bêta et montre une particularité de la forme de ce noyau qui change selon la valeur de x comme il est représenter pour un h_n fixé. Cette particularité en faite à une influence sur la quantité du lissage appliqué par les estimateurs à noyau bêta.

4.3 Calcul du biais

On a

$$Biais\left(\widehat{f}_n(x)\right) = \mathbb{E}\left(\widehat{f}_n(x)\right) - f(x)$$

pour calculer le biais il suffit donc de calculer $\mathbb{E}\left(\widehat{f}_n(x)\right)$

$$\begin{aligned}
\mathbb{E}\left(\widehat{f}_n(x)\right) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X_i)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X_i)\right) \\
&= \int_0^1 K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(y) f(y) dy \\
&= \int_0^1 f(y) K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(y) dy
\end{aligned}$$

Si on considère que
$K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}$ est la densité d'une v.a $\xi_x \sim$ βéta $(\frac{x}{h_n}+1, \frac{(1-x)}{hn}+1)$ alors

$$\mathbb{E}\left(\widehat{f}_n(x)\right) = \mathbb{E}\left(f(\xi_x)\right) \tag{4.3}$$

D'après *Chen*([9],page 86) et *Johnson*, *Kotz* et *Balakrishman* [29] : si μ_{ξ_x} et $\sigma^2_{\xi_x}$ désignent respectivement la moyenne et la variance de ξ_x, Ils montrent alors qu'il existe une constante M telle que

$$\begin{gathered}
\mu_{\xi_x} = x + h_n(1-2x) + \Delta_1(x) \text{ et} \\
\sigma^2_{\xi_x} = h_n x(1-x) + \Delta_2(x)
\end{gathered}$$

où $\Delta_j(x) \leq M h_n^2 \quad , j = 1,2$. Ainsi le terme $\Delta_j(x) = O(h_n^2) \quad , j = 1,2$ donc

$$\mu_{\xi_x} = x + h_n(1-2x) + O(h_n^2) \tag{4.4}$$

$$\sigma^2_{\xi_x} = h_n x(1-x) + O(h_n^2) \tag{4.5}$$

faisons un développement de Taylor avec reste d'intégral d'ordre 1 pour $f(\xi_x)$ au voisinage de x

$$f(\xi_x) = f(x) + f^{'}(x)(\xi_x - x) + r_1(\xi_x - x) \tag{4.6}$$

avec

$$r_1(\xi_x - x) = \int_x^{\xi_x} (\xi_x - t) f^{''}(t) dt$$

Remarquons que :

$$\begin{aligned} r_1(\xi_x - x) &= \int_x^{\xi_x} (\xi_x - t) f^{''}(t) dt - \int_x^{\xi_x} (\xi_x - t) f^{''}(x) dt + \int_x^{\xi_x} (\xi_x - t) f^{''}(x) dt \\ &= \int_x^{\xi_x} (\xi_x - t) \left(f^{''}(t) - f^{''}(x) \right) dt + \int_x^{\xi_x} (\xi_x - t) f^{''}(x) dt \end{aligned}$$

avec le changement de variable $u = t - x, r_1(\xi_x - x)$ devient :

$$r_1(\xi_x - x) = \int_0^{\xi_x - x} (\xi_x - u - x) \left(f^{''}(u+x) - f^{''}(x) \right) du + \int_0^{\xi_x - x} (\xi_x - u - x) f^{''}(x) du$$

posons

$$r(\xi_x - x) = \int_0^{\xi_x - x} (\xi_x - u - x) \left(f^{''}(u+x) - f^{''}(x) \right) du$$

d'où

$$\begin{aligned} r_1(\xi_x - x) &= r(\xi_x - x) + f^{''}(x) \int_0^{\xi_x - x} (\xi_x - u - x) du \\ &= r(\xi_x - x) + f^{''}(x) \left[\xi_x u - xu - \frac{u^2}{2} \right]_0^{\xi_x - x} \end{aligned}$$

$$\begin{aligned} r_1(\xi_x - x) &= r(\xi_x - x) + f''(x)\left[\xi_x(\xi_x - x) - x(\xi_x - x) - \frac{1}{2}(\xi_x - x)^2\right] \\ &= r(\xi_x - x) + f''(x)\left[(\xi_x - x)^2 - \frac{1}{2}(\xi_x - x)^2\right] \\ &= r(\xi_x - x) + \frac{1}{2}f''(x)(\xi_x - x)^2 \end{aligned}$$

remplaçons l'expression de $r_1(\xi_x - x)$ précédente dans l'équation (4.6) on obtient :

$$f(\xi_x) = f(x) + f'(x)(\xi_x - x) + \frac{1}{2}f''(x)(\xi_x - x)^2 + r(\xi_x - x) \tag{4.7}$$

faisons le changement de variable $s = \dfrac{u}{\sqrt{h_n}}, \quad du = \sqrt{h_n}\, ds$

$$\begin{aligned} r(\xi_x - x) &= \int_0^{\frac{\xi_x - x}{\sqrt{h_n}}} (\xi_x - \sqrt{h_n}s - x)\left(f''(\sqrt{h}s + x) - f''(x)\right)\sqrt{h_n}ds \\ &= \sqrt{h_n}\int_0^{\frac{\xi_x - x}{\sqrt{h_n}}} (\sqrt{h_n}\,\frac{(\xi_x - x)}{\sqrt{h_n}} - \sqrt{h_n}s)\left(f''(x + \sqrt{h_n}s) - f''(x)\right)ds \end{aligned}$$

Soit g la densité de la v.a $Y = \dfrac{(\xi_x - x)}{\sqrt{h_n}}$, désignons par y une réalisation de Y

$$\begin{aligned} r(\xi_x - x) &= \sqrt{h_n}\int_0^{y} (\sqrt{h_n}\, y - \sqrt{h_n}s)\left(f''(x + \sqrt{h_n}s) - f''(x)\right)ds \\ &= h_n\int_0^{y} (y - s)\left(f''(x + \sqrt{h_n}s) - f''(x)\right)ds \end{aligned}$$

On peut voir $r(\xi_x - x)$ comme une fonction de y d'où :

$$\begin{aligned} \mathbb{E}\left(r(\xi_x - x)\right) &= h_n\int\int_0^{y} (y - s)\left(f''(x + \sqrt{h_n}s) - f''(x)\right)ds\, g(y)dy \\ &\leq k\sqrt{h_n}s \end{aligned}$$

comme f'' est continue uniformément sur $[0,1]$, d'aprés le théorème de convergence dominée l'intégrale sur la partie droite converge uniformément vers 0.c'est-à-dire

$$\mathbb{E}\left(r(\xi_x - x)\right) \xrightarrow[h_n \to 0]{C.V.U} 0$$

puis en prenant l'espérance de chaque côté de (4.7) et à partir de (4.4) et (4.5) on a

$$\mathbb{E}\left(f(\xi_x)\right) = f(x) + f^{'}(x)\mathbb{E}\left((\xi_x - x)\right) + \frac{1}{2}f^{''}(x)\mathbb{E}\left((\xi_x - x)^2\right)$$

or

$$\begin{aligned}
\mathbb{E}\left((\xi_x - x)\right) &= \mathbb{E}(\xi_x) - x \\
&= \mu_{\xi_x} - x \\
&= x + h_n(1-2x) + O(h_n^2) - x \\
&= h_n(1-2x) + O(h_n^2)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left((\xi_x - x)^2\right) &= \mathbb{E}\left((\xi_x)^2 - 2\xi_x x + x^2\right) \\
&= \mathbb{E}((\xi_x)^2) - 2x\mathbb{E}(\xi_x) + x^2 \\
&= \mathbb{V}ar(\xi_x) + [\mathbb{E}(\xi_x)]^2 - 2x\mu_{\xi_x} + x^2 \\
&= \sigma_{\xi_x}^2 + \mu_{\xi_x}^2 - 2x\mu_{\xi_x} + x^2 \\
&= h_n x(1-x) + h_n^2(1-2x)^2 + O(h_n^2)
\end{aligned}$$

d'où

$$\begin{aligned}
\mathbb{E}\left(f(\xi_x)\right) &= f(x) + f^{(1)}(x)h_n(1-2x) + \frac{1}{2}f^{(2)}(x)\left[h_n x(1-x) + h_n^2(1-2x)^2\right] + O(h_n^2) \\
&= f(x) + f^{(1)}(x)h_n(1-2x) + \frac{1}{2}f^{(2)}(x)h_n x(1-x) + o(h_n) \qquad (4.8)
\end{aligned}$$

donc le biais de l'estimateur à noyau bêta est :

$$biais\left(\widehat{f_n}(x)\right) = \left\{(1-2x)f^{(1)}(x) + \frac{1}{2}x(1-x)f^{(2)}(x)\right\}h_n + o(h_n) \qquad (4.9)$$

où le reste est uniformément $o(h_n)$ pour $x \in [0,1]$, le biais est de $O(h_n)$ dans $[0,1]$, indiquant que $\widehat{f_n}$ n'est pas biaisé aux bornes.

4.4 Calcul de la variance

On a

$$\begin{aligned}
\mathbb{V}ar\left(\widehat{f}_n(x)\right) &= \mathbb{V}ar\left(\frac{1}{n}\sum_{i=1}^{n} K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X_i)\right) \\
&= \frac{1}{n^2}\, n\mathbb{V}ar\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X)\right) \\
&= \frac{1}{n}\left\{\mathbb{E}\left[\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X)\right)^2\right] - \left[\mathbb{E}\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X)\right)\right]^2\right\} \\
&= \frac{1}{n}\left\{\mathbb{E}\left[\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X)\right)^2\right] - \left[\mathbb{E}\left(\widehat{f}_n(x)\right)\right]^2\right\}
\end{aligned}$$

d'une part on a

$$\mathbb{E}\left[\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(X)\right)^2\right] = \int_0^1 \left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(y)\right)^2 f(y)dy$$

et

$$\begin{aligned}
\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(y)\right)^2 &= \frac{1}{\beta\left(\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1\right)}(y^{\frac{x}{h_n}})^2\left((1-y)^{\frac{(1-x)}{h_n}}\right)^2 \\
&= \frac{1}{\beta\left(\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1\right)} y^{\frac{2x}{h_n}}(1-y)^{\frac{2(1-x)}{h_n}}
\end{aligned}$$

Si on multiplie et on divise par le même terme $\beta\left(\frac{2x}{h_n}+1,\frac{2(1-x)}{h_n}+1\right)$ on obtient

$$\begin{aligned}
\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(y)\right)^2 &= \frac{\beta\left(\frac{2x}{h_n}+1,\frac{2(1-x)}{h_n}+1\right)}{\beta\left(\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1\right)}\,\frac{1}{\beta\left(\frac{2x}{h_n}+1,\frac{2(1-x)}{h_n}+1\right)}(y^{\frac{x}{h_n}})^2 \\
&\quad \left((1-y)^{\frac{(1-x)}{h_n}}\right)^2 \\
&= A_{h_n}(x)\,\frac{1}{\beta\left(\frac{2x}{h_n}+1,\frac{2(1-x)}{h_n}+1\right)} y^{\frac{2x}{h_n}}(1-y)^{\frac{2(1-x)}{h_n}}
\end{aligned}$$

où

$$A_{h_n}(x) = \frac{\beta\left(\frac{2x}{h_n}+1, \frac{2(1-x)}{h_n}+1\right)}{\beta^2\left(\frac{x}{h_n}+1, \frac{(1-x)}{h_n}+1\right)} \tag{4.10}$$

donc on trouve que

$$\begin{aligned}
\mathbb{E}\left[\left(K_{\frac{x}{h_n}+1,\frac{(1-x)}{h_n}+1}(y)\right)^2\right] &= \int_0^1 A_{h_n}(x)\,\frac{1}{\beta\left(\frac{2x}{h_n}+1, \frac{2(1-x)}{h_n}+1\right)} y^{\frac{2x}{h_n}}(1-y)^{\frac{2(1-x)}{h_n}} f(y)dy \\
&= A_{h_n}(x)\int_0^1 \frac{1}{\beta\left(\frac{2x}{h_n}+1, \frac{2(1-x)}{h_n}+1\right)} y^{\frac{2x}{h_n}+1-1}(1-y)^{\frac{2(1-x)}{h_n}+1-1} \\
&\quad f(y)dy \\
&= A_{h_n}(x)\mathbb{E}[f(\gamma_x)] \qquad (4.11)
\end{aligned}$$

où γ_x est une v.a qui suit une loi béta$(\frac{2x}{h_n}+1, \frac{2(1-x)}{h_n}+1)$

Pour déterminer l'ordre de grandeur de $A_{h_n}(x)$, on a besoin du lemme suivant (voir (3.5) *Chen*(2000)[9], p 77)

Lemme 4.1 Pour h_n assez petit,

$$A_{h_n}(x) \leq \frac{1}{2\sqrt{\pi}}[x(1-x)]^{\frac{-1}{2}} h_n(h_n^{-1}+1)^{\frac{3}{2}}, \qquad \forall x \in [0,1]$$

et aussi

$$A_{h_n}(x) \sim \begin{cases} \frac{1}{2\sqrt{\pi}}[x(1-x)]^{-1/2}h_n^{-1/2} & \text{si} \quad \frac{x}{h_n} \quad \text{et} \frac{(1-x)}{h_n} \to \infty \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}h_n^{-1} & \text{si} \quad \frac{x}{h_n} \to \kappa \quad \text{et} \frac{(1-x)}{b} \to \kappa \end{cases}$$

pour une constante positive κ

La variance est donc le résultat du lemme(4.1) et de l'équation (4.11)

$$\mathbb{V}ar\left(\widehat{f}_n(x)\right) = \begin{cases} \frac{1}{2\sqrt{\pi}}\frac{n^{-1}h_n^{-1/2}}{[x(1-x)]^{1/2}}[f(x)+O(n^{-1})] & \text{si} \quad \frac{x}{h_n} \quad \text{et} \quad \frac{(1-x)}{h_n} \to \infty \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}n^{-1}h_n^{-1}[f(x)+O(n^{-1})] & \text{si} \quad \frac{x}{h_n} \to \kappa \quad \text{et} \quad \frac{(1-x)}{h_n} \to \kappa \end{cases}$$

La variance asymptotique est d'ordre plus grand $(n^{-1}h_n^{-1})$ près des bornes que à l'intérieur $(n^{-1}h_n^{-1/2})$

4.5 Estimateur à noyau non symétrique de la fonction de régression pour des données censurées

Dans cette partie l'objectif est d'établir la convergence forte uniforme pour un estimateur à noyau de la régression pour des données censurées quand la condition de symétrie sur le noyau est retirée. Donc l'estimateur est celui défini dans (3.9), la seul différence est de remplacer le noyau par un noyau non symétrique c'est-à-dire

$$m_n(x) = \frac{\frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\delta_i T_i}{\bar{G}_n(T_i)}K_d\left(\frac{x-X_i}{h_n}\right)}{\frac{1}{nh_n^d}\sum_{i=1}^{n}K_d\left(\frac{x-X_i}{h_n}\right)} =: \frac{r_{1,n}(x)}{\ell_n(x)} \tag{4.12}$$

Pour $d > 1$ et pour un noyau non symétrique, posons les hypothèses suivantes :

Hypothèses

H1) La fenêtre h_n satisfait :

$i)$ $\lim_{n\to+\infty} h_n = 0$, $\lim_{n\to+\infty}\frac{nh_n^d}{\log n} = +\infty$,

$ii)$ $\sqrt{\frac{\log\log n}{n}} = o(h_n)$

H2) Le noyau K_d est bornée et satisfait :

$i)$ $\int_{\mathbb{R}^d}\|t\|K_d(t)dt < +\infty$,

$ii)$ $\int_{\mathbb{R}^d}(t_1+t_2+...+t_d)K_d^2(t)dt < +\infty$ et $\int_{\mathbb{R}^d}K_d^2(t)dt < +\infty$

$iii)$ $\forall(t,s)\in\mathcal{C}^2 \quad |K_d(t)-K_d(s)| \leq \|t-s\|^{\gamma}$ pour $\gamma > 0$.

H3) La fonction $r_1(\cdot)$ défini dans (3.5) est continûment différentiable, et $\sup_{x\in C}\left|\frac{\partial r_1}{\partial x_i}(x)\right| < +\infty$ pour $i = 1,...,d$.

H4) La fonction $r_2(x) := \int_{\mathbb{R}}\frac{y^2}{\bar{G}(y)}\, f_{X,Y}(x,y)dy$ est continûment différentiable, et $\sup_{x\in C}\left|\frac{\partial r_2}{\partial x_i}(x)\right| < +\infty$ pour $i = 1,...,d$.

H5) $\exists D_1 > 0$ et $\exists D_2 > 0$ tel que $\sup_{u,v\in C}\left|\ell_{ij}(u,v)\right| < D_1$ et $\sup_{u\in C}|\ell(u)| < D_2$, où ℓ_{ij} est la densité conjointe de (X_i,X_j).

H6) La densité $\ell(.)$ est continûment différentiable et $\sup_{x\in C}\left|\frac{\partial \ell}{\partial x_i}(x)\right| < +\infty$ pour $i = 1,...,d$. de plus il existe $\xi > 0$ tel que $\ell(x) > \xi \quad \forall x \in C$.

Remarque 4.1 On voit que les hypothèses H1-H5 sont semblables à celles du théorème (3.1). Les principales différences sont dues au fait que pour un noyau non symétrique nous n'avons pas besoin de conditions sur les dérivées secondes parce que, techniquement, les calculs sont basés sur l'expression de Taylor d'ordre un.

Théorème 4.1 Sous les hypothèses **H1-H6**, on a

$$\sup_{x\in\mathcal{C}}|m_n(x)-m(x)| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right)+O(h_n)\ \ p.s \quad quand\ \ n\to\infty.$$

Démonstration 4.1 L'idée de base est d'écrire le processus $|m_n(x)-m(x)|$ sous la forme

$$\begin{aligned}|m_n(x)-m(x)| &= \left|\frac{r_{1,n}(x)}{\ell_n(x)}-\frac{r_1(x)}{\ell(x)}\right| \\ &= \left|\left(\frac{r_{1,n}(x)}{\ell_n(x)}-\frac{\tilde{r}_{1,n}(x)}{\ell_n(x)}\right)+\left(\frac{\tilde{r}_{1,n}(x)}{\ell_n(x)}-\frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{\ell_n(x)}\right)\right. \\ &\quad \left.+\left(\frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{\ell_n(x)}-\frac{r_1(x)}{\ell_n(x)}\right)+\left(\frac{r_1(x)}{\ell_n(x)}-\frac{r_1(x)}{\ell(x)}\right)\right|\end{aligned}$$

et on a

$$\begin{aligned}\sup_{x\in\mathcal{C}}|m_n(x)-m(x)| &\le \sup_{x\in\mathcal{C}}\left|\frac{r_{1,n}(x)}{\ell_n(x)}-\frac{\tilde{r}_{1,n}(x)}{\ell_n(x)}\right|+\sup_{x\in\mathcal{C}}\left|\frac{\tilde{r}_{1,n}(x)}{\ell_n(x)}-\frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{\ell_n(x)}\right| \\ &\quad +\sup_{x\in\mathcal{C}}\left|\frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{\ell_n(x)}-\frac{r_1(x)}{\ell_n(x)}\right|+\sup_{x\in\mathcal{C}}\left|\frac{r_1(x)}{\ell_n(x)}-\frac{r_1(x)}{\ell(x)}\right| \\ &\le \sup_{x\in\mathcal{C}}\left(\frac{1}{\ell_n(x)}\right)\left\{\sup_{x\in\mathcal{C}}|r_{1,n}(x)-\tilde{r}_{1,n}(x)|+\sup_{x\in\mathcal{C}}|\tilde{r}_{1,n}(x)-\mathbb{E}(\tilde{r}_{1,n}(x))|\right. \\ &\quad \left.+\sup_{x\in\mathcal{C}}|\mathbb{E}(\tilde{r}_{1,n}(x))-r_1(x)|+\sup_{x\in\mathcal{C}}\left|r_1(x)\frac{\ell(x)-\ell_n(x)}{\ell(x)}\right|\right\} \\ &\le \frac{1}{\inf_{x\in\mathcal{C}}(\ell_n(x))}\left\{\underbrace{\sup_{x\in\mathcal{C}}|r_{1,n}(x)-\tilde{r}_{1,n}(x)|}_{I_1}+\underbrace{\sup_{x\in\mathcal{C}}|\tilde{r}_{1,n}(x)-\mathbb{E}(\tilde{r}_{1,n}(x))|}_{I_2}\right. \\ &\quad \left.+\underbrace{\sup_{x\in\mathcal{C}}|\mathbb{E}(\tilde{r}_{1,n}(x))-r_1(x)|}_{I_3}+\sup_{x\in\mathcal{C}}\left(|r_1(x)|\left(\frac{1}{\ell(x)}\right)\right)\underbrace{\sup_{x\in\mathcal{C}}|\ell(x)-\ell_n(x)|}_{I_4}\right\} \\ &\le \frac{1}{\inf_{x\in\mathcal{C}}\ell_n(x)}\left\{I_1+I_2+I_3+\sup_{x\in\mathcal{C}}\left(|r_1(x)|\,\xi^{-1}\right)I_4\right\}\end{aligned}$$

Calcul de I_1

$$
\begin{aligned}
I_1 &= \sup_{x\in C}|r_{1,n}(x)-\tilde{r}_{1,n}(x)| \\
&= \sup_{x\in C}\left|\frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\delta_i T_i}{\bar{G}_n(T_i)}K_d\left(\frac{x-X_i}{h_n}\right)-\frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\delta_i T_i}{\bar{G}(T_i)}K_d\left(\frac{x-X_i}{h_n}\right)\right| \\
&= \sup_{x\in C}\left|\frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\mathbb{I}_{\{Y_i\le C_i\}}\,(Y_i\wedge C_i)}{\bar{G}_n(Y_i\wedge C_i)}K_d\left(\frac{x-X_i}{h_n}\right)-\frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\mathbb{I}_{\{Y_i\le C_i\}}\,(Y_i\wedge C_i)}{\bar{G}(Y_i\wedge C_i)}K_d\left(\frac{x-X_i}{h_n}\right)\right| \\
&\le \sup_{x\in C}\frac{1}{nh_n^d}\sum_{i=1}^{n}\left|\frac{Y_i}{\bar{G}_n(Y_i)}K_d\left(\frac{x-X_i}{h_n}\right)-\frac{Y_i}{\bar{G}(Y_i)}K_d\left(\frac{x-X_i}{h_n}\right)\right|, \quad \text{car}(Y_i\wedge C_i)=Y_i,\ \delta_i=1 \\
&\le \sup_{x\in C}\frac{1}{nh_n^d}\sum_{i=1}^{n}\left|Y_iK_d\left(\frac{x-X_i}{h_n}\right)\left[\frac{1}{\bar{G}_n(Y_i)}-\frac{1}{\bar{G}(Y_i)}\right]\right| \\
&\le \sup_{x\in D}\frac{1}{nh_n^d}\sum_{i=1}^{n}\left|Y_iK_d\left(\frac{x-X_i}{h_n}\right)\right|\left|\frac{\bar{G}(Y_i)-\bar{G}_n(Y_i)}{\bar{G}_n(Y_i)\bar{G}(Y_i)}\right| \\
&\le \frac{1}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)}\,\frac{\tau_F}{nh_n^d}\sum_{i=1}^{n}\left|K_d\left(\frac{x-X_i}{h_n}\right)\right|\sup_{t\le\tau_F}(|\bar{G}_n(t)-\bar{G}(t)|)
\end{aligned}
$$

D'aprés l'utilisation de la loi des grand nombre et la loi du logarithme itéré (Deheuvels et Einmahl 2000 [14]) on obtient : $\sup_{t\le\tau_F}(|\bar{G}_n(t)-\bar{G}(t)|)\le\sqrt{\frac{\log\log n}{n}}$. D'où

$$
\begin{aligned}
I_1 &\le \frac{\tau_F}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)}\,\frac{1}{nh_n^d}\sum_{i=1}^{n}\left|K_d\left(\frac{x-X_i}{h_n}\right)\right|\sqrt{\frac{\log\log n}{n}} \\
&\le \frac{\tau_F}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)}\,\mathbb{E}\left(\frac{1}{h_n^d}K_d\left(\frac{x-X_i}{h_n}\right)\right)\sqrt{\frac{\log\log n}{n}} \\
&\le C\sqrt{\frac{\log\log n}{n}} \\
&\le C\ o(h_n), \qquad (\text{ d'aprés } \mathbf{H1}\ ii)\ ) \\
&= O(h_n)
\end{aligned}
$$

Calcul du I_3

$$\begin{aligned} I_3 &= \sup_{x\in C}|\mathbb{E}(\tilde{r}_{1,n}(x)) - r_1(x)| \\ &= \sup_{x\in C}\left|\mathbb{E}\left(\frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\delta_i T_i}{\bar{G}(T_i)}K_d\left(\frac{x-X_i}{h_n}\right)\right) - r_1(x)\right| \\ &= \sup_{x\in C}\left|\mathbb{E}\left(\frac{1}{h_n^d}\frac{\delta_1 T_1}{\bar{G}(T_1)}K_d\left(\frac{x-X_1}{h_n}\right)\right) - r_1(x)\right| \end{aligned}$$

et

$$\begin{aligned} \mathbb{E}\left(\frac{1}{h_n^d}\frac{\delta_1 T_1}{\bar{G}(T_1)}K_d\left(\frac{x-X_1}{h_n}\right)\right) &= \mathbb{E}\left\{\mathbb{E}\left(\frac{1}{h_n^d}\frac{\delta_1 T_1}{\bar{G}(T_1)}K_d\left(\frac{x-X_1}{h_n}\right)/X_1\right)\right\} \\ &= \mathbb{E}\left\{\frac{1}{h_n^d}K_d\left(\frac{x-X_1}{h_n}\right)\mathbb{E}\left(\frac{\delta_1 T_1}{\bar{G}(T_1)}|X_1\right)\right\} \\ &= \int_{\mathbb{R}^d}\frac{1}{h_n^d}K_d\left(\frac{x-u}{h_n}\right)\mathbb{E}\left(\frac{\delta_1 T_1}{\bar{G}(T_1)}|X_1=u\right)\ell(u)\ du \end{aligned}$$

mais,

$$\begin{aligned} \mathbb{E}\left(\frac{\delta_1 T_1}{\bar{G}(T_1)}/X_1=u\right) &= \mathbb{E}\left(\frac{\mathbb{I}_{\{Y_1\le C_1\}}Y_1}{\bar{G}(Y_1)}/X_1=u\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\frac{\mathbb{I}_{\{Y_1\le C_1\}}Y_1}{\bar{G}(Y_1)}/Y_1\right)/X_1=u\right) \\ &= \mathbb{E}\left(\frac{Y_1}{\bar{G}(Y_1)}\mathbb{E}\left(\mathbb{I}_{\{Y_1\le C_1\}}/Y_1\right)/\ X_1=u\right) \\ &= \mathbb{E}\left(\frac{Y_1}{\bar{G}(Y_1)}\mathbb{P}\left(Y_1\le C_1/Y_1\right)/\ X_1=u\right) \\ &= \mathbb{E}\left(\frac{Y_1}{\bar{G}(Y_1)}\mathbb{P}\left(Y_1\le C_1\right)/\ X_1=u\right) \\ &= \mathbb{E}\left(\frac{Y_1}{\bar{G}(Y_1)}\bar{G}(Y_1)/\ X_1=u\right) \\ &= \mathbb{E}\left[Y_1/X_1=u\right]=m(u) \end{aligned}$$

alors on a

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{h_n^d}\frac{\delta_1 T_1}{\bar{G}(T_1)}K_d\left(\frac{x-X_1}{h_n}\right)\right) &= \int_{\mathbb{R}^d}\frac{1}{h_n^d}K_d\left(\frac{x-u}{h_n}\right)m(u)\ell(u)\ du \\
&= \int_{\mathbb{R}^d}\frac{1}{h_n^d}K_d\left(\frac{x-u}{h_n}\right)\frac{r_1(u)}{\ell(u)}\ell(u)\ du \\
&= \int_{\mathbb{R}^d}\frac{1}{h_n^d}K_d\left(\frac{x-u}{h_n}\right)r_1(u)\ du \\
&= \int_{\mathbb{R}^d}K_d(t)\,r_1(x-h_nt)\ dt, \qquad t=\left(\frac{x-u}{h_n}\right)
\end{aligned}$$

donc

$$\begin{aligned}
I_3 &= \sup_{x\in C}\left|\int_{\mathbb{R}^d}K_d(t)\,r_1(x-h_nt)\ dt - r_1(x)\right| \\
&= \sup_{x\in C}\left|\int_{\mathbb{R}^d}K_d(t)\,[r_1(x-h_nt)-r_1(x)]\,dt\right|
\end{aligned}$$

Un développement de Taylor d'ordre 1 au voisinage de x' donne

$$r_1(x-h_nt)-r_1(x) = -h_n\left(t_1\frac{\partial r_1}{\partial x_1}(x')+\ldots+t_d\frac{\partial r_1}{\partial x_d}(x')\right)$$

avec x' entre $x-h_nt$ et x. Ainsi

$$\begin{aligned}
I_3 &= \sup_{x\in C}\left|\int_{\mathbb{R}^d}K_d(t)\,[r_1(x-h_nt)-r_1(x)]\,dt\right| \\
&= \sup_{x\in C}\left|-h_n\int_{\mathbb{R}^d}K_d(t)\left(t_1\frac{\partial r_1}{\partial x_1}(x')+\ldots+t_d\frac{\partial r_1}{\partial x_d}(x')\right)dt\right| \\
&\leq h_n\sup_{x\in C}\int_{\mathbb{R}^d}\left|K_d(t)(t_1\frac{\partial r_1}{\partial x_1}(x')+\ldots+t_d\frac{\partial r_1}{\partial x_d}(x'))dt\right| \\
&\leq h_n\,C,
\end{aligned}$$

Les hypothèses **H1** i)**, H2** i) et **H3** donnent

$$I_3 = \sup_{x\in C}|\mathbb{E}\,(\tilde{r}_{1,n}(x))-r_1(x)| = O(h_n) \qquad p.s.$$

Calcul du I_2 Pour contrôler I_2, on utilise un recouvrement de $\mathcal{C}$ (comme il est compact) par un nombre fini $\mathfrak{s}_n$ de boule $\mathcal{B}_k(x_k^*, h_n^{d\eta})$ centré en $x_k^* = (x_{1,k}^*, ...x_{d,k}^*)$, $k \in \{1, ..., \mathfrak{s}_n\}$, avec $\eta > \frac{1}{d} + \frac{1}{\gamma}$, (γ est le même comme dans la condition **H2**). Puisque $\mathcal{C}$ est bornée il existe une constante $M > 0$ tel que $\mathfrak{s}_n \leq \frac{M}{h_n^{d\eta}}$.
En suite on définit , pour $x \in \mathcal{C}$:

$$\Delta_i(x) = \frac{1}{nh_n^d} \frac{\delta_i T_i}{\bar{G}(T_i)} K_d\left(\frac{x - X_i}{h_n}\right) - \mathbb{E}\left(\frac{1}{nh_n^d} \frac{\delta_1 T_1}{\bar{G}(T_1)} K_d\left(\frac{x - X_1}{h_n}\right)\right)$$

Il est évident, d'aprés (3.8) que

$$\sum_{i=1}^{n} \Delta_i(x) = \tilde{r}_{1,n}(x) - \mathbb{E}(\tilde{r}_{1,n}(x)).$$

écrivons $\Delta_i(x) - \Delta_i(x_k^*) =: \tilde{\Delta}_i(x)$, on a clairement $|\Delta_i(x)| \leq \left|\tilde{\Delta}_i(x)\right| + \left|\Delta_i(x_k^*)\right|$.
d'une part on a

$$\begin{aligned}
\tilde{\Delta}_i(x) &= \frac{1}{nh_n^d} \frac{\delta_i T_i}{\bar{G}(T_i)} K_d\left(\frac{x - X_i}{h_n}\right) - \mathbb{E}\left(\frac{1}{nh_n^d} \frac{\delta_1 T_1}{\bar{G}(T_1)} K_d\left(\frac{x - X_1}{h_n}\right)\right) \\
&\quad - \frac{1}{nh_n^d} \frac{\delta_i T_i}{\bar{G}(T_i)} K_d\left(\frac{x_k^* - X_i}{h_n}\right) + \mathbb{E}\left(\frac{1}{nh_n^d} \frac{\delta_1 T_1}{\bar{G}(T_1)} K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right) \\
&= \frac{1}{nh_n^d} \frac{\delta_i T_i}{\bar{G}(T_i)} \left[K_d\left(\frac{x - X_i}{h_n}\right) - K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right] \\
&\quad - \mathbb{E}\left(\frac{1}{nh_n^d} \frac{\delta_1 T_1}{\bar{G}(T_1)} \left[K_d\left(\frac{x_k^* - X_1}{h_n}\right) - K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right]\right)
\end{aligned}$$

Comme K_d est hölderienne (par l'hypothèse **H2** *iii*))on a :

$$\left[K_d\left(\frac{x_k^* - X_1}{h_n}\right) - K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right] \leq \left\|\frac{x - x_k^*}{h_n}\right\|^{\gamma}$$

$$
\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\sum_{i=1}^{n}\tilde{\Delta}_i(x)\right| &\leq \sup_{x\in\mathcal{C}}\sum_{i=1}^{n}\left\{\frac{1}{nh_n^d}\frac{\delta_i|T_i|}{\bar{G}(T_i)}\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}-\mathbb{E}\left(\frac{1}{nh_n^d}\frac{\delta_1|T_1|}{\bar{G}(T_1)}\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}\right)\right\}\\
&\leq \sup_{x\in\mathcal{C}}\left\{\frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\delta_i|T_i|}{\bar{G}(T_i)}\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}-\mathbb{E}\left(\frac{n}{nh_n^d}\frac{\delta_1|T_1|}{\bar{G}(T_1)}\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}\right)\right\}\\
&\leq \sup_{x\in\mathcal{C}}\left\{\mathbb{E}\left(\frac{1}{h_n^d}\frac{\delta_1|T_1|}{\bar{G}(T_1)}\right)\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}-\mathbb{E}\left(\frac{1}{h_n^d}\frac{\delta_1|T_1|}{\bar{G}(T_1)}\right)\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}\right\}\\
&\leq \sup_{x\in\mathcal{C}}\left\{2\mathbb{E}\left(\frac{1}{h_n^d}\frac{\delta_1|T_1|}{\bar{G}(T_1)}\right)\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}\right\}\\
&\leq \sup_{x\in\mathcal{C}}\left\{2\mathbb{E}\left(\frac{1}{h_n^d}\frac{|Y_1|}{\bar{G}(Y_1)}\right)\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}\right\}\\
&\leq \sup_{x\in\mathcal{C}}\left(\frac{2\mathbb{E}(|Y_1|)}{\bar{G}(\tau_F)}\frac{1}{h_n^d}\left\|\frac{x-x_k^*}{h_n}\right\|^{\gamma}\right)\\
&\leq \frac{\mathbb{E}(|Y_1|)}{\bar{G}(\tau_F)}\frac{h_n^{d\gamma\eta}}{h_n^{\gamma+d}}
\end{aligned}
$$

L'hypothèse **H1** et la condition sur η donnent : $\sup_{x\in\mathcal{C}}\left|\sum_{i=1}^{n}\tilde{\Delta}_i(x)\right| = o(1) \quad p.s.$

d'autre part, pour toute $\varepsilon > 0$, on a

$$
\mathbb{P}\left\{\max_{k=1,\ldots,\mathfrak{s}_n}\left|\sum_{i=1}^{n}\Delta_i(x_k^*)\right|>\varepsilon\right\}\leq\sum_{i=1}^{\mathfrak{s}_n}\mathbb{P}\left\{\left|\sum_{i=1}^{n}\Delta_i(x_k^*)\right|>\varepsilon\right\}. \tag{4.13}
$$

et soit $U_i = nh_n^d\Delta_i(x_k^*)$. On a $\mathbb{E}(U_i)=0$ et

$|U_i| \leq 2\tau_F\overline{K}\,\upsilon =: M_1$, avec $\overline{K}$ est la borne supérieur de K et $\upsilon = \dfrac{1}{\bar{G}(\tau_F)}$. Pour appliquer l'inégalité de Bernstein, nous calculons $S^2 = Var(U_i)$, c'est-à-dire

$$
S^2 = \mathbb{E}\left[\frac{\delta_1^2T_1^2}{\bar{G}^2(T_1)}K_d^2\left(\frac{x_k^*-X1}{h_n}\right)\right]-\mathbb{E}^2\left[\frac{\delta_1T_1}{\bar{G}(T_1)}K_d\left(\frac{x_k^*-X_1}{h_n}\right)\right]
$$

D'autre part, en utilisant les propriétés de l'espérance conditionnelle, on obtient

$$
\begin{aligned}
\mathbb{E}\left[\frac{\delta_1^2T_1^2}{\bar{G}^2(T_1)}K_d^2\left(\frac{x_k^*-X_1}{h_n}\right)\right] &= \mathbb{E}\left[K_d^2\left(\frac{x_k^*-X_1}{h_n}\right)\mathbb{E}\left(\frac{\delta_1^2T_1^2}{\bar{G}^2(T_1)}/X_1\right)\right]\\
&= \int_{\mathbb{R}^d}K_d^2\left(\frac{x_k^*-u}{h_n}\right)r_2(u)\quad du\\
&\leq \frac{h_n^d}{\bar{G}(\tau_F)}\int_{\mathbb{R}^d}K_d^2(t)r_2(x_k^*-h_nt),\quad \left(t=\frac{x_k^*-u}{h_n}\right)
\end{aligned}
$$

avec $r_2(u)$ est défini dans l'hypothèse **H4,**. Par un développement de Taylor autour de x_k^*, sous les hypothèse **H2** *ii*) et **H4,** on obtient

$$\mathbb{E}\left[\frac{\delta_1^2 T_1^2}{\bar{G}^2(T_1)} K_d^2\left(\frac{x_k^* - X_1}{h_n}\right)\right] = O(h_n^d)$$

d'autre part, à partir de l'hypothèse **H3,**

$$\begin{aligned}\mathbb{E}^2\left[\frac{\delta_1 T_1}{\bar{G}(T_1)} K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right] &= \mathbb{E}^2\left[K_d\left(\frac{x_k^* - X_1}{h_n}\right)\mathbb{E}\left(\frac{\delta_1 T_1}{\bar{G}(T_1)}/X_1\right)\right] \\ &= \left[\int_{\mathbb{R}^d} K_d\left(\frac{x_k^* - u}{h_n}\right) r_1(u)dt\right]^2 \\ &= O(h_n^{2d})\end{aligned}$$

enfin

$$S^2 = O(h_n^d).$$

Puis en appliquant l'inégalité de Bernstein, on a

$$\begin{aligned}\mathbb{P}\left\{\left|\sum_{i=1}^n \Delta_i(x_k)\right| > \varepsilon\right\} &= P\left(\left|\sum_{i=1}^n U_i\right| > \varepsilon h_n^d n\right) \leq 2\exp\left\{-\frac{\varepsilon^2 h_n^d n}{2(c+\varepsilon M_1)}\right\} \\ &= : A\end{aligned}$$

donc

$$\begin{aligned}\mathbb{P}\left\{\max_{k=1,\ldots,\mathfrak{s}_n}\left|\sum_{i=1}^n \Delta_i(x_k)\right| > \varepsilon\right\} &\leq \mathfrak{s}_n A \leq M h_n^{-d\eta} A \\ &\leq M\left(nh_n^d\right)^{-\eta} n^{\eta} e^{-\frac{\varepsilon^2 nh_n^d}{2(c+\varepsilon M_1)}} \qquad (4.14) \\ &= M\left(nh_n^d\right)^{-\eta} n^{\eta - \frac{\varepsilon^2}{2(c+\varepsilon M_1)}\frac{nh_n^d}{\log n}} \qquad (4.15)\end{aligned}$$

De l'hypothèse **H1** i) , le dernier terme dans (4.15) est le terme général d'une série convergente. Puis par le lemme de Borel-Cantelli le premier terme de (4.15) tend vers zéro presque sûrement. Or si nous remplaçons ε par $\varepsilon_0\sqrt{\frac{\log n}{nh_n^d}}$ on peut choisir ε_0 de telle sorte

que le terme $n^{\eta-\frac{\varepsilon_0^2}{2(c+\varepsilon_0 M_1)}}$ soit le terme général d'une série convergente (ceci est vrai pour tout $\varepsilon_0 > M_1(\eta+1)+\sqrt{M_1^2(1+\eta)^2+2c(1+\eta)}$), et en suite

$$\sup_{x\in\mathcal{C}}|\tilde{r}_{1,n}(x)-\mathbb{E}(\tilde{r}_{1,n}(x))| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right)$$

pour I_4, on utilise les même étapes que I_2 et I_3 pour l'obtention

$$\sup_{x\in\mathcal{C}}|\ell(x)-\ell_n(x)| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right)+O(h_n)\ \ p.s$$

en fin, l'hypothèse **H4** conclue la démonstration du résultat principal.

Chapitre 5

Simulation

Sommaire

5.1 Introduction

L'objectif de ce chapitre, est de pouvoir observer et comparer le résultat des estimations de la fonction densité et de régression avec la méthode du noyau. En regardant l'influence de plusieurs paramètres tels que le nombre de données génerées (n) ainsi que la valeur choisie pour h, et le noyau K.
On va représenter les résultats obtenus pour les différents jeux de données ainsi que pour les différentes valeurs de h. En comparant visuellement les figures puis en calculant l'erreur quadratique moyenne (MSE en anglais) et l'erreur quadratique moyenne intégrée (MISE en anglais).

5.2 Influence de choix du paramètre de lissage h

La partie la plus importante dans l'estimateur à noyau est de sélectioner le paramètre h. Il existe plusieurs situation où il est satisfaisant de sélectioner ce parmètre par une représentation graphique de plusieurs densités en utilisant différentes fenêtres, et après choisir la densité la plus acceptable. L'une des stratégies pour faire cela est de commencer par une petite (où grande) fenêtre puis augmenter (ou diminuer) jusqu'à ce qu'à atteindre la plus appropriée. Et pour cela on va réaliser les simulations suivant ces étapes :

(1) générer 1000 observations de loi normale centrée réduite $X_i \sim \mathcal{N}(0,1)$ sur l'intervalle $[-4,4]$.

(2) calculer l'estimateur à noyau $\widehat{f_n}(x)$ basé sur les données X_i par le choix :

- de noyau Triangulaire
- de noyau Gaussien
- de noyau d'Epanechnikov.
- de noyau Biweight.

(3) comparer les différentes formes obtenus à la vraie densité dans les cas : $h = 0.1, h = 0.3, h = 0.5, h = 0.7$

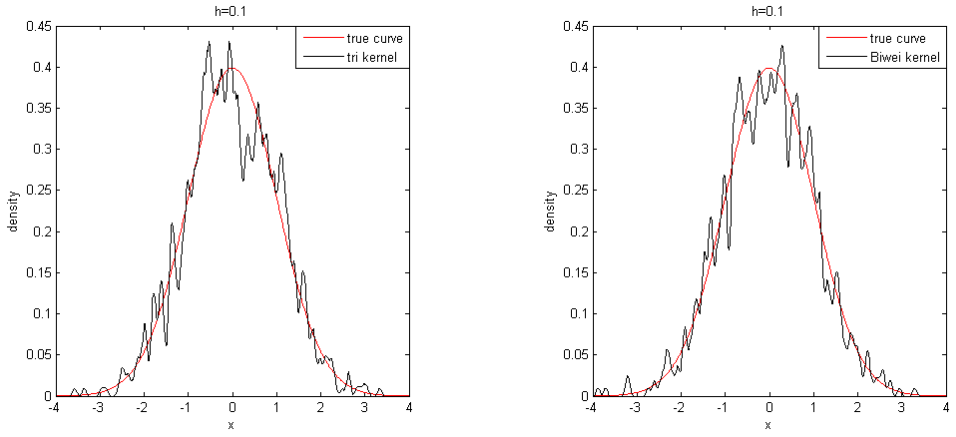

FIGURE 5.1 – Estimation de la loi $\mathcal{N}(0,1)$, avec les noyaux Triangu et Biweight $h = 0.1$

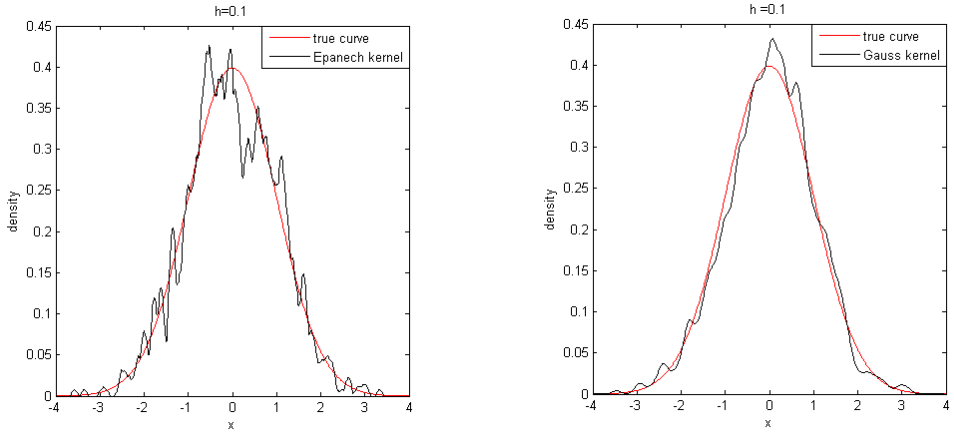

FIGURE 5.2 – Estimation de la loi $\mathcal{N}(0,1)$, avec les noyaux Epanech et Gaussien $h = 0.1$

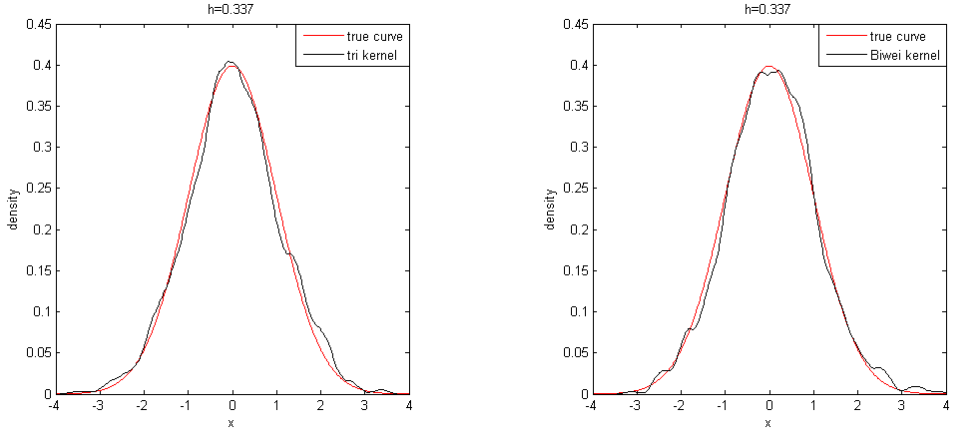

FIGURE 5.3 – Estimation de la loi $\mathcal{N}(0,1)$, avec les noyaux Triangu et Biweight $h = 0.337$

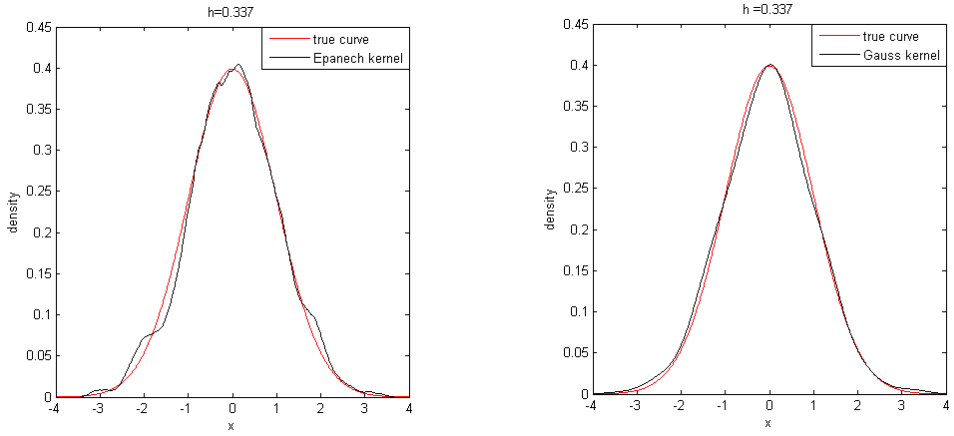

FIGURE 5.4 – Estimation de la loi $\mathcal{N}(0,1)$, avec les noyaux Epanech et Gaussien $h = 0.337$

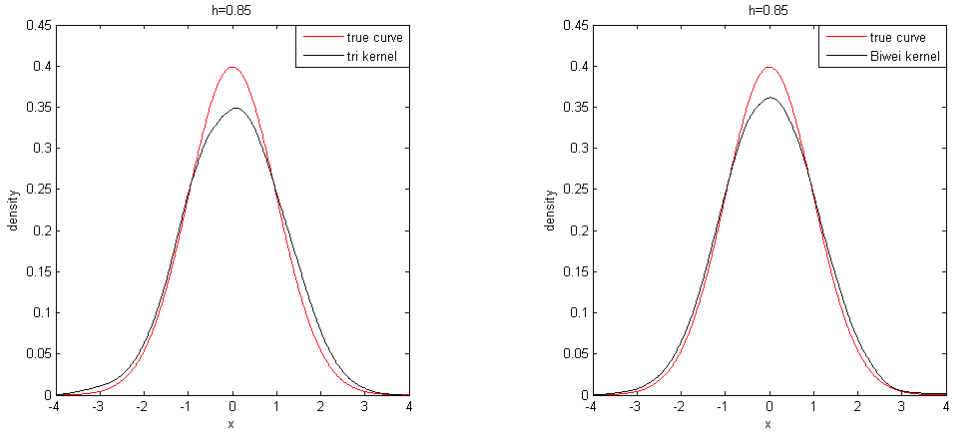

FIGURE 5.5 – Estimation de la loi $\mathcal{N}(0,1)$, avec les noyaux Triangu et Biweight $h = 0.85$

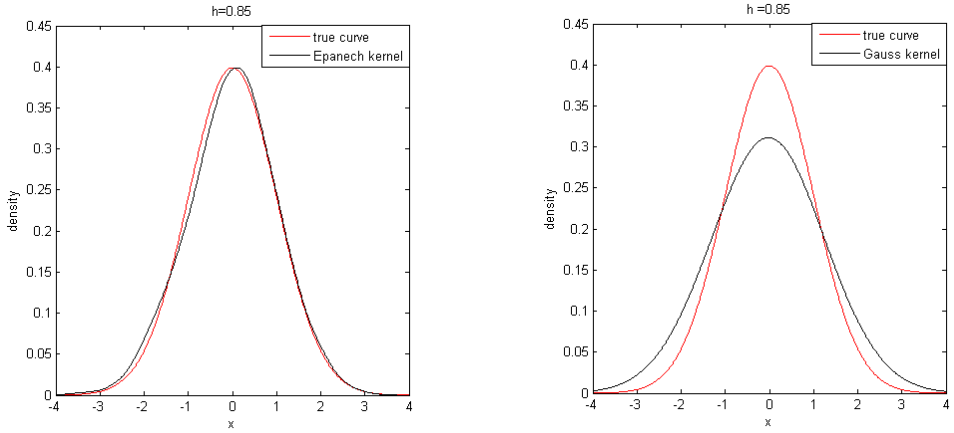

FIGURE 5.6 – Estimation de la loi $\mathcal{N}(0,1)$, avec les noyaux Epanech et Gaussien $h = 0.85$

Discussion On peut voir l'influence de la largeur de la fenêtre sur l'estimation de la densité. Dans le cas $h = 0.1$ on a une courbe sous-lissée pour tout les noyaux. Par contre, dans le cas $h = 0.85$ on remarque que l'allure de la distribution est sur-lissée et parfois aplatie. Et le meilleur résultat est obtenu pour $h = 0.337$.

Une autre façon pour étudier l'influence de choix du paramètre de lissage h d'une densité est d'utiliser différentes méthodes de sélections parmi lesquelles on a choisi :

- la méthode du AMISE.
- la validation croisés non biaisé("UCV").
- La méthode de maximum de vraisemblance avec validation croisée ("MLCV").

pour faire ces simulations on a utilisé le package **kedd**[26]. Les résultats obtenus sont présentés dans le tableau suivant :

n	Paramètre	Biwei	Epanech	Gauss
100	h_{MISE}	1.996342	1.824546	0.824385
	h_{UCV}	1.242409	1.097562	0.248211
	h_{MLCV}	1.977000	1.851000	0.548900
500	h_{MISE}	1.496872	1.391253	0.6287794
	h_{UCV}	0.862090	0.741884	0.336807
	h_{MLCV}	0.963600	1.485494	0.360300
1000	h_{MISE}	1.496997	1.314552	0.594076
	h_{UCV}	0.893814	0.745068	0.314895
	h_{MLCV}	0.879600	0.748100	0.326400
5000	h_{MISE}	1.497322	0.914885	0.413211
	h_{UCV}	0.567007	0.505596	0.204130
	h_{MLCV}	0.543700	0.491300	0.191900

TABLE 5.1 – Résultats des simulations de la loi $\mathcal{N}(0,1)$, pour déterminer le paramètre h

Le tableau (5.2) represente l'erreur moyenne quadratique (MSE) pour l'estimateur de la fonction de densité $\widehat{f}_{n,i}(x)$ correspondant aux valeurs de h précédentes du tableau (5.1). Pour calculer le MSE on reproduit 100 fois l'estimateur $\widehat{f}_n(x)$ et on calcule la médiane sur l'intervalle $[-3,3]$, qui est

$$\underset{x\in[-3,3]}{\text{med}} \frac{1}{100}\sum_{i=1}^{100}\left[\widehat{f}_{n,i}(x)-f(x)\right]^2$$

n	*Mthode*	Epanech	Biwei	Gauss
100	*AMISE*	0.0015	0.0012	0.0014
	UCV	0.0008323	0.0008384	0.0011
	MLCV	0.0016	0.0013	0.000769
500	*AMISE*	0.00062112	0.0005027	0.00060198
	UCV	0.00025668	0.000246003	0.00023652
	MLCV	0.00076421	0.00026868	0.00023818
1000	*AMISE*	0.00051874	0.00042053	0.00046272
	UCV	0.00013968	0.00014309	0.00012497
	MLCV	0.00015217	0.00015843	0.00014736
5000	*AMISE*	0.00013479	0.0003986	0.00013777
	UCV	0.00004040	0.000042723	0.000038102
	MLCV	0.000039799	0.000042902	0.000042624

TABLE 5.2 – Résultats de la médiane de MSE sur $x \in [-3,3]$

Discussion Les résultats obtenus dans tableau (5.2) montrent que les valeurs de MSE sont proches pour les différents noyaux utilisés surtout entre le noyau "Biweight" et "Epanechnikov". Les valeurs obtenues pour le noyau "Epanechnikov" sont légèrement meilleurs aux autres noyaux. La convergence du MSE vers zéro est indépendante du noyau mais dépend essentiellement de la taille n de l'échantillon.

5.3 comparaison de l'estimateur à noyau beta de la fonction de régression avec l'estimateur à noyau gaussien dans le cas censurée

L'objectif de cette section, est de comparer par simulation, pour $d = 1$, l'impact des noyaux symétriques et non symétriques sur le biais aux bornes dans l'estimateur de la fonction de régression. On considère deux types de relations entre les Y_i et X_i : l'une linéaire et l'autre non linéaire. Dans chaque cas on a organisées ces simulations selon les étapes suivantes :

(1) générer deux suites indépendantes X_i et ε_i de v.a.i.i.d de taille n qui suivent la loi $\mathcal{N}(0,1)$ c'est-à-dire $X_i \sim \mathcal{N}(0,1)$, $\varepsilon_i \sim \mathcal{N}(0,1)$

(2) simuler un n-échantillon de v.a réelle de censure $C_i \sim \mathcal{E}(\lambda)$

(3) calculer Y_i dans les deux cas linéaire et non linéaire, puis prendre $T_i = \min(Y_i, C_i)$ et $\delta_i = \mathbb{I}_{\{Y_i \leq C_i\}}$.

(4) calculer l'estimateur $m_n(x)$ basé sur les données simulées $(X_i, T_i, \delta_i), i = 1, n$

Modèle (1) : on commence par le modèle linéaire : $Y_i = 2\,X_i - 1 + 0.2\,\varepsilon_i$ en choisissant :

– le noyau Gaussien

$$K_g\left(\frac{x - X_i}{h_n}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\left(\frac{x - X_i}{h_n}\right)^2 / 2\right)$$

– le noyau béta

$$K_{x/h_n,(1-x)/h_n}(X_i) = beta(x/h_n, (1-x)/h_n)(X_i)$$

où $beta(p,q)(u) = \frac{1}{\beta(p,q)} u^p (1-u)^{q-1}, \quad u \in [0,1],\ \beta(n,p) = \int_0^1 t^{n-1}(1-t)^{p-1} dt$

et on compare les courbes obtenues dans les deux cas avec la vraie courbe qui est $m(x) = \mathbb{E}(Y/X = x) = 2x - 1$ pour des taux de censure égaux au $10\%, 20\%,$et 30% et pour une taille d'échantillon $n = 100, n = 500$.

Les graphes suivants montrent le comportement de l'estimateur de la fonction de régression dans le cas où le pourcentage de censure augmente.

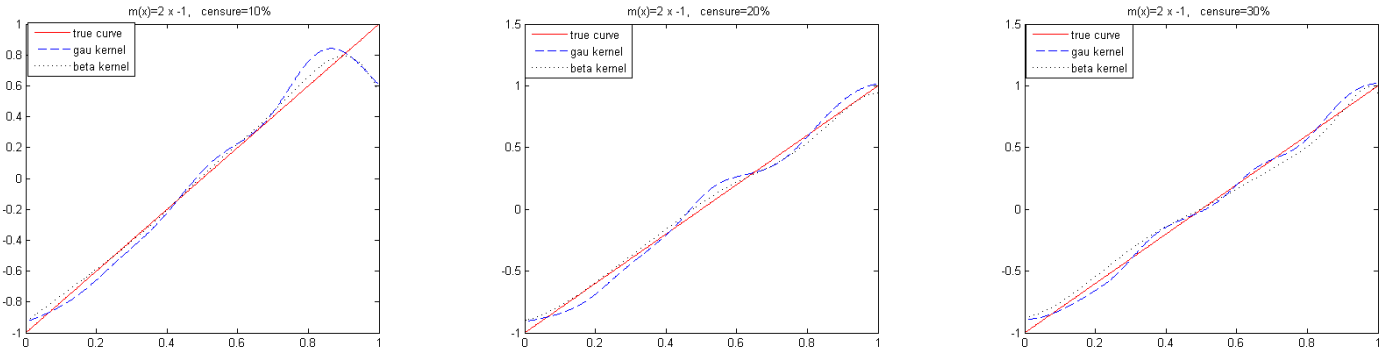

FIGURE 5.7 – $m(x) = 2\,x - 1$, pour.censure= 10%, 20%, et 30% $n = 100$

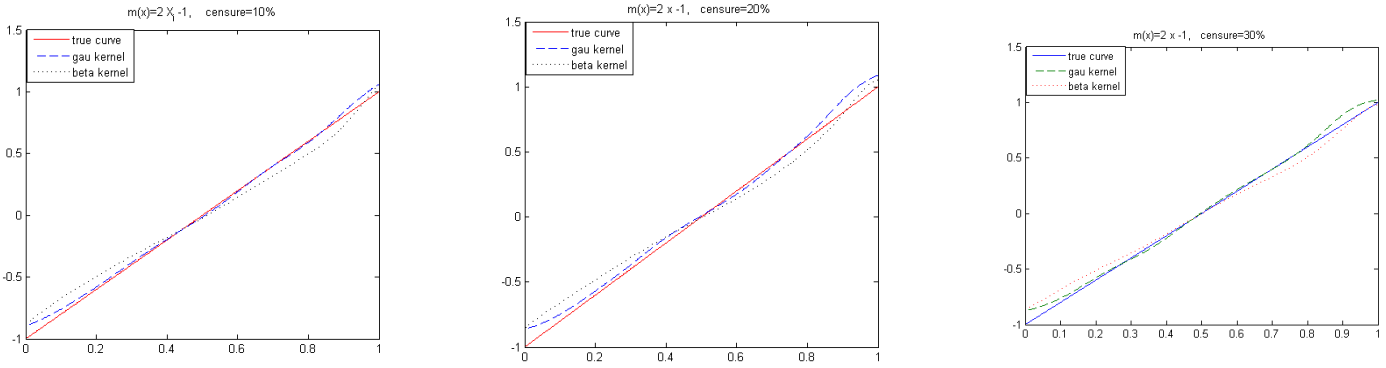

FIGURE 5.8 – $m(x) = 2\,x - 1$, pour.censure= 10%, 20%, et 30% $n = 500$

On remarque que le comportement du noyau bêta et pour un taux de censure faible est meilleur sur les bornes, ce qui est conforté par le tableau suivant :

pourcentage de censure		$n = 50$		$n = 100$		$n = 500$	
		$x = 0$	$x = 1$	$x = 0$	$x = 1$	$x = 0$	$x = 1$
10%	MSE_{Gauss}	1.1603	0.2403	1.3743	0.1149	1.3197	0.0323
	MSE_{Beta}	1.1115	0.2200	1.3245	0.1134	1.2532	0.0350
20%	MSE_{Gauss}	1.3091	0.3133	1.2774	0.1465	1.2426	0.0527
	MSE_{Beta}	1.2474	0.2787	1.2035	0.1338	1.1789	0.0536
30%	MSE_{Gauss}	1.1543	0.9877	0.9510	0.3730	1.3199	0.1014
	MSE_{Beta}	1.0788	0.7691	0.9484	0.3394	1.2498	0.1080

TABLE 5.3 – MSE aux bornes : $x = 0$ et $x = 1$, pour le modèle $m(x) = 2\,x - 1$

Tableau (5.3), représente l'erreur moyenne quadratique de l'estimateur par rapport à la valeur théorique. Pour chacune des valeurs de n on reproduit 100 fois l'estimateur $m_n(x)$ et on calcule $\frac{1}{100}\sum_{i=1}^{100}[m_{n,i}(x) - m(x)]^2$; et on donne le MSE aux bornes, c'est-à-dire, au point $x = 0$ et $x = 1$.
Nous donnons aussi , dans le tableau (5.4), le MSE en prenant la médiane sur $x \in [0,\ 1]$.

on note, dans ce cas, que le noyaux symétrique (gaussien) a un comportement légèrement meilleur en rapport avec une vitesse de convergence plus rapide dans le sens que le taux de convergence du noyau symétrique (voir théorème (3.1)) tend vers zéro plus rapidement que pour les noyaux non symétriques (théorème (4.1), ce qui est visible en particulier pour les petites valeurs n.

pourcentage de censure		$n = 50$	$n = 100$	$n = 500$
10%	MSE_{Gauss}	0.0252	0.0136	0.0129
	MSE_{Beta}	0.0359	0.0185	0.0127
20%	MSE_{Gauss}	0.0254	0.0133	0.0119
	MSE_{Beta}	0.0444	0.0242	0.0129
30%	MSE_{Gauss}	0.0534	0.0321	0.0089
	MSE_{Beta}	0.0610	0.0551	0.0221

TABLE 5.4 – la médiane de MSE sur $x \in [0,1]$, pour le modèle $m(x) = 2\,x - 1$

Ces deux tableaux montre l'influence du pourcentage de censure sur la qualité de l'estimateur qui apparait clairement dans le cas de la forte censure où elle devient un peu moins bonne que lorsqu'on a un pourcentage moins élevé de censure.
On termine par un calcules d'erreur quadratique moyenne intégré $MISE = \int_{\mathbb{R}} MSE(x)\ dx$ sur l'intervalle $[0,1]$ et on représente les résultats comme suit :

pourcentage de censure		$n = 50$	$n = 100$	$n = 500$
10%	$MISE_{Gauss}$	0.0489	0.0432	0.0421
	$MISE_{Beta}$	0.0457	0.0421	0.0414
20%	$MISE_{Gauss}$	0.0620	0.0500	0.0428
	$MISE_{Beta}$	0.0532	0.0466	0.0419
30%	$MISE_{Gauss}$	0.0658	0.0569	0.0474
	$MISE_{Beta}$	0.0548	0.0510	0.0452

TABLE 5.5 – l'erreur quadratique moyenne intégré (MISE) sur $x \in [0,1]$, $m(x) = 2\,x - 1$

On remarque dans le tableau (5.5) qu'il ya très peut de différence entre les deux noyaux, on peut voir aussi que plus le pourcentage de censure augment plus la (MISE) augmente.

Modèle (2) : cas de régression non linéaire en choisissant le modèle suivant :

$$Y_i = \sin(6\,X_i) + 0.2\,\varepsilon_i$$

qui est le cas du sinus, avec les taux de censure égaux au $10\%, 20\%,$et 30% et pour une taille d'échantillon $n = 100, n = 500$.

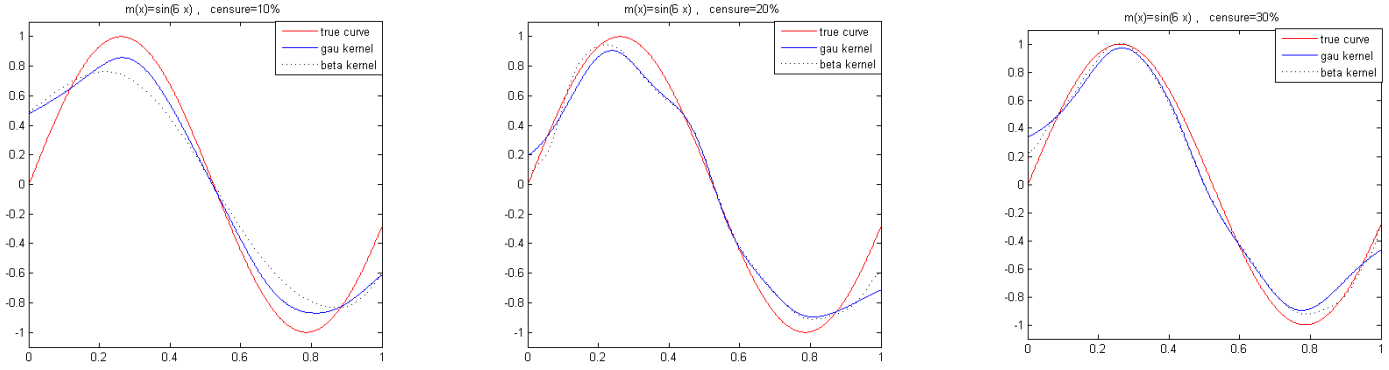

FIGURE 5.9 – $m(x) = \sin(6\,x)$, pour.censure= 10%, 20%,et 30% $n = 100$

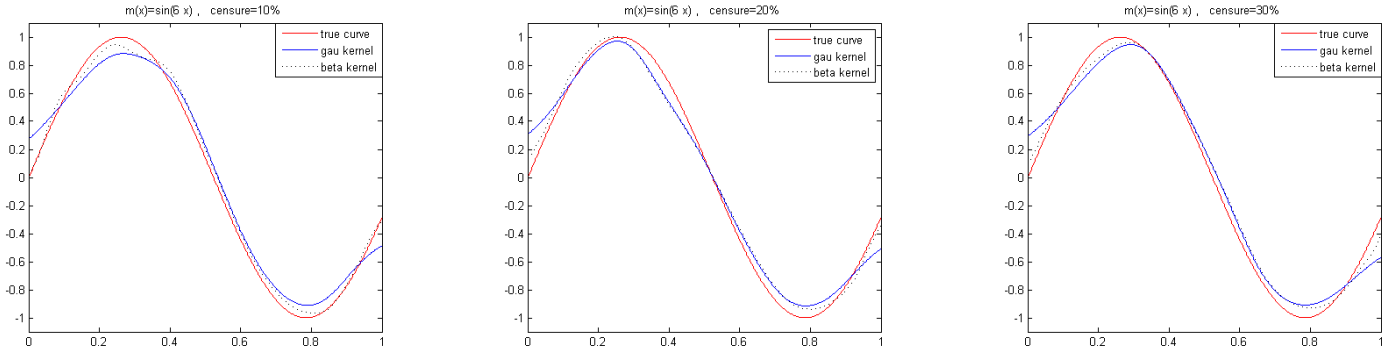

FIGURE 5.10 – $m(x) = \sin(6\,x)$, pour.censure= 10%, 20%,et 30% $n = 500$

pourcentage de censure		$n = 50$		$n = 100$		$n = 500$	
		$x = 0$	$x = 1$	$x = 0$	$x = 1$	$x = 0$	$x = 1$
10%	MSE_{Gauss}	0.4495	0.1449	0.4492	0.1328	0.4470	0.1261
	MSE_{Beta}	0.3928	0.1434	0.3817	0.1280	0.3779	0.1209
20%	MSE_{Gauss}	0.4662	0.1546	0.4802	0.1341	0.4628	0.1240
	MSE_{Beta}	0.3625	0.1540	0.4003	0.1314	0.3899	0.1186
30%	MSE_{Gauss}	0.3036	0.1383	0.4038	0.1265	0.4177	0.1209
	MSE_{Beta}	0.2485	0.1367	0.3345	0.1237	0.3505	0.1149

TABLE 5.6 – MSE aux bornes : $x = 0$ et $x = 1$, pour le modèle $m(x) = \sin(6\,x)$

pourcentage de censure		$n = 50$	$n = 100$	$n = 500$
10%	MSE_{Gauss}	0.0477	0.0392	0.0322
	MSE_{Beta}	0.1027	0.0908	0.0872
20%	MSE_{Gauss}	0.0519	0.0397	0.0340
	MSE_{Beta}	0.1004	0.0917	0.0866
30%	MSE_{Gauss}	0.0613	0.0452	0.0338
	MSE_{Beta}	0.0954	0.0954	0.0906

TABLE 5.7 – la médiane de MSE sur $x \in [0,1]$, pour le modèle $m(x) = \sin(6\,x)$

pourcentage de censure		$n = 50$	$n = 100$	$n = 500$
10%	MSE_{Gauss}	0.0099	0.0089	0.0083
	MSE_{Beta}	0.0132	0.0121	0.0115
20%	MSE_{Gauss}	0.0108	0.0090	0.0083
	MSE_{Beta}	0.0137	0.0121	0.0117
30%	MSE_{Gauss}	0.0128	0.0109	0.0085
	MSE_{Beta}	0.0154	0.0138	0.0118

TABLE 5.8 – l'erreur quadratique moyenne intégré (MISE) sur $x \in [0,1],\ m(x) = \sin(6\,x)$

Une autre fois encore, on remarque un meilleur comportement du noyau bêta aux bornes. En conclusion, bien que le comportement de l'estimateur à noyau non symétrique est plus proche de la vraie fonction de régression aux bornes, l'estimateur à noyau symétrique reste meilleur - au sens de l'erreur quadratique moyenne intégrée *MISE*- dans tout l'intervalle de l'étude.

Conclusion

La progression dans les calculs ainsi que les moyens rendant facile leur réalisation (i.e outils informatiques) disponibles actuellement pour le statisticien, ont eu un impact significatif dans la recherche statistique et spécialement dans les procédures d'analyse des données non paramétriques. En particulier, la recherche théorique et appliquée sur l'estimation de densité non paramétrique a eu une influence remarquable sur des sujets pertinents, tels que la regression non paramétrique.

Nous avons dans ce mémoire essayé de faire une synthèse des résultats existant concernant l'estimateur non paramétrique de la densité et la fonction de régression dans le cas complet et censurées. Nous avons donné quelques estimateurs de la densité dans le cas complet comme l'histogramme, l'histogramme mobile, et l'estimateur à noyau (l'estimateur de *Parzen-Rosenbllat*) pour lequel nous avons donné la vitesse de convergence au sens du *MISE* pour l'estimateur à noyau qui est plus rapide que pour l'histogramme, étant d'ordre $n^{-4/5}$ au lieu de $n^{-2/3}$, et nous avons mis en évidence non seulement le rôle du paramètre de lissage h_n, mais aussi le choix du noyau K. Nous avons effectué un rappel sur trois méthodes de sélection du paramètre h_n (le plug-in, "UCV", "MLCV"). Par la suite nous avons abordé l'estimateur non paramétrique de la fonction de régression en insistant sur l'estimateur de *Nadaraya-Watson*.

Dans le cas des données censurées, nous avons rappelé l'estimateur de la densité introduit par *Blum* et *Susarla* (1980) et l'estimateur de la régression qui avait suscité l'intérêt d'un certain nombre de chercheurs à savoir *Carbonez* et *al* (1985), *Kohler* et *al* (2002), *Guessoum* et *Ould-Said*(2009), ces dernier ont étudié la convergence uniforme presque sure et ils ont obtenu une vitesse de convergence de l'ordre de $O\left(\max\left\{\sqrt{\frac{\log n}{nh_n}}, h_n^2\right\}\right)$, dont nous avons apporté notre contribution, en utilisant les même technique de démonstration de l'estimateur de la fonction de régressin pour des données censurées pour une covariable X dans $\mathbb{R}^d$ et dans le cas où l'hypothèse de symétrie du noyau n'est pas remplie et nous avons trouvé $O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) + O\left(h_n^2\right)$. Pour conforter notre résultat nous avons fait des comparaisons par des simulations entre un noyau symétrique (gaussien) et un noyau non symétrique (bêta).

Cette étude nous a permis de tracer quelques lignes perspectives, tel que le problème d'estimation non paramétrique de la fonction de régression dans le cas où les données présentent une forme de dépendance, comme le cas de l'α-mélange et d'établir d'autres types de données incomplètes.

Annexe

L'objectif de *l'annexe* est de donner une idée sur la programmation mathématique sous langage Matlab utilisée dans ce mémoire.
La fonction `fr` permet de simuler la fonction de répartition dans un intervalle $[a,b]$ avec un pas h sur la base d'un échantillon simulé X.

Code .1

```
function[]=fr(n,a,b,h)     ♯ a,b les bornes d'intervalle

X=randn(1,n)     ♯ la loi normal

Xord=sort(X)

x=a:h:b

F(1)=0

for i=1:length(x)

for k=1:(n-1)     ♯ n la taille de l'échantillon

if x(i)< Xord(1)

F(i)=0

elseif (x(i)>=Xord(k)) (x(i)< Xord(k+1))

F(i)=k/n

elseif x(i)> Xord(k+1)

F(i)=1

end

end

end

♯ Representation graphique

plot(x,F)

xlabel('x')

legend('Fonction de répartiton empirique')

end

fr(7,-3,3,0.01)
```

La fonction `rec` permet de simuler le noyau rectangulaire.

Code .2

```
function[]=rec(h,l,a,b)     ♯ a,b les bornes d'intervalle
u=-l:h:l
n=length(u)
for i=1:n
if(u(i)>=a)(u(i)<=b)
K(i)=1/(b-a)
else
K(i)=0
end
end
♯ Representation graphique
plot(u,K)
xlabel('u')
ylabel('K(u)')
legend('noyau de Rosenblatt')
end
rec(0.01,1.5,-1,1)
```

La fonction `tri` permet de simuler le noyau triangulaire.

Code .3

```
function[]=tri(h,l,a,b)     ♯ a,b les bornes d'intervalle
u=-l:h:l
n=length(u)
for i=1:n
if(u(i)>=a)(u(i)<=b)
K(i) = 1-abs(u(i))
else
```

```
K(i) =0
end
end
♯ Representation graphique
plot(u,K)
xlabel('u')
ylabel('K(u)')
legend('noyau Triangulaire')
end
tri(0.01,1.5,-1,1)
```

La fonction `epan` permet de simuler le noyau d'Epanechnikov.

Code .4

```
function[]=epan(h,l,a,b)     ♯ a,b les bornes d'intervalle
u=-l:h:l
n=length(u)
for i=1:n
if(u(i)>=a)(u(i)<=b)
K(i)=(3/4)*(1-(u(i))^2)
else
K(i) =0
end
end
♯ Representation graphique
plot(u,K)
xlabel('u')
ylabel('K(u)')
legend('noyau d'Epanechnikov')
end
epan(0.01,1.5,-1,1)
```

La fonction `biw` permet de simuler le noyau biwieght.

Code .5

```
function[]=biw(h,l,a,b)     ♯ a,b les bornes d'intervalle
u=-l:h:l
n=length(u)
for i=1:n
if(u(i)>=a)(u(i)<=b)
K(i) =(15/16)*(1-u(i)^2)^2
else
K(i) =0
end
end
♯ Representation graphique
plot(u,K)
xlabel('u')
ylabel('K(u)')
legend('noyau Biwieght')
end
epan(0.01,1.5,-1,1)
```

La fonction `gau` permet de simuler le noyau gaussien.

Code .6

```
function[]=gau(h,a,b)     ♯ a,b les bornes d'intervalle
u=a:h:b
n=length(u)
for i=1:n
if(u(i)>=a)(u(i)<=b)
K(i) =(1/sqrt(2*pi))*exp((-1/2)*u(i)^2)
else
```

```
K(i) =0
end
end
♯ Representation graphique
plot(u,K)
xlabel('u')
ylabel('K(u)')
legend('noyau Gaussien')
end
gau(0.01,-4,4)
```

La fonction `noyaubeta` permet de simulé le noyau bêta de paramètre $(\frac{x}{h_n}+1)$ et $(\frac{(1-x)}{h_n}+1)$.

Code .7

```
function[]=noyaubeta(x,h)      ♯ x : la position, h : la fenêtre
p=(x/h)+1
q=((1-x)/h)+1
u(1)=0
for i=1:501
u(i)=u(1)+(i-1)*0.002;
Beta(i)=betapdf(u(i),p,q);
end
♯ Representation graphique
plot(u,Beta)
xlabel('t')
ylabel('K(t)')
end
noyaubeta(0.5,0.2)
```

La fonction `kerdens` permet de simulé l'estimateur de *Parzen-Rosenblatt* .

Code .8

```
function[]=kerdens(n,h)     ♯ n  le nombre d'observation ,h  la fenêtre
X=randn(n,1);
u=-3;
for j=1:601
f=0;
for i=1:n
f=f+(1/sqrt(2*pi))*exp (-(1/2)*((X(i)-u)/h)^2); ♯ estimateur à noyau gaussien
end
g(j)=(f/(n*h));
z(j)=(1/sqrt(2*pi))*exp((-0.5)*u^2) ♯ la vraie densité
u=u+0.01;
end
u=-3:0.01:3;
plot(u,z,'r',u,g,'k');
♯ Representation graphique
xlabel('x')
ylabel('fn(x)')
legend('true curve','estimate');
end
kerdens(100,0.3)
```

La fonction `mse` permet de simulé le *MSE*

Code .9

```
function[]=mse(n,h,m)
for s=1:m
rand('state',sum(100*clock))
X=randn(n,1);
```

```
u=-3;
for j=1:601
f=0;
for i=1:n
f=f+(1/sqrt(2*pi))*exp(-(1/2)*((X(i)-u)/h)^2); ♯ estimateur à noyau gaussien
end
g(j)=(f/(n*h));
z(j)=(1/sqrt(2*pi))*exp((-0.5)*u^2) ♯ la vraie densité
g(s,j)=g(j);
err(s,j)=(g(s,j)-z(j))^2;
u=u+0.01;
end
end
vecmse=mean(err);
mse=median(vecmse)
end
mes(100,0.3,100)
```

Code .10

```
library("kedd") ♯ charger le package "kedd"
x<-rnorm(1000,mean=0,sd=1) ♯ loi normale(0,1)
h.amise(x, deriv.order = 0, kernel = "gaussian")
h.ucv(x, deriv.order = 0, kernel = "gaussian")
h.mlcv(x, deriv.order = 0, kernel = "gaussian")
h.amise(x, deriv.order = 0, kernel = "epanechnikov")
h.ucv(x, deriv.order = 0, kernel = "epanechnikov")
h.mlcv(x, deriv.order = 0, kernel = "epanechnikov")
h.amise(x, deriv.order = 0, kernel = "biweight")
h.ucv(x, deriv.order = 0, kernel = "biweight")
h.mlcv(x, deriv.order = 0, kernel = "biweight")
```

Bibliographie

[1] Beran, R. (1981). *Nonparametric regression with randomly censored survival data*, Technical Report university of California, Berkeley.

[2] Bierens, H. J. (1987). *Kernel estimators of regression function*. Advances in Econometrics, Cambridge Univ. Press.

[3] Bouezmarni T. and Rolin J-M. (2003). *Consistency of the beta kernel density function estimator.* Canad. J. Statist., 31 , No. 1, 89-98.

[4] Bowman, A. W. (1984) *An alternative method of crooss-validation for the smoothing density estimates estimator.* Biometrika, Vol. 71, pp. 353-360.

[5] Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, J. C. (1983). *Tree structured methods for classification andregression*. Chapman et Hall.

[6] Carbon M,Francq C (2006). *Estimation non paramétrique de la densité et de la regression-Prévision non paramétrique* Labora.Proba.Statist,France.

[7] Carbonez, A., Györfi, L. and Van der Meulin, E.C. (1995). *Partition-estimate of a regression function under random censoring*. Statist. Decisions, 1321-37.

[8] Chen S,X (1999). *Beta kernel estimators for density functions.* Comput Statist. Data Anal,31 (1999) 131-145.

[9] Chen S,X (2000). *Beta kernel smoothers for regression curves.* Statist.Sinica 10 (2000) 73-91.

[10] Cover, T.M. and Hart, P.E. (1967). *Nearest neighbor pattern classification*. IEEE Transactions of Information Theory. 13, 21-27.

[11] Dabrowska, D.M. (1987). *Nonparametric regression with censored survival data. Scand. J. Statist.14 :181–197.*

[12] Dabrowska, D.M. (1989). *Uniform consistency of the kernel conditionnal Kaplan-Meier estimate*. Ann. Statistics, 17 :1157–1167.

[13] Dalalyan A.S. *Note de cours statistique avencée : Méthodes non-paramétrique Ecole Centrale de Paris*.

[14] P. Deheuvel and J. H.J. Einmahl. *Functional limit laws for the increments of Kaplan-Meier productlimit processes and applications*. *Ann. Probab.*, 28, 1301-1335,2000.

[15] Devroye, L.P. (1978). *The uniform convergence of the Nadaraya-Watson regression function estimate.* The Canadian J. of Statistics. 6.2, 179-191.

[16] Devroye, L.P. and Wagner, T. J. (1980). *Distribution free consistency results in nonparametric discrimination and regression function estimation*. Ann. Statist. 8, 231-239.

[17] Diehl, S. and Stute, W. (1988) *Kernel Density and Hazard Function Estimation in the Presence of Censoring. Journal of Multivariate Analysis, Vol. 25, pp. 299-310.*

[18] Duin, R. P. W. (1976). *On the choice of smoothing parameters of Parzen estimators of probability density function IEEE Transactions on Computers, C-25, 11751179.*

[19] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis : Theory and Practice. Springer*.

[20] Ferraty,F., Mass,A. and Vieu, P. (2007). *Nonparametric regression of functional data :Inference and practical aspects.* Aust. N. Z. J. Stat. 49.3 267-286.

[21] Gășser, T. and Műller, H. G. (1979). *Kernel estimation of regression function. In : Smoothing Techniques for Curve Estimation*, Lecture Notes in Math., 757, 23-68. Springer-Verlag, Berlin.

[22] Gordon, L. and Olshen, R.A.(1980). *Consistent nonparametric regerssion from recursive partionning schemes*. J.Multivariate Anal. 10, 611-627.

[23] Guessoum Z (2009). *Regression non paramétrique dans les modèle censurés*.Thèse de doctorat USTHB.

[24] Guessoum, Z. Ghettab, S. (2013) *Beta kernel regression estimator : Some comparisons with symmetric kernel. ASMDA(2013) Mataro Espagne.* .

[25] Guessoum, Z. and Ould-Said, E. (2009) *On non-paramétric estimation of régression function under random censorship model. Statist. Decisions*.

[26] Guidoum, A.C (2013) *kedd : A Package for Simulation of Kernel Estimator and Bandwidth Selection for Density and its Derivatives in R* R package version 1.0. http ://cran.r-project.org/package=kedd.

[27] Habbema, J. D. F., Hermans, J., and Van den Broek, K. (1974). *A stepwise discrimination analysis program using density estimation*. *Compstat 1974 : Proceedings in Computational Statistics. Physica Verlag, Vienna*.

[28] Huber C. *Modèles pour des durées de survie*.
http :// www.biomedicale.univ-paris5.fr/survie/enseign/survie sansi.pdf

[29] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Wiley, New York.

[30] Kaplan, E. L. and Meier, P. (1958). *Nonparametric estimation from incomplete observations.* J. Amer. Statist. Assoc., 53, 457-481.

[31] Köhler, M., Màthé, K. and Pintér, M. (2002). *Prediction from randomly Right Censored Data.text J. Multivariate Anal., 80, 73-100.*

[32] Lejeune M (2010). *Statistique La théorie et ses applications*. Springer-Verlag France,Paris.

[33] Loquin K , Strauss O (2006). *Fuzzy Histograms and Density Estimation.* Advances in Soft Computing 6, 45-52 (2006). Springer-Verlag Berlin Heidelberg.

[34] Matias C (2012) *Introduction à l'estimation non paramétrique. Note de cours Université de Évry*.

[35] Nelson, W. (1972). *Theory and applications of hazard plotting for censored failure data.* Technometrics 14 945-966.

[36] Nadaraya, E. A (1964). On estimating regression. Theor. Probab. Appl. 9, 141-142.

[37] Parzen E (1962). *on Estimation of a Probability Density Functions and Mode.* Annals. Mathe. Statist.,Vol 33 ,Issue 3 (Sep,1962), 1065-1076.

[38] Padgett, W. J., and MC Nichols, D. T. (1984). *Nonparametric density estimation from censored data. Comm. Statist. A-Theory Methods 13 1581-1611*.

[39] R Development Core Team (2011). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http ://www.R-project.org/.

[40] R Development Core Team (2011). *Writing R extensions*. version 2.13.1 (2011-07-08), ISBN 3-900051-11-9.

[41] Rosenblatt (1956). *Remarks on Some Nonparametric Estimates of Density Functions.* Annals. Mathe. Statist.,Vol 27 , (Sep,1956), 832-837.

[42] Rudemo, M. (1982) *Empirical choice of histogram and kernel density estimators.* Scandinavian Journal of Statistics, Vol. 9, No. 2, pp. 65-78.

[43] Scoutt D.W(1979). *On Optimal and Data-Based Histograms.* Biometrika, Vol 66, No , (Dec,1979),pp 605-610.

[44] Silverman B.W, (1986). *Density Estimation for statistics data analysis,Chapman and Hall,1986*. ISBN 0-412-24620-1.

[45] Tsybakov A.B (2004). *Introduction à l'estimation non paramétrique*. Springer-Verlag ,New York-Berlin. ISBN 3-540-40592-5.

[46] Zhang B (1996). *Some Asymptotic Results for Kernel Density Estimation under Random Censorship.*Author(s) : Biao Source : *Bernoulli, Vol. 2, No. 2 (Jun., 1996), pp. 183-198.*

Glossaire

- $i.i.d$: indépendent identiquement distribuées
- $f.d.r$: fonction de répartition
- MSE : Mean Squard Errore (l'erreur moyenne quadratique)
- $MISE$: Mean Integred Squard Errore (l'erreur moyenne quadratique intégrée)
- $\mathbb{P}$: la mesure de probabilité attachée à l'espace probabiliste $(\Omega, \mathcal{A})$
- $(\Omega, \mathcal{A}, \mathbb{P})$: l'espace probabilité où les ariables aléatoire sont considérées
- $\mathbb{E}(X)$: la variance de la variable aléatoire X
- $\mathbb{V}ar(X)$: la variance de la variable aléatoire X
- $:=$ la définition d'une quantité
- $^{\mathcal{L}}$: la convergence en loi
- ^{p} : la convergence en probabilité
- $^{p.s}$: la convergence presque sûe
- $^{m.q}$: la convergence en moyenne quadratique
- $\mathbb{I}_A$: la fonction indicatrice qui vaut 1 sur l'ensemble A et 0 ailleurs
- $a_n = O(b_n), n \to \infty$: signifie que $\limsup_{n\to\infty} |a_n/b_n|$ avec a_n et b_n deux suite réelles
- $a_n = o(b_n), n \to \infty$: signifie que $\lim_{n\to\infty} a_n/b_n = 0$ avec a_n et b_n deux suite réelles