
UNIVERSITE Dr. MOULAY TAHAR DE SAÏDA
FACULTE DES SCIENCES
DEPARTEMENT DE MATHEMATIQUES

Mémoire de master

Estimation non-paramétrique de la densité par histogramme généralisé

KADA Smail

REMERCIEMENT

En préambule à ce mémoire, nous souhaitons adresser nos remerciements les plus sincères aux personnes qui ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire.

Nous souhaitons remercier en premier lieu l'encadreur de ce mémoire, *M^r KADI Mokhtar*, pour son aide précieuse et pour le temps qu'il a bien voulu nous consacrer.

Pour son attention de tout instant sur nos travaux, pour ses conseils avisés et son écoute qui ont été prépondérants pour la bonne réussite de ce mémoire. Son énergie et sa confiance ont été des éléments moteurs pour nous. Nous avons pris un grand plaisir à travailler avec lui.

Nous exprimons nos gratitude à toutes les personnes rencontrées lors des recherches que nous avons effectuées et qui ont accepté de répondre à nos interrogations avec gentillesse.

Enfin, nous remercions nos parents, frères et soeurs, et nos familles pour leur soutien au cours de ces deux années et sans lesquels nous ne sommes pas là aujourd'hui, ainsi que tous nos proches et amis qui ont aidé et toujours soutenu et encouragé au cours de la réalisation de ce mémoire.

Merci à tous et à toutes
KADA Smail

Table des matières

| | |
|---|-----------|
| Introduction | 5 |
| 1 Généralités sur les histogrammes | 7 |
| 1.1 Généralités | 7 |
| 1.1.1 Histogramme | 7 |
| 1.1.2 Relation entre densité de probabilité et histogramme des données | 9 |
| 1.1.3 Estimateur de densité par histogramme | 10 |
| 1.1.4 Propriétés statistiques de l'estimateur par histogramme | 11 |
| 1.1.5 Utilisation de l'estimateur de densité par histogrammes | 13 |
| 1.1.6 Histogrammes de la loi $\mathcal{U}(0, 1)$ | 16 |
| 1.1.7 Histogrammes de la loi $\mathcal{N}(0, 1)$ | 18 |
| 1.2 Estimateur à noyau | 21 |
| 1.2.1 Définition des estimateurs à noyau | 21 |
| 1.2.2 Propriétés statistiques de l'estimateur à noyau | 24 |
| 1.2.3 Un résultat théorique sur la convergence des estima- teurs à noyau | 26 |
| 2 Estimation non paramétrique de la densité par histogramme généralisé | 27 |
| 2.1 Estimation de la densité | 27 |
| 2.1.1 Les histogrammes ou estimateurs à noyaux | 27 |
| 2.1.2 Les estimations par fonctions orthogonales | 28 |
| 2.1.3 Les méthodes dites du maximum de vraisemblance | 29 |

| | | |
|----------|--|-----------|
| 2.1.4 | Les méthodes mixtes, les histogrammes lissés | 30 |
| 2.2 | Estimations par histogrammes uniformes | 32 |
| 2.2.1 | Moments des estimations | 34 |
| 2.2.2 | Moyenne quadratique intégrée | 37 |
| 2.2.3 | Développements pour la loi normale $\mathcal{N}(0,1)$, et les noyaux symétriques | 40 |
| 2.2.4 | Développement par les noyaux de Gram-Charlier | 42 |
| 2.2.5 | Développements par les noyaux de Legendre | 43 |
| 2.2.6 | Développement par les noyaux de Laguerre | 45 |
| 2.2.7 | Comportement asymptotique des estimations, critère du M.I.S.E | 47 |
| 3 | Estimation fonctionnelle | 49 |
| 3.1 | Données fonctionnelles | 50 |
| 3.2 | Estimation de la fonction de répartition | 51 |
| 3.2.1 | Estimation ponctuelle | 51 |
| 3.2.2 | Intervalle de confiance | 54 |
| 4 | Exemples sur les développements des noyaux | 55 |
| 4.1 | Développements pour les noyaux symétriques | 55 |
| 4.2 | Développements pour les noyaux de Gram-charlier | 58 |
| 4.3 | Développements pour les noyaux de Legendre | 58 |
| | Bibliographie | 62 |

Introduction

L'estimation non-paramétrique de la densité a fait l'objet de multiples travaux par des méthodes diverses dans les années passées. Par la suite, l'avènement de la statistique moderne a permis à cette dernière une évolution importante, surtout l'estimation non paramétrique par la méthode du noyau qui a été intensément traitée par plusieurs auteurs.

Rosenblatt (1956) [18] donne l'expression de l'erreur quadratique moyenne de l'estimateur de la densité basé sur le noyau uniforme dans le cas des variables indépendantes et identiquement distribuées.

Parzen (1962) [14] établit la normalité asymptotique de cet estimateur.

L'estimateur à noyau dépend de deux paramètres, à savoir le noyau, fonction déterminant la forme du voisinage contenant les observations en étude, et la fenêtre de lissage qui contrôle la taille de ce voisinage.

Le choix de la fenêtre est crucial pour la performance de l'estimation. La fenêtre optimale relativement à cette perte dépend généralement d'une quantité inconnue. Pour contourner ce problème délicat qu'est le choix de la largeur de la fenêtre.

Dans ce travail, nous nous intéressons essentiellement à l'estimation non-paramétrique par histogramme généralisé qui est le plus ancien des estimateurs non-paramétriques de densité. L'origine des histogrammes est attribuée à John Graunt au XVII^e siècle répondant à l'objectif d'une représentation de la distribution de données. À ce titre, il peut être considéré comme un estimateur de la densité de probabilité sous-jacente à un ensemble fini d'observations.

Ce travail est divisé en quatre chapitres. Dans le premier chapitre, nous rappelons quelques définitions sur l'histogramme et de ses propriétés, et, quelques notions sur l'estimateur à noyau.

Dans le deuxième chapitre, on présente la méthode de l'histogramme généralisé pour le cas réelle, ensuite dans le chapitre suivant, on présente l'estimation dans le cas fonctionnel.

Enfin, dans le dernier chapitre, on présente quelques exemples de cette méthode d'estimation ; on étudie le comportement des estimations appliquées aux densités usuelles, et particulièrement à la distribution normale ; on obtient ainsi les valeurs optimales de la suite $\{\delta_n\}$ quand définir après pour les noyaux de Parzen-Rosenblatt, pour les petits échantillons.

Chapitre 1

Généralités sur les histogrammes

1.1 Généralités

1.1.1 Histogramme

Construire un histogramme à partir d'un ensemble d'observations (précises) (x_1, \dots, x_n) consiste à partitionner l'intervalle de référence $\Omega = [e_{min}, e_{max}]$ en $p \in \mathbb{N}$ cellule A_k , $k \in \{1, \dots, p\}$, et à compter le nombre d'observations appartenant à chaque cellule A_k . Si toutes les cellules de l'histogramme ont la même largeur, on dit que l'histogramme est régulier. On note $\Delta \in \mathbb{R}^+$, la largeur des cellules qui est alors appelée le pas ou la granularité de l'histogramme. La valeur de Δ est donnée par :

$$\Delta = \frac{e_{max} - e_{min}}{p}. \quad (1.1)$$

Le nombre d'observations appartenant à chaque cellule A_k est appelé accumulateur de la cellule A_k est noté Acc_k . Il est défini par :

$$Acc_k = \sum_{i=1}^n \mathbb{1}_{A_k}(x_i) \quad (1.2)$$

où $\mathbb{1}_{A_k}(x)$ est la fonction indicatrice de l'ensemble A_k définie par :

$$\mathbb{1}_{A_k}(x) = \begin{cases} 1 & \text{si } x \in A_k \\ 0 & \text{sinon} \end{cases} \quad (1.3)$$

La Figure 1.1 présente un histogramme de 100 observations tirées aléatoirement d'une loi normale centrée réduite $\mathcal{N}(0, 1)$. Ces observations sont réparties sur un intervalle de référence $\Omega = [-5, 5]$. La granularité de l'histogramme est fixée à $\Delta = 0.8$.

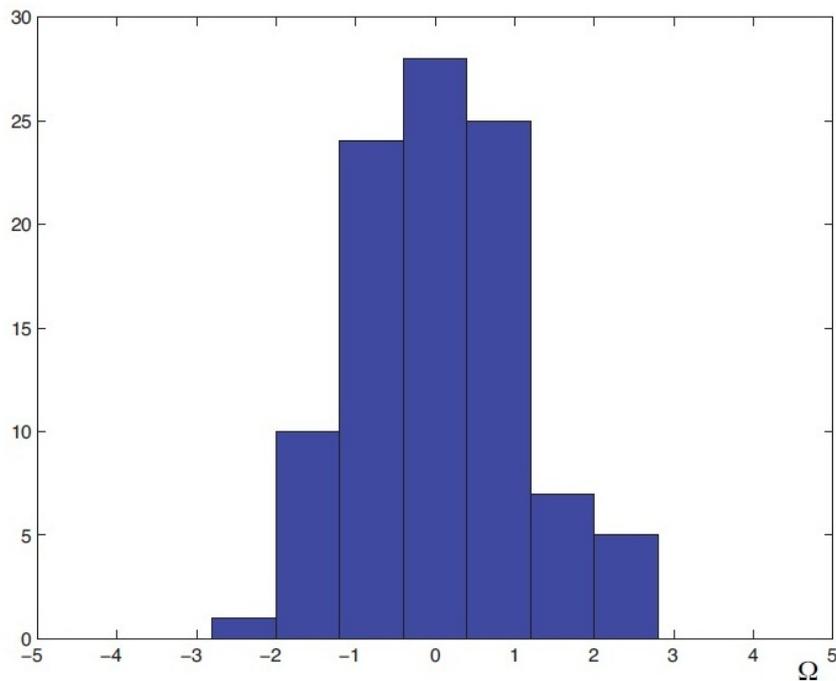


Figure 1.1 : Histogramme de 100 observations issues d'une loi normale $\mathcal{N}(0, 1)$ construit sur une partition de granularité $\Delta = 0.8$ d'un intervalle de référence $\Omega = [-5, 5]$.

1.1.2 Relation entre densité de probabilité et histogramme des données

L'histogramme et densité de probabilité sont liés par des conditions aux limites : une densité de probabilité peut être vue comme la limite d'un histogramme lorsque le nombre d'observations tend vers l'infini et que la granularité de l'histogramme tend vers zéro. La Figure 1.2 illustre, cette relation, en considérant le même processus d'observations qu'en section 1.1 avec un grand nombre d'observations ($n = 100000$) et une plus faible granularité ($\Delta = 0.2$). En comparant la Figure 1.1 avec la Figure 1.2, on voit clairement que l'allure de cet histogramme se rapproche de l'allure de la densité de probabilité ($\mathcal{N}(0, 1)$) quand le nombre d'observations n augmente et la granularité Δ diminue.

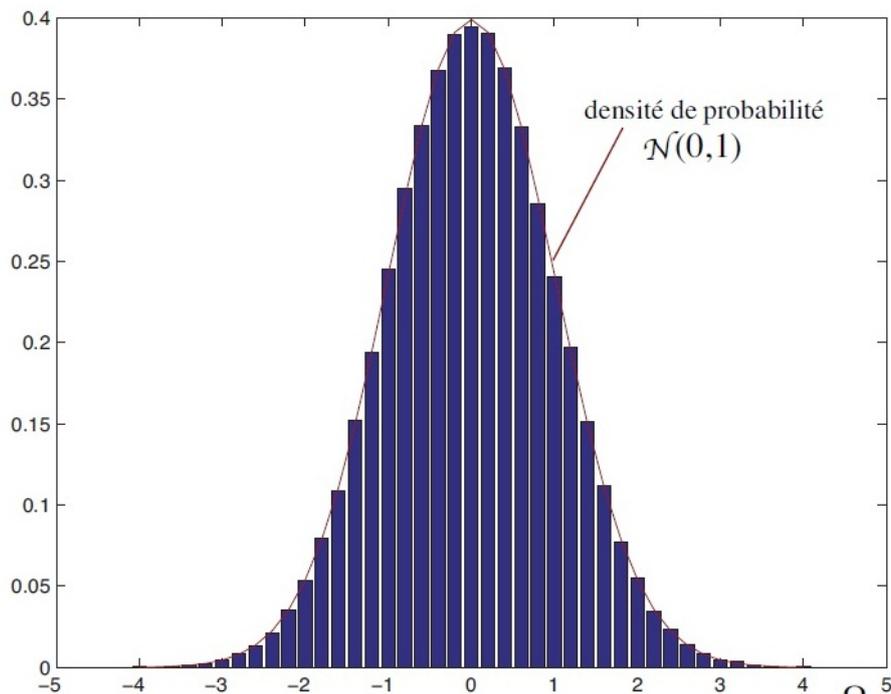


Figure 1.2 : Histogramme de 100000 observations construit avec une granularité $\Delta = 0.2$.

1.1.3 Estimateur de densité par histogramme

Soit un ensemble fini d'observations (x_1, \dots, x_n) de n variables aléatoires (X_1, \dots, X_n) indépendantes (au sens de probabilité) et identiquement distribuées (de même loi de probabilité) (i.i.d) de densité de probabilité commune f . Pour tout $k \in \{1, \dots, p\}$, soit Acc_k l'accumulateur associé à la cellule A_k , défini par l'expression (1.2), la probabilité de A_k (basée sur les observations x_i), notée $P(A_k)$, est donnée par :

$$P(A_k) = \frac{Acc_k}{n}. \quad (1.4)$$

En émettant l'hypothèse, généralement irréaliste, que les observations se répartissent uniformément dans la cellule A_k (de largeur Δ), on peut alors construire un estimateur de la densité f , pour tout $x \in \Omega$, par :

$$\hat{f}_\Delta(x) = \frac{1}{\Delta} \sum_{k=1}^p P(A_k) \mathbf{1}_{A_k}(x). \quad (1.5)$$

Cet estimateur peut aussi s'écrire

$$\hat{f}_\Delta(x) = \frac{1}{n\Delta} \sum_{k=1}^p Acc_k \mathbf{1}_{A_k}(x). \quad (1.6)$$

Dans la suite, nous émettons l'hypothèse que les cellules $A_k, \forall k \in \{1, \dots, p\}$ forment une partition de Ω (c'est-à-dire $\Omega = \bigcup_{k=1}^p A_k$ et $\forall i, j \in \mathbb{N} : i \neq j, A_i \cap A_j \neq \phi$) et définissons pour chaque cellule A_k , son centre a_k telles que $\forall k \in \{1, \dots, p\}, A_k = [a_k - \frac{\Delta}{2}, a_k + \frac{\Delta}{2}]$ et $\forall k \in \{1, \dots, p-1\}, a_{k+1} = a_k + \Delta$.

1.1.4 Propriétés statistiques de l'estimateur par histogramme

Nous présentons, dans cette partie, les propriétés statistiques de l'estimateur par histogramme \hat{f}_Δ défini par l'expression 1.6¹. En statistiques, il est nécessaire de mesurer la qualité d'un estimateur. Pour cela, on évalue, d'une part, l'écart entre la moyenne de l'estimateur et la densité à estimer, ce critère d'évaluation est appelé *biais*, et d'autre part, la variance de l'estimateur (due au caractère aléatoire d'observations) qui caractérise la dispersion des valeurs de l'estimateur dans l'ensemble d'observations. On essaye généralement de réduire au mieux ces deux quantités. Le biais de l'estimateur \hat{f}_Δ 1.6 est donné pour tout $x \in (a_k, a_{k+1}]$, $\forall k \in \{1, \dots, p\}$, par :

$$\begin{aligned} \text{biais}(\hat{f}_\Delta(x)) &= \mathbb{E}(\hat{f}_\Delta(x)) - f(x) \\ &= \frac{1}{2}f'(\Delta - 2(x - a_k)) + o(\Delta^2), \end{aligned} \tag{1.7}$$

où o est un terme résiduel et f' est la dérivée de f . f' doit être une fonction de $L(\Omega)$ absolument continue et carrée intégrable .

La variance de l'estimateur est donnée , pour tout $x \in \Omega$, par :

$$\begin{aligned} \text{var}(\hat{f}_\Delta(x)) &= \mathbb{E} \left((\hat{f}_\Delta(x))^2 \right) - \left(\mathbb{E}(\hat{f}_\Delta(x)) \right)^2 \\ &= \frac{f(x)}{n\Delta} + o(n^{-1}). \end{aligned} \tag{1.8}$$

Cette variance tend vers zéro quand le produit $n\Delta$ tend vers l'infini quand le nombre d'observation n tend vers l'infini.

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

1. (pour une étude plus détaillée voir, par exemple, les livres de Bosq et Lecoutre [3] et de Simonoff [20]).

1.1.4 Propriétés statistiques de l'estimateur par histogramme 12

La convergence en *moyenne quadratique* de l'estimateur \hat{f}_Δ^n 1.6 a été établie par Lecoutre [11]. Nous avons pour tout $x \in (a_k, a_{k+1}]$:

$$\begin{aligned} MSE(\hat{f}_\Delta(x)) &= var(\hat{f}_\Delta(x)) + biais^2(\hat{f}_\Delta(x)) \\ &= \frac{f(x)}{n\Delta} + \frac{f'(x)^2}{4}(\Delta - 2(x - a_k))^2 + o(n^{-1}) + o(\Delta^3). \end{aligned} \quad (1.9)$$

Cette erreur quadratique moyenne tend vers zéro quand Δ tend vers zéro et $n\Delta$ tend vers l'infini quand n tend vers l'infini. Ce critère d'erreur quadratique moyenne est un critère local. On lui préfère généralement un critère plus global obtenu en l'intégrant sur tout le domaine Ω . Ce critère porte le nom d'erreur quadratique moyenne intégrée (MISE).

La convergence en *moyenne quadratique intégrée* de l'estimateur \hat{f}_Δ^n défini par 1.6 a été établie par Lecoutre [11]. Nous avons :

$$MISE = \frac{1}{n\Delta} + \frac{\Delta^2 R(f')}{12} + o(n^{-1}) + o(\Delta^3), \quad (1.10)$$

où $R(s) = \int_\Omega (s(x))^2 dx$, pour toute fonction s carrée intégrable.

Cette erreur quadratique moyenne intégrée tend vers zéro quand Δ tend vers zéro et $n\Delta$ tend vers l'infini quand n tend vers l'infini.

L'utilisation du critère MISE permet de définir une *granularité optimale* de l'histogramme notée Δ^* . Cette valeur optimale est la valeur qui minimise ce critère pour un nombre d'observations et une loi donnée. Cette valeur optimale est de la forme :

$$\Delta^* = \left[\frac{6}{R(f')} \right]^{\frac{1}{3}} n^{-\frac{1}{3}} \quad (1.11)$$

Notons que cette valeur est inutilisable en pratique car elle fait intervenir une connaissance a priori de densité inconnue f via l'intégrale du carré de sa dérivée ($R(f')$). Puisque, justement, cette technique est dédiée à l'estimation de densités dont la loi est inconnue, la valeur de Δ^* est généralement

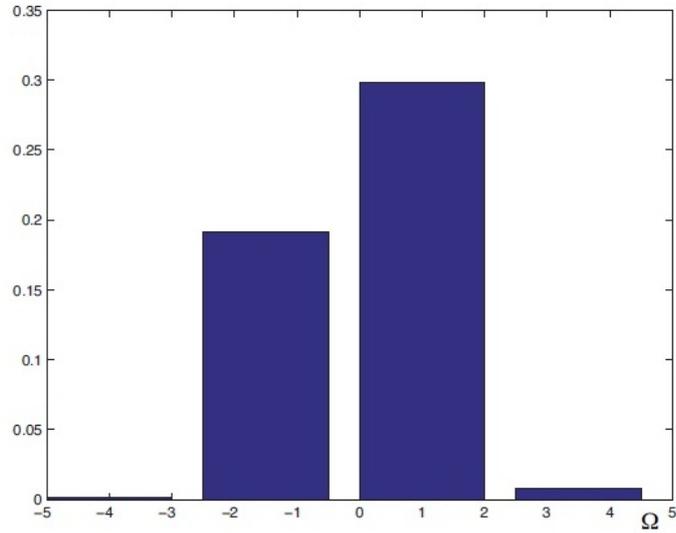
inconnue. En remplaçant, la valeur de Δ^* 1.11 dans l'expression 1.10, on obtient la valeur optimale de l'*erreur quadratique moyenne intégrée asymptotique* (AMISE) notée $AMISE^*$:

$$AMISE^* = \left[\frac{9R(f')}{16} \right]^{\frac{1}{3}} n^{-\frac{2}{3}} \quad (1.12)$$

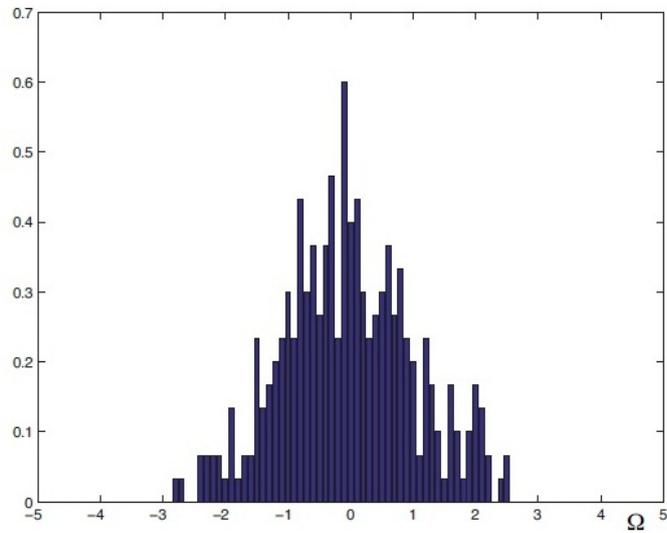
Geffroy a montré dans [8] la convergence *uniforme et presque complète* de l'estimateur \hat{f}_Δ 1.6. Tandis que la *convergence en probabilité* et presque complète en norme $L^1(\Omega)$ a été établie par Abou-Jaoudé [1].

1.1.5 Utilisation de l'estimateur de densité par histogrammes

Utiliser un histogramme pour estimer la densité présente l'intérêt de la simplicité. Cette simplicité a cependant une contre partie désavantageuse. la représentation par histogramme de la densité sous-jacente à un ensemble fini d'observations n'est robuste au choix ni de la granularité ni de l'intervalle de référence de la partition sur laquelle l'histogramme est bâti. Nous illustrons ce manque de robustesse sur un exemple de 300 observations tirées aléatoirement d'une distribution normale $\mathcal{N}(0, 1)$. Sur la Figure 1.3, nous avons choisi deux valeurs de largeur de cellule : $\Delta = 2.5$ (Figure 1.3 (a)) et $\Delta = 0.1$ (Figure 1.3 (b)). On peut, par exemple, s'intéresser à la localisation du mode principal (c'est-à-dire la valeur la plus fréquemment prise par ces observations) sur l'histogramme. Pour $\Delta = 2.5$, la localisation du mode est claire dans l'intervalle $[0, 2]$. Tandis que pour $\Delta = 0.1$, la localisation de ce mode est dans l'intervalle $[-0.14, -0.06]$.



(a) $\Delta = 2.5$



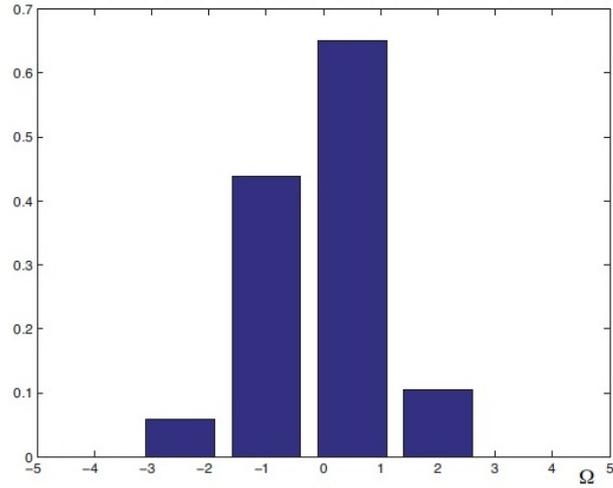
(b) $\Delta = 0.1$

Figure 1.3 : Histogramme de 300 observations construit avec une granularité $\Delta = 2.5$ (a) et $\Delta = 0.1$ (b).

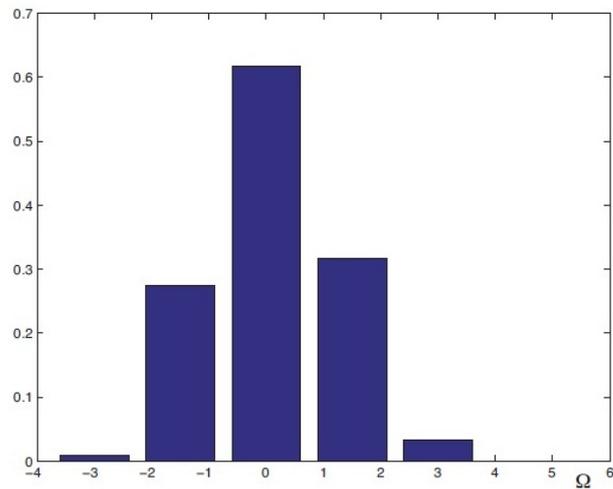
Sur la Figure 1.4, nous avons choisi deux positions différentes de l'intervalle de référence Ω : $\Omega = [-5, 5]$ (Figure 1.4 (a)) et $\Omega = [-4, 6]$ (Figure 1.4 (b)). La

1.1.5 Utilisation de l'estimateur de densité par histogrammes 15

Figure 1.4 montre l'influence du positionnement de Ω , sur la représentation d'observations. Sur la Figure 1.4 (b), la distribution semble symétrique alors qu'elle semble ne pas l'être sur la Figure 1.4 (a).



(a) $\Omega = [-5, 5]$



(b) $\Omega = [-4, 6]$

Figure 1.4 : Histogramme de 300 observations construit sur la portion $\Omega = [-5, 5]$ (a) et $\Omega = [-4, 6]$ (b).

1.1.6 Histogrammes de la loi $\mathcal{U}(0, 1)$

On se place ici dans le cas où l'ensemble \mathcal{F} des densités possibles est réduit à la densité $f = \mathbb{1}_{[0,1]}$ de la loi $\mathcal{U}(0, 1)$. Evidemment, dans ce cas, il n'est pas utile d'estimer la densité f puisqu'elle est connue ! Néanmoins, cette étude exploratoire nous montrera divers défauts et qualités de l'histogramme, d'où son intérêt. Nous partons donc d'un échantillon $(x_n, n \geq 1)$ de la loi $\mathcal{U}(0, 1)$. Quelle que soit la taille n de l'échantillon (x_1, \dots, x_n) , les nombres x_1, \dots, x_n appartiennent à l'intervalle $]0, 1[$ et sont tous différents parce que la probabilité qu'il n'en soit pas ainsi est nulle.

L'intervalle $[0, 1]$ est divisé en 50 classes : $[0, 0.02], \dots,]0.98, 1]$.

Ces classes sont notées $I_j, j = 1, \dots, N$ dans les formules ci-dessous. l désigne leur longueur commune ($l = 0.02$). n et N ayant été choisis, on rappelle que l'histogramme correspondant est constant sur chacune des classes I_j où il vaut

$$\bar{f}_{n,N,j} = \frac{1}{nl} \sum_{i=1}^n \mathbb{1}_{I_j}(x_i) \quad (1.13)$$

(il s'agit de l'effectif de la classe I_j divisés par le produit nl). L'histogramme est la fonction de \mathbb{R} dans \mathbb{R}^+ définie par :

$$x \mapsto \bar{f}_{n,N}(x) = \sum_{j=1}^N \bar{f}_{n,N,j} \cdot \mathbb{1}_{I_j}(x), \quad (1.14)$$

ou, si l'on n'écrit pas la variable x

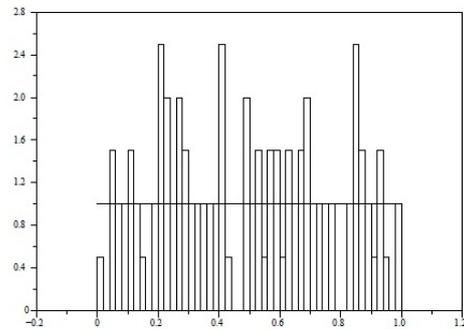
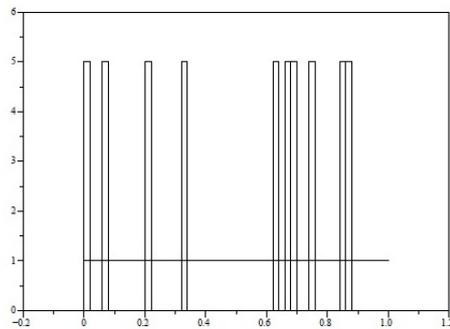
$$\bar{f}_{n,N} = \sum_{j=1}^N \bar{f}_{n,N,j} \mathbb{1}_{I_j} \quad (1.15)$$

En fait, les notations 1.13, 1.14 et 1.15 sont trop imprécises car elles ne montrent pas que l'histogramme est calculé à partir de l'échantillon (x_1, \dots, x_n) ,

autrement dit, qu'il dépend du hasard. En fait, $\bar{f}_{n,N}(x)$ aurait dû être noté

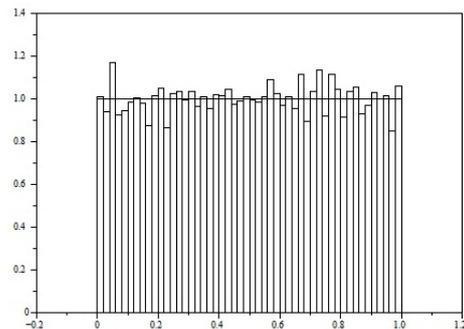
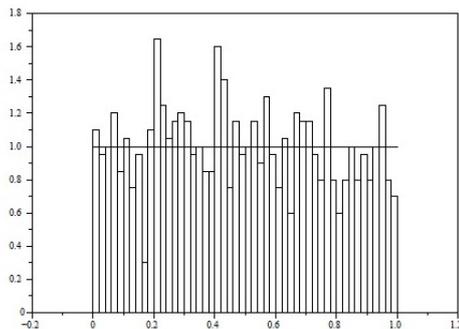
$$\bar{f}_{n,N}(x_1, \dots, x_n; x) = \sum_{j=1}^N \frac{1}{nl} \left(\sum_{i=1}^n \mathbb{1}_{I_j}(x_i) \right) \cdot \mathbb{1}_{I_j}(x) \quad (1.16)$$

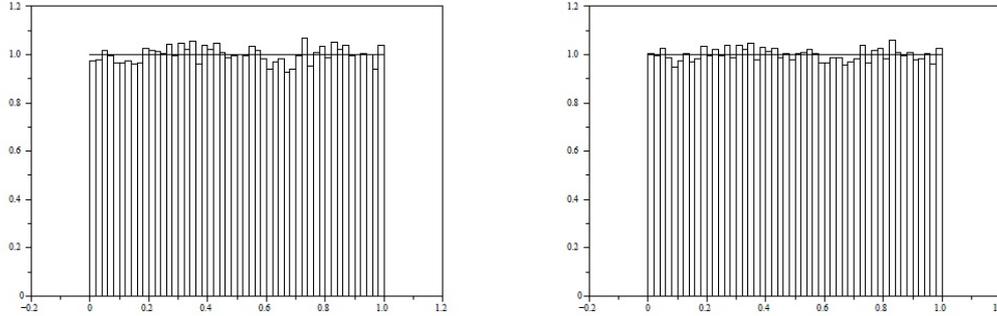
Voici maintenant les graphes des 6 histogrammes annoncés, complétés par le graphe de f .



Premier graphe : $nl = 10 \times 0.02 = 0.2$. Dans une classe qui contient 1 point, la valeur de l'histogramme est $\frac{1}{0.2} = 5$. Il y a 10 classes de cette sorte ci-dessus.

Deuxième graphe : $nl = 100 \times 0.02 = 2$. Il y a des classes contenant 0,1,2,3,4 et 5 points.



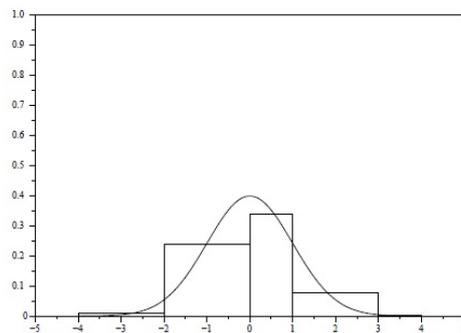
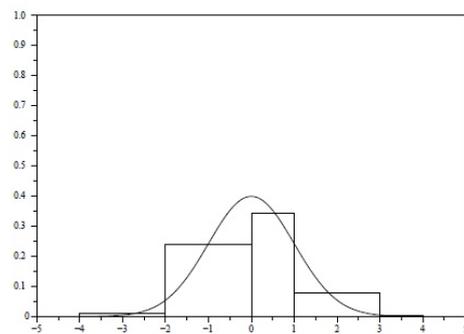
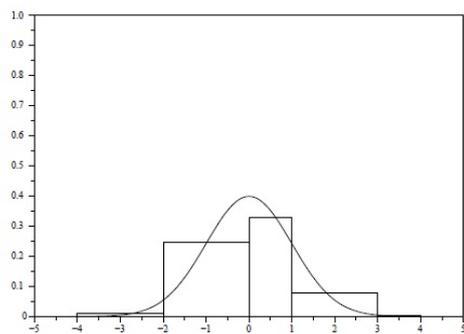
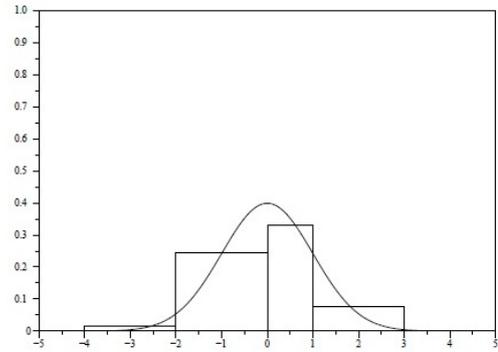
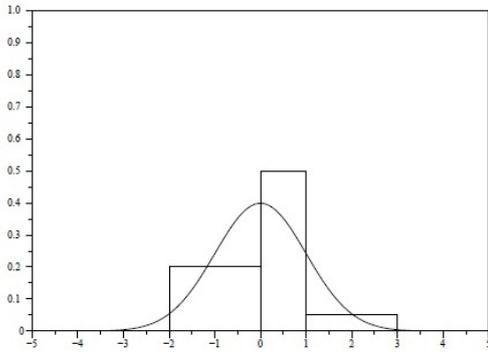


On constate que ces histogrammes se rapprochent de plus en plus du graphe de f au sens de la convergence uniforme quand la taille n de l'échantillon grandit. La convergence semble lente.

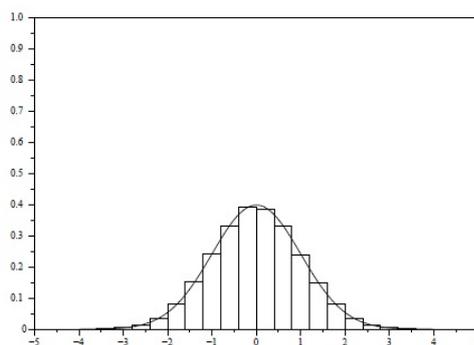
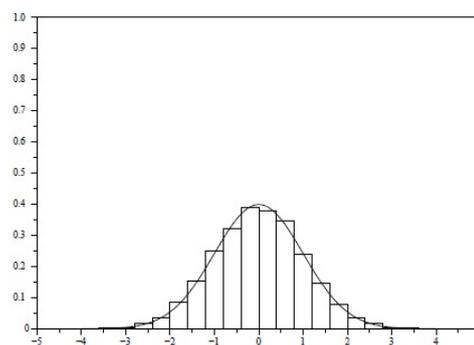
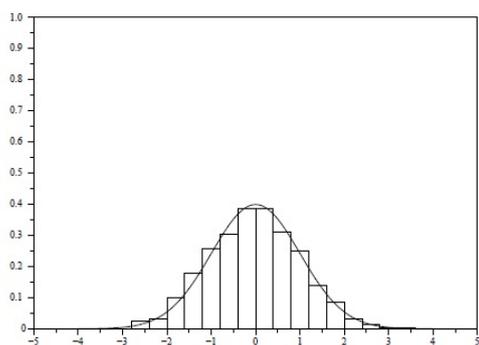
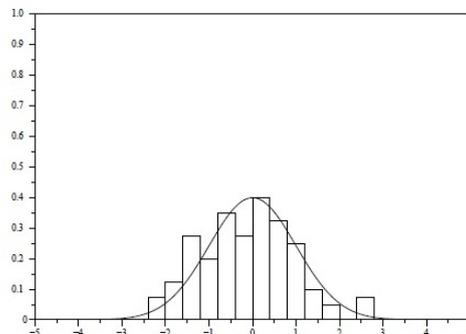
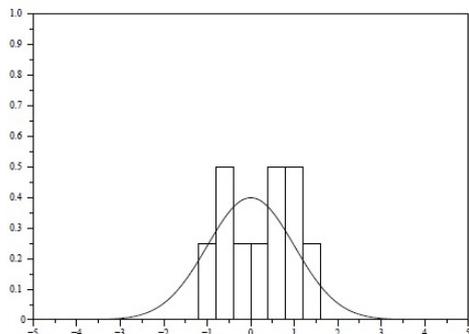
1.1.7 Histogrammes de la loi $\mathcal{N}(0, 1)$

Nous supposons maintenant que l'ensemble \mathcal{F} des densités possibles est réduit à la densité $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ de la loi normale centrée réduite, toujours dans un but exploratoire, le cas précédemment étudié de la loi $\mathcal{U}[0, 1]$, trop simple, masquant les problèmes de convergence des histogrammes. Nous utilisons donc dans cette division des échantillons de $\mathcal{N}(0, 1)$ et traçons le graphe de l'histogramme ainsi que le graphe de la densité de cette loi en donnant successivement à la taille de l'échantillon les valeurs 10, 100, 1000, 10000 et 100000, l'histogramme étant construit sur l'intervalle $[-4, 4]$ partagé en $[-4, -2]$, $[-2, 0]$, $[0, 1]$, $[1, 3]$ et $[3, 4]$ dans le premier cas, en 20 intervalles de même longueur dans le second.

premier partage de $[-4, 4]$



Second partage de $[-4, 4]$



On constate qu'il n'y a apparemment aucun espoir que l'histogramme converge vers f . En fait, les deux suites d'histogrammes semblent avoir une limite qui

est une fonction en escalier. Effectivement, il est facile de démontrer, de nouveau à l'aide de la loi forte des grands nombres, que l'histogramme converge uniformément sur $[-4, 4]$, au sens de la convergence presque sûre, vers la fonction en escalier

$$x \mapsto \varphi(x) = \sum_{j=1}^N \frac{\alpha_j}{l_j} \mathbb{1}_{I_j}(x)$$

où (I_1, \dots, I_N) désigne le partage en classes de l'intervalle $[-4, 4]$ et où, pour $j = 1, \dots, N$, α_j désigne la probabilité pour qu'une variable aléatoire X suivant la loi $\mathcal{N}(0, 1)$ prenne ses valeurs dans l'intervalle I_j , dont la longueur est notée l_j , soit

$$\forall j \in \{1, \dots, N\}, \alpha_j = P(x \in I_j) = \int_{I_j} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

On remarquera que φ n'est pas une densité de probabilité, son intégrale valant

$$\sum_{j=1}^N \alpha_j = P(X \in [-4, 4]) < 1$$

1.2 Estimateur à noyau

1.2.1 Définition des estimateurs à noyau

Supposons que l'on ne dispose que d'un échantillon observé de taille 1, soit (x_1) (sous l'hypothèse qu'il provient d'une loi à densité f inconnue). Si K désigne une densité de probabilité qui a un maximum en 0, la fonction

$$x \mapsto K(x - x_1),$$

qui est encore une densité de probabilité, admet un maximum en x_1 . Comme

il est raisonnable de penser que la densité inconnue f prend de grandes valeurs (c'est à dire des valeurs voisines de son maximum) au voisinage de x_1 puisque x_1 a été tiré au hasard suivant la loi de densité f , on pourra décider de choisir $x \mapsto K(x - x_1)$ comme estimateur de f . En effet, la loi de densité $x \mapsto K(x - x_1)$ a au moins la propriété de rendre les plus probables les valeurs autour de x_1 . Cette raison n'est pas très convaincante, mais il ne faut pas oublier que pour choisir cet estimateur de f , on n'a pratiquement pas d'information ! On peut de plus donner de la souplesse à l'estimateur en lui adjoignant un paramètre supplémentaire $h > 0$ qui s'appelle la *largeur de la fenêtre* et en choisissant comme nouvel estimateur de f la fonction un peu plus sophistiquée suivante, qui est toujours une densité de probabilité :

$$x \mapsto \frac{1}{h} \cdot K\left(\frac{x - x_1}{h}\right) \quad (1.17)$$

Le choix de h permet d'accorder plus ou moins d'importance à l'observation : en effet, plus h est petit, plus le maximum de l'estimateur est grand. On peut alors continuer dans cet esprit : si l'échantillon observé est de taille 2, soit (x_1, x_2) , et comme l'on n'a aucune raison d'accorder plus d'importance à x_1 qu'à x_2 , on choisira comme estimateur de f la fonction

$$x \mapsto \frac{1}{2h} \cdot \left(K\left(\frac{x - x_1}{h}\right) + K\left(\frac{x - x_2}{h}\right) \right)$$

et plus généralement, on associera à l'échantillon observé (x_1, \dots, x_n) , de taille n , l'estimateur

$$x \mapsto \hat{f}_n(x) = \frac{1}{nh_n} \left(K\left(\frac{x - x_1}{h_n}\right) + \dots + K\left(\frac{x - x_n}{h_n}\right) \right), \quad (1.18)$$

qui est encore une densité de probabilité . La formule (1.18) permet de surcroît de faire varier la largeur de la fenêtre avec n , parce qu'elle a été notée h_n et non h .

est appelé *estimateur à noyau* de la densité inconnue f . On voit qu'il dépend

uniquement des choix du noyau K et de la largeur de la fenêtre h_n et qu'il se calcule à partir de l'échantillon observé (x_1, \dots, x_n) .

Définition 1.2.1. (Noyau sommatif) Un noyau sommatif est une fonction $K : \Omega \rightarrow \mathbb{R}^+$, qui vérifie la propriété de sommativité, c'est-à-dire :

$$\int_{\Omega} K(u) du = 1. \quad (1.19)$$

On peut définir, à partir de noyau sommatif K , un noyau sommatif K_{Δ}^x traduit en $x \in \Omega$ et dilaté d'une largeur de bande (ou paramètre de lissage) $\Delta > 0$, par :

$$\forall u \in \Omega, K_{\Delta}^x(u) = \frac{1}{\Delta} K\left(\frac{u-x}{\Delta}\right) \quad (1.20)$$

Par convention, $K(u) = K_1^0(u)$.

La Figure 1.5 présente quatre noyaux sommatifs ayant une largeur de bande $\Delta = 1$. L'intervalle de référence est $\Omega = [-2, 2]$

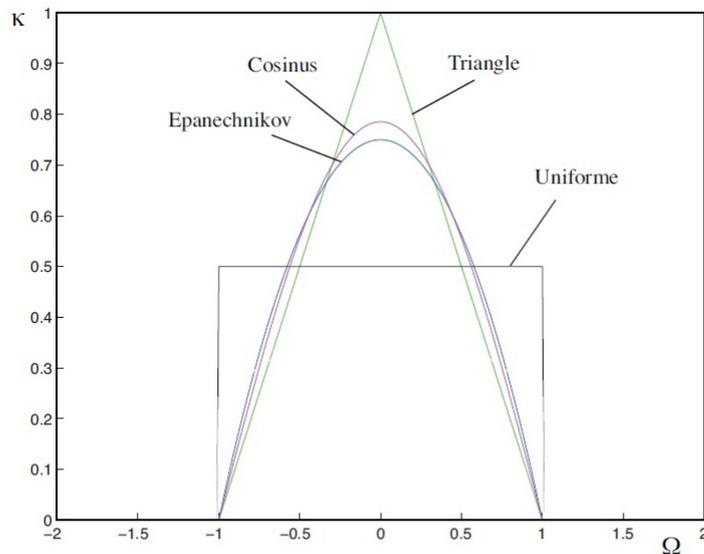


Figure 1.5 : Superposition de quatre noyaux sommatifs.

La plupart des noyaux sommatifs couramment utilisés en estimation fonctionnelle sont monomodaux, symétriques et centrés (c'est-à-dire définissant un voisinage autour de l'origine). Dans la suite de ce manuscrit, nous appelons $K(\Omega)$ l'ensemble des noyaux sommatifs monomodaux, symétriques et centrés. Un noyau sommatif K peut être vu comme une densité de probabilité induisant une mesure de probabilité, notée P_K , définie par :

$$\forall A \in \Omega, P_K(A) = \int_A K(u)du. \quad (1.21)$$

1.2.2 Propriétés statistiques de l'estimateur à noyau

Nous présentons, dans cette partie, les propriétés statistiques de l'estimateur de densité de Parzen-Rosenblatt.

Nous considérons ici les noyaux sommatifs K centrés de variance finie de $K(\Omega)$ c'est-à-dire vérifiant :

$$\int_{\Omega} uK(u)du = 0 \text{ et } \int_{\Omega} u^2K(u)du = \sigma_K^2 > 0. \quad (1.22)$$

σ_K^2 est appelée la *variance du noyau* K .

Le biais de l'estimateur $\hat{f}_{K\Delta}^n$ est donné pour tout $x \in \Omega$, par :

$$\text{biais}(\hat{f}_{K\Delta}^n(x)) = \frac{\Delta^2 \sigma_K^2 f''(x)}{2} + o(\Delta^4) \quad (1.23)$$

où f'' est la dérivée seconde de la fonction de densité f . f'' doit être une fonction absolument continue.

L'expression (1.23) montre que ce biais converge vers zéro quand la largeur de bande Δ tend vers zéro. Il découle de l'expression (1.23) que le biais de $\hat{f}_{K\Delta}^n$ ne dépend pas directement du nombre d'observations mais seulement de la fonction de noyau K et donc l'augmentation du nombre d'observations ne peut pas, à elle seule, diminuer le biais.

La variance de cet estimateur est donnée, pour tout $x \in \Omega$, par :

$$\text{var}(\hat{f}_{K\Delta}^n(x)) = \frac{f(x)R(K)}{n\Delta} + o(n^{-1}) \quad (1.24)$$

où $R(K) = \int_{\Omega} (K(u))^2 du$.

Cette variance tend vers zéro quand $n\Delta$ tend vers l'infini quand n tend vers l'infini.

La convergence en moyenne quadratique de l'estimateur $\hat{f}_{K\Delta}^n$ a été établie par Bosq et Lecoutre [3]. Nous avons, pour tout $x \in \Omega$:

$$MSE(\hat{f}_{K\Delta}^n(x)) = \frac{f(x)R(K)}{n\Delta} + \frac{\Delta^4 \sigma_K^4 (f''(x))^2}{4} + o(n^{-1}) + o(\Delta^6) \quad (1.25)$$

En intégrant l'expression (1.25) sur tout le domaine Ω , on obtient l'erreur quadratique moyenne intégrée de $\hat{f}_{K\Delta}^n$:

$$MISE = \frac{R(K)}{n\Delta} + \frac{\Delta^4 \sigma_K^4 R(f'')}{4} + o(n^{-1}) + o(\Delta^6) \quad (1.26)$$

où $R(f'') = \int_{\Omega} (f''(u))^2 du$.

La largeur de bande optimale de $\hat{f}_{K\Delta}^n$, notée Δ^* , est donnée par :

$$\Delta^* = \left(\frac{R(K)}{\sigma_K^4 R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}} \quad (1.27)$$

Notons que Δ^* est une quantité déterministe qui dépend du nombre d'observations n .

Elle dépend aussi de la densité inconnue f via l'intégrale du carré de sa dérivée seconde ($R(f'')$) et ne peut pas donc être utilisée telle qu'elle dans les calculs. Une façon classique de corriger ce défaut consiste à remplacer la quantité $R(f'')$, dans l'expression (1.27), par un estimateur approprié. Cette correction a fait l'objet de beaucoup de travaux (voir, par exemple, Hall et Marron [9], Jones et Sheather [10]). En remplaçant la valeur de Δ^* (1.27)

1.2.3 Un résultat théorique sur la convergence des estimateurs à noyau 26

dans l'expression 1.26, on obtient la valeur du AMISE*.

$$AMISE^* = \frac{5}{4} (\sigma_K R(K))^{\frac{4}{5}} (R(f''))^{\frac{1}{5}} n^{-\frac{4}{5}} \quad (1.28)$$

Epanechnikov [5] a proposé un *noyau optimal* de l'estimateur $\hat{f}_{K\Delta}^n(x)$ qui minimise le critere AMISE* (1.28) par rapport au noyau K . La solution K_{opt} obtenue est donnée par :

$$K_{opt}(u) = \begin{cases} \frac{3}{4}(1 - u^2) & \text{si } |u| \leq 1, \\ 0 & \text{sinon} \end{cases} \quad (1.29)$$

Ce noyau est appelé le noyau d'Epanechnikov.

1.2.3 Un résultat théorique sur la convergence des estimateurs à noyau

Appelons noyau de Geffroy [8] toute densité de probabilité bornée K dont l'ensemble des points de discontinuité est de mesure de Lebesgue nulle et telle que l'application

$$x \mapsto \sup(|K(u)|; x - 1 < u < x + 1) \quad (1.30)$$

soit intégrable sur \mathbb{R} (toujours par rapport à la mesure de Lebesgue ou pour l'intégrale de Riemann impropre).

Théorème 1.2.1. [12] *Pour tout échantillon théorique $(X_n, n \geq 1)$ suivant une loi inconnue de densité f uniformément continue sur \mathbb{R} , la suite $(\check{f}_n, n \geq 1)$ des estimateurs à noyau de Geffroy K définis par converge presque sûrement au sens de la convergence uniforme, c'est à dire que*

$$d_{sup}(\check{f}_n, f) \xrightarrow{p.s.} 0 \text{ quand } n \rightarrow +\infty. \quad (1.31)$$

si et seulement si

$$h_n \rightarrow 0 \text{ et } \frac{\log n}{n \cdot h_n} \rightarrow 0 \text{ quand } n \rightarrow +\infty. \quad (1.32)$$

Chapitre 2

Estimation non paramétrique de la densité par histogramme généralisé

Dans tout ce chapitre, on supposera que $\{X_n, n \geq 1\}$ est une suite de v.a indépendantes de même loi sur \mathbb{R} , possédant une densité $f(x)$, relativement à la mesure de lebesgue sur \mathbb{R} , continue.

2.1 Estimation de la densité

2.1.1 Les histogrammes ou estimateurs à noyaux

Leur forme la plus générale est

$$f_n(x) = \frac{1}{B_n} \sum_{i=1}^n K_{n,i}(x, X_i);$$

le plus classique d'entre eux est l'estimateur de Parzen-Rosenblatt (Rosenblatt(1956) [18], Parzen(1962) [14]), qui sera l'objet des développements ultérieurs de cette formule :

$$f_n(x) = \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\delta_n}\right) \quad (2.1)$$

2.1.2 Les estimations par fonctions orthogonales

Étant donné une suite $\{\varphi_n, n \geq 0\}$, orthonormale relativement au produit $\int \varphi(x)\psi(x)h(x)dx$, h étant une fonction fixée, si, de plus $f \in L^2(h)$ peut être représentée sous la forme

$$f(x) = \sum_{i=0}^{\infty} a_i \varphi_i(x),$$

on estime f par

$$f_n(x) = \sum_{i=0}^{m(n)} \hat{a}_i \varphi_i(x),$$

les \hat{a}_i étant des estimation de a_i données par exemple par

$$\hat{a}_i = \sum_{j=1}^n \varphi_i(x_j) h(x_j).$$

Cette méthode a été surtout développée à partir des travaux de Cencov (1962) [4]; citons également Van Ryzin (1966) [21], Schwarz-Stuart (1967) [19], Watson(1969) [23].

Il est intéressant de noter que les développements de la densité en série relativement à une base orthogonale sont beaucoup plus anciens que les travaux de Cencov.

On peut remarquer qu'il n'y a pas de différence formelle entre les estimations générales par la méthode du noyau et les estimations par fonctions orthogonales; en effet, pour ces dernières :

$$f_n(x) = \frac{1}{n} \sum_{j=1}^n \left(\sum_{i=0}^{m(n)} \varphi_i(x) \varphi_i(x_j) h(x_j) \right)$$

Certains auteurs ont utilisé ces similitudes(Watson (1969) [23]).Cependant les méthodes de traitement restent en général assez différentes.

2.1.3 Les méthodes dites du maximum de vraisemblance

On trouve plusieurs types d'estimations différentes sous cette appellation ; Citons, comme exemple typique, la note de Robertson (1967) [17], suivie par celles de Prakasa-Rao (1969) [15], Wegman (1969) [24], concernant principalement l'estimation de densités unimodales ; un exemple de ces estimations est de la manière suivante :

a) On pose

$$g_n(x) = \sum_{i=1}^n \frac{1}{M(A_i)} U_{A_i}(x)$$

où, si $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ est la statistique ordonnée de l'échantillon, si M est le mode supposé différent des valeurs prises par l'échantillon, $A_1 = (X_{1,n}, X_{2,n}), \dots, A_r = (X_{p,n}, M), A_{r+1} = (M, X_{p+1,n}), \dots, A_n = (X_{n-1,n}, X_{n,n})$, $M(A_i)$ désignant la mesure de Lebesgue de A_i ; $M(a, b) = b - a$. U_A désigne ici la fonction indicatrice de A

b) L'estimation $f_n(x)$ est alors la fonction définie de manière unique, telle que :

- $\forall a, \{x | f_n(x) > a\}$ est un intervalle contenant M ,
- $\int_{\mathbb{R}} f_n \theta(f_n) dx = \int_{\mathbb{R}} g_n \theta(f_n) dx, \forall \theta, \text{ avec } \theta(0) = 0,$
- $\int_A (g_n - f_n) dx \leq 0, \text{ si } A \text{ est un intervalle contenant } M.$

Pour l'estimation précédente, Robertson [17] prouve que f_n est un estimateur du maximum de vraisemblance parmi toutes les fonctions vérifiant la dernière expression.

Une nouvelle approche de l'estimation par le maximum de vraisemblance (qui dans le cas ordinaire ne peut donner de résultat, la vraisemblance n'étant pas bornée, consiste, non pas à limiter de façon très restrictive l'espace de fonctions pouvant être prises comme estimations, mais à maximiser :

$$\prod_{i=1}^n f(x_i) \exp(-\phi(f)),$$

au lieu de

$$\prod_{i=1} f(x_i),$$

parmi les fonctions d'une variété $H(\mathbb{R})$. ϕ est ici une fonctionnelle donnée.

Il est intéressant de constater que les estimations (dites du maximum de vraisemblance avec fonction de pénalisation) ainsi obtenues, fournissent dans des cas assez généraux des estimations polynomiales par morceaux (spline), les extrémités des intervalles (noeuds) étant des points de l'échantillon.

2.1.4 Les méthodes mixtes, les histogrammes lissés

Ces méthodes, d'introduction plus récente, réalisent le plus souvent un mélange des estimations précédentes, afin d'optimiser les résultats de convergence.

Un exemple typique est l'histo-spline, obtenu en réalisant dans un premier stade un histogramme simple, obtenu par le découpage du support de la distribution en r intervalle disjoints, puis en ajustant une estimation polynomiale par morceaux (spline) dont les noeuds sont les extrémités des intervalles, qui soit globalement de classe C^p , et qui respecte les probabilités empiriques de ces intervalles. Citons à ce sujet les travaux de Wahba (1971) [22].

L'efficacité de l'estimation :

La première difficulté réside dans la manière de mesurer cette efficacité. Le critère le plus commode semble être celui de l'écart quadratique moyen :

$$\begin{aligned} M^2 &= \int_{-\infty}^{+\infty} \mathbb{E}((f_n(x) - f(x))^2) dx \\ &= \mathbb{E} \left(\int_{-\infty}^{+\infty} (f_n(x) - f(x))^2 dx \right) \\ &= B_1^2 + B_2^2, \end{aligned} \tag{2.2}$$

où

$$B_1^2 = \int_{-\infty}^{+\infty} (\mathbb{E}(f_n(x)) - f(x))^2 dx, \quad B_2^2 = \int_{-\infty}^{+\infty} V(f_n(x)) dx. \quad (2.3)$$

Une estimation réalisant un minimum de M^2 sera dite estimation du M.I.S.E (minimum integrated square error). D'autres mesures d'écart ont été étudiées, notamment par Bickel-Rosenblatt(1973) [22] :

$$\sup \frac{|f_n(x) - f(x)|}{f(x)}, \quad \int_{\mathbb{R}} \frac{(f_n(x) - f(x))^2}{f(x)} dx$$

Il ne faut pas cependant juger l'efficacité d'une estimation, comme celà est fait trop souvent, en prenant un exemple particulier pour f , comme une loi normale, et en classant les estimation par ordre d'efficacité (pour ce cas particulier). Les estimations dépendent beaucoup du bon choix des paramètres de départ, et lorsque ceux-ci sont trop nombreux, leur robustesse est gravement mise en défaut.

Un exemple typique d'estimation par la méthode des fonctions orthogonales est obtenu en prenant $h(x) = (2\pi)^{\frac{1}{2}} e^{-\frac{1}{2}x^2}$, et comme fonction φ_i , les polynôme d'Hermite étant définis par :

$$\begin{aligned} H_i(x) &= (-1)^i h(x) \frac{d^i}{dx^i} h^{-1}(x) \\ &= \int_{-\infty}^{+\infty} H_m(x) H_n(x) h^{-1}(x) dx \\ &= \begin{cases} 0 & \text{si } m \neq n \\ n! & \text{si } m = n \end{cases} . \end{aligned} \quad (2.4)$$

On posera

$$\varphi_n(x) = \frac{1}{n!} H_n(x) h^{-1}(x)$$

Le développement de f dans cette base orthonormale est le développement de Gram-Charlier de f ; pour estimer une loi normale $\mathcal{N}(0,1)$, ce type de

développement est idéal, puisqu'on pourra se limiter au premier terme $a_0 = 1$, $a_n = 0$ si $n \geq 1$. On peut alors montrer que le développement de Gram-Charlier correspondant à une loi $\mathcal{N}(0, \frac{1}{2})$ est divergent. Ainsi, comme on le voit dans cet exemple, il suffit de diminuer l'écart type de la loi dans la proportion $1/\sqrt{2}$, pour obtenir des estimations non valables.

2.2 Estimations par histogrammes uniformes

On considère ici les estimations du type suivant :

$$f_n(x) = \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\delta_n}\right), \text{ avec les hypothèses suivantes :}$$

$\{X_n, n \geq 1\}$ est une suite de v.a indépendantes de même loi, possédant une densité f sur \mathbb{R} .

$\{\delta_n, n \geq 1\}$ est une suite positive décroissante, de limite 0.

K est une fonction bornée, telle que $\int_{-\infty}^{+\infty} K(y)dy = 1$; on notera en générale :

$$[[y^p K^r]] = \int_{-\infty}^{+\infty} y^p K^r(y)dy, \text{ si cette quantité est définie.}$$

Les noyaux usuels se limitent à la liste suivante :

A) Noyaux polynomiaux ou de Legendre :

En général $K(y) = P(y)$, $|y| < C$, où P est un polynôme, $K(y) = 0$ pour $|y| \geq C$. Comme exemple importants :

(1) Noyau unité : $K(y) = 1$, si $|y| < \frac{1}{2}$.

(2) Noyau d'Epanechnikov : $K(y) = 1 - y^2$, si $|y| < 1$.

(3) Noyau de Legendre d'ordre 1 : $K(y) = \frac{3}{8}(3 - 5y^2)$, si $|y| < 1$.

(4) Noyau de Legendre d'ordre 2 : $\frac{15}{128}(15 - 70y^2 + 63y^4)$, si $|y| < 1$

B) Noyaux de Gram-Charlier :

En général $K(y) = P(y)e^{-\frac{1}{2}y^2}$, P étant un polynôme. comme exemples importants :

(5) Noyau normal : $K(y) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2}$

(6) Noyau de Gram-Charlier d'ordre 1 : $K(y) = \frac{1}{2}(3 - y^2)(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2}$

(7) Noyau de Gram-Charlier d'ordre 2 :

$$\frac{1}{8}(y^4 - 10y^2 + 15)(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2}$$

c) Noyaux de Laguerre :

En général $K(y) = P(|y|)e^{-|y|}$, P étant un polynôme. comme exemples importants :

(8) Noyau de Picard : $K(y) = \frac{1}{2}e^{-|y|}$

(9) Noyau de Laguerre d'ordre 1 : $K(y) = \frac{1}{2}(\frac{1}{2}y^2 - 3|y| + 3)e^{-|y|}$

D) Noyaux divers :

(10) Noyau de Cauchy d'ordre r : $K(y) = \frac{2^{2r-2}\Gamma(r)^2}{\pi(2r-1)} \frac{1}{(1+x^2)^r}$

(11) $K(y) = \frac{(1-|y|^a)(a+1)}{2a}$, si $|y| < 1$; pour $a = 1$, noyau triangulaire $K(y) = 1 - |y|$

(12) $K(y) = \frac{e^{-|y|^a}}{2\Gamma(1/a)}$

(13) Noyau de Fejer-de la Vallée poussin : $K(y) = \pi^{-1} \left(\frac{1}{y} \sin y\right)^2$

(14) Noyau de Jackson-de la Vallée poussin : $K(y) = 3\pi^{-1} \left(\frac{1}{y} \sin y\right)^4$

(15) Noyau de Fourier : $K(y) = \pi^{-1} \left(\frac{1}{y} \sin y\right)$

(16) $K(y) = \cos y$, $|y| < \frac{1}{2}\pi$.

Les noyau 13,14 et 15 présentent peu d'intérêt pratique, nous ne les citons ici que pour mémoire.

Nous étudions maintenant les propriétés en général de convergence en moyenne quadratique des estimations :

2.2.1 Moments des estimations

Par un changement de variable, on obtient l'identité :

$$\mathbb{E} \left(K^p \left(\frac{X - x}{u} \right) \right) = u \int_{-\infty}^{+\infty} K^p(y) f(x + uy) dy \quad (2.5)$$

supposons que f soit n fois différentiable en x :

$$f(x + h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(x) + \dots + \frac{h^n}{n!} f^{(n)}(x) + e_n(x, h)h^n \quad (2.6)$$

Dans le cas où f est $n + 1$ fois dérivable dans l'intervalle $(x, x + h)$,

$$e_n(x, h)h^n = \int_0^h \frac{(h - t)^n}{n!} f^{(n+1)}(x + t) dt \quad (2.7)$$

On obtient alors formellement, on intégrant (2.6) dans (2.5) :

$$\frac{1}{u} \mathbb{E} \left(K^p \left(\frac{X - x}{u} \right) \right) = \sum_{r=0}^n u^r \frac{f^{(r)}(x)}{r!} [[y^r K^p]] + u^n R^n(x, u), \quad (2.8)$$

avec :

$$R_n(x, h) = \int_{-\infty}^{+\infty} y^n K^p(y) e_n(x, yh) dy$$

Pour obtenir que, dans (2.7), $\lim_{u \rightarrow 0} R_n(x, u) = 0$, il est nécessaire de faire des hypothèses sur K et f :

Proposition 2.2.1. *:Si f est n fois différentiable en x ,*

1) *Si K est nul en dehors d'un compact (cas des noyaux 1,2,6), ou,*

2) Si f est définie sur \mathbb{R} et y admet une majoration de la forme :

$$f(y) \leq C(1 + |y|^R), R \geq 0, \quad (2.9)$$

et si $[[y^r K^p]]$ est défini $\forall r \geq 0$, ou,

3) Si f vérifie la condition du 2), sauf au voisinage d'un nombre fini de points où elle admet des discontinuités infinies, et si K vérifie l'hypothèse :

$$\lim_{y \rightarrow \infty} y^r K^p(y) = 0, \forall r \geq 0, \quad (2.10)$$

Alors (2.8) est vérifié, avec $\lim_{u \rightarrow 0} R_n(x, u) = 0$.

Par la suite, nous supposerons toujours que K est borné.

Corollaire 2.2.1. : Si f est N fois différentiable en x , on a :

$$\mathbb{E}(f_n(x)) = f(x) + \sum_{r=1}^N \delta_n^r \frac{f^{(r)}(x)}{r!} [[y^r K]] + R_N(x, \delta_n) \delta_n^N, \quad (2.11)$$

avec $\lim_{u \rightarrow 0} R_n(x, u) = 0$,

$$\begin{aligned} (\mathbb{E}(f_n(x)) - f(x))^2 &= \sum_{r=2}^{N+1} \frac{\delta_n^r}{r!} \left(\sum_{p=1}^{r-1} C_r^p f^{(p)}(x) f^{(r-p)}(x) [[y^p K]] [[y^{r-p} K]] \right) \\ &\quad + S_{N+1}(x, \delta_n) \delta_n^{N+1}, \end{aligned} \quad (2.12)$$

avec $\lim_{u \rightarrow 0} S_{N+1}(x, u) = 0$.

$$\begin{aligned} V(f_n(x)) &= \frac{1}{n\delta_n} \left(\sum_{r=0}^N \delta_n^r \frac{f^{(r)}(x)}{r!} [[y^r K^2]] \right) \\ &\quad - \frac{1}{n\delta_n} \left(\sum_{r=0}^{N-1} \frac{\delta_n^{r+1}}{r!} \left(\sum_{p=0}^r C_r^p f^{(p)}(x) f^{(r-p)}(x) [[y^p K]] [[y^{r-p} K]] \right) \right) \\ &\quad + \frac{1}{n\delta_n} T_n(x, \delta_n) \delta_n^N, \end{aligned} \quad (2.13)$$

avec $\lim_{u \rightarrow 0} T_n(x, u) = 0$

Remarque 2.2.1. :Les resultats (2.11), (2.12), (2.13) se déduisent aisément de (2.8), les v.a étant indépendantes, et car :

$$V \left(K \left(\frac{X-x}{u} \right) \right) = \mathbb{E} \left(K^2 \left(\frac{X-x}{u} \right) \right) - \mathbb{E} \left(K \left(\frac{X-x}{u} \right) \right)^2.$$

Ces resultats restent valables lorsque $N = 0$, c'est-à-dire lorsque f est continue en x ; on supprimera alors les sommations impossibles ; on obtient dans ce cas :

$$\mathbb{E}(f_n(x)) - f(x) = R_0(x, \delta_n), V(f_n(x)) = \frac{[[K^2]]}{n\delta_n}(f(x) + T_0(x, \delta_n)).$$

Ces résultats sont valables pour les noyaux 1-2-3-4-5-6-7-8-9-11-12-16 en particulier.

Ils n'ont plus de sens, les moments correspondants n'étant pas définis, pour les noyaux 13-14-15.

Il est intéressant, lorsque f est analytique en x , d'obtenir un développement en série, à la place de (2.8) :

$$\frac{1}{u} \mathbb{E} \left(K^p \left(\frac{X-x}{u} \right) \right) = \sum_{r=0}^{\infty} u^r \frac{f^r(x)}{r!} [[y^r K^p]] \quad (2.14)$$

Proposition 2.2.2. :Si f est analytique au voisinage du point x ,

- 1) Si r est le rayon de convergence du développement de f en série entière au voisinage de x , si $K(y) = 0$ pour $|y| > C$, alors (2.14) est vérifié et convergent pour $|u| < r/C$.
- 2) Si f est analytique sur \mathbb{R} , et si $r = \infty$, si K n'est pas nul en dehors d'un compact de \mathbb{R} , (2.14) est vérifié et convergent pour

$$|u| < \frac{1}{\limsup_{n \rightarrow \infty} \left(\frac{f^{(n)}(x)}{n!} \int_{-\infty}^{+\infty} |y|^n |K(y)| dy \right)^{1/n}}$$

ou si $K \geq 0$,

$$|u| < \frac{1}{\limsup_{n \rightarrow \infty} \left(\frac{f^{(n)}(x)}{n!} [[y^n K]] \right)^{1/n}} \quad (2.15)$$

Corollaire 2.2.2. :Avec les hypothèses de la proposition 2.2.2 entraînant la validité de (2.14) :

$$\mathbb{E}(f_n(x)) = f(x) + \sum_{r=1}^{\infty} \delta_n^r \frac{f^{(r)}(x)}{r!} [[y^r K]], \quad (2.16)$$

$$(\mathbb{E}(f_n(x)) - f(x))^2 = \sum_{r=2}^{\infty} \frac{\delta_n^r}{r!} \left(\sum_{p=1}^{r-1} C_r^p f^{(p)}(x) f^{(r-p)}(x) [[y^p K]] [[y^{r-p} K]] \right) \quad (2.17)$$

$$V(f_n(x)) = \frac{1}{n\delta_n} \left(\sum_{r=0}^{\infty} \delta_n^r \frac{f^{(r)}(x)}{r!} [[y^r K^2]] \right) - \frac{1}{n\delta_n} \left(\sum_{r=0}^{\infty} \frac{\delta_n^{r+1}}{r!} \left(\sum_{p=0}^{\infty} C_r^p f^{(p)}(x) f^{(r-p)}(x) [[y^p K]] [[y^{r-p} K]] \right) \right) \quad (2.18)$$

2.2.2 Moyenne quadratique intégrée

Nous estimons ici $M^2 = \mathbb{E} \left(\int_I (f_n(x) - f(x))^2 dx \right) = B_1^2 + B_2^2$, où :

$$B_1^2 = \int_I (\mathbb{E}(f_n(x)) - f(x))^2 dx, \quad B_2^2 = \int_I V(f_n(x)) dx.$$

Proposition 2.2.3. :Si f est N fois continûment différentiable dans un ouvert contenant I , et dans l'un ou l'autre des cas suivants :

- 1) I est un compact, et K est un noyau nul en dehors d'un compact ;
- 2) $f^{(p)} \in L^1(V(I)) \cap L^2(V(I))$, $\forall 0 \leq p \leq N + 1$, $V(I)$ étant un voisinage ouvert de I , et ou bien K est nul en dehors d'un compact, ou bien

$I = \mathbb{R}$; alors :

$$B_1^2 = \sum_{r=2}^{N+1} \frac{\delta_n^r}{r!} \left(\sum_{p=1}^{r-1} C_r^p [[y^p K]] [[y^{r-p} K]] \int_I f^{(p)}(x) f^{(r-p)}(x) dx \right) + o(\delta_n^{N+1}) \quad (2.19)$$

$$B_2^2 = \frac{1}{n\delta_n} \left(\sum_{r=0}^N \frac{\delta_n^r}{r!} [[y^r K^2]] \int_I f^{(r)}(x) dx \right) - \frac{1}{n\delta_n} \left(\sum_{r=0}^{N-1} \frac{\delta_n^{r+1}}{r!} \left(\sum_{p=0}^r C_r^p [[y^p K]] [[y^{r-p} K]] \int_I f^{(p)}(x) f^{(r-p)}(x) dx \right) + o(\delta_n^N) \right) \quad (2.20)$$

Remarque 2.2.2. 1) Si $f^{(p)}$, $0 \leq p \leq N$, s'annule aux extrémités de I , [2.19](#), [2.20](#) se simplifient considérablement. En effet :

$$\int_I f^{(p)}(x) f^{(r-p)}(x) = - \int_I f^{(p+1)}(x) f^{(r-p-1)}(x);$$

$$\sum_{p=1}^{2r} \text{ et } \sum_{p=0}^{2r+1} (-1)^p C_{2r+1}^p [[y^p K]] [[y^{2r-p+1} K]]$$

ont un coefficient nul ; on obtient en conséquence :

$$B_1^2 = \sum_{r=1}^{(\frac{1}{2}(N+1))} \frac{\delta_n^{2r}}{2r!} \left(\int (f^{(r)}(x))^2 dx \right) \left(\sum_{p=1}^{2r-1} C_{2r}^p (-1)^{p+r} [[y^p K]] [[y^{2r-p} K]] \right) + o(\delta_n^{N+1}) \quad (2.21)$$

$$\begin{aligned}
B_2^2 &= \frac{1}{n\delta_n} \left([[K^2]] - \delta_n \sum_{r=0}^{(\frac{1}{2}(N+1))} \frac{\delta_n^{2r}}{(2r)!} \left(\int (f^{(r)}(x))^2 dx \right) \right) \\
&\cdot \left(-\frac{1}{n} \sum_{p=0}^{2r} C_{2r}^p (-1)^{p+r} [[y^p K]] [[y^{2r-p} K]] \right) \\
&\quad + \frac{1}{n\delta_n} o(\delta_n^N)
\end{aligned} \tag{2.22}$$

2) Si, en plus des hypothèses précédentes, K est symétrique relativement à 0, $[[y^{2r+1}K]]$ sera nul pour tout r . on obtient

$$\begin{aligned}
B_1^2 &= \sum_{r=2}^{(\frac{1}{2}(N+1))} \frac{\delta_n^{2r}}{(2r)!} \left(\int (f^{(r)}(x))^2 dx \right) \left((-1)^r \sum_{p=1}^{r-1} C_{2r}^{2p} [[y^{2p} K]] [[y^{2r-2p} K]] \right) \\
&\quad + o(\delta_n^{N+1})
\end{aligned} \tag{2.23}$$

$$\begin{aligned}
B_2^2 &= \frac{1}{n\delta_n} \left([[K^2]] - \delta_n \sum_{r=0}^{(\frac{1}{2}(N+1))} \frac{\delta_n^{2r}}{(2r)!} \left(\int (f^{(r)}(x))^2 dx \right) \right) \\
&\cdot \left(-(-1)^r \frac{1}{n} \sum_{p=0}^r C_{2r}^{2p} [[y^{2p} K]] [[y^{2r-2p} K]] \right) \\
&\quad + \frac{1}{n\delta_n} o(\delta_n^N)
\end{aligned} \tag{2.24}$$

Proposition 2.2.4. :Si f est analytique sur I , et si les condition de la proposition 2.2.2 entrainant 2.14 pour $u = \delta_n$ sont vérifiées en tout point de I , si de plus les conditions de la proposition 2.2.3 2) sont vérifiées pour $N = \infty$, c'est-à-dire $\forall p \geq 0, f^{(p)} \in L^1(V(I)) \cap L^2(V(I))$, $V(I)$ étant un voisinage ouvert de I , et ou bien K est nul en dehors d'un compact, ou bien $I = \mathbb{R}$, si

de plus

$$|\delta_n| < \frac{1}{\limsup_{n \rightarrow \infty} \left(\left(\int_I |f^n(x)|^2 dx \right)^{\frac{1}{2}} \frac{1}{n!} \int_{-\infty}^{+\infty} |y|^n |K(y)| dy \right)^{\frac{1}{n}}} \quad (2.25)$$

$$B_1^2 = \sum_{r=2}^{\infty} \frac{\delta_n^r}{r!} \left(\sum_{p=1}^{r-1} C_r^p [[y^p K]] [[y^{r-p}]] \int_I f^{(p)}(x) f^{(r-p)}(x) dx \right) \quad (2.26)$$

$$B_2^2 = \frac{1}{n\delta_n} \left(\sum_{r=0}^{\infty} \frac{\delta_n^r}{r!} [[y^r K^2]] \int_I f^{(r)}(x) dx \right) \quad (2.27)$$

$$- \frac{1}{n\delta_n} \left(\sum_{r=0}^{\infty} \frac{\delta_n^{r+1}}{r!} \left(\sum_{p=0}^r C_r^p [[y^p]] [[y^{r-p} K]] \int_I f^{(p)}(x) f^{(r-p)}(x) dx \right) \right)$$

2.2.3 Développements pour la loi normale $\mathcal{N}(0,1)$, et les noyaux symétriques

Soit $f(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$; dans la suite, il sera utile de déterminer

$$I_n = \int_{-\infty}^{+\infty} x^{2n} f(x)^2 dx;$$

$$I_n = \int_{-\infty}^{+\infty} \frac{x^{2n}}{2\pi} e^{-x^2} dx = \int_{-\infty}^{+\infty} \frac{(\frac{1}{2})^{n_u} 2^{2n}}{\sqrt{2}(2\pi)} e^{-\frac{1}{2}u^2} du = \frac{(\frac{1}{2})^{n+1}}{\sqrt{\pi}} \mathbb{E}(X^{2n}) = \frac{(\frac{1}{2})^{2n+1} (2n)!}{\sqrt{\pi n!}}$$

tout d'abord

$$\int_{-\infty}^{+\infty} |f^{(2n)}(x)| dx \leq (2n)! \sum_{r=0}^n \frac{\mathbb{E}(X^{2n-2r})}{2^r r! (2n-2r)!} \leq \frac{(2n)!}{n!} \sum_{r=0}^n \frac{n!}{2^r r! (n-r)!} \leq \frac{(2n)!}{n!}$$

D'autre part, comme

2.2.3 Développements pour la loi normale $\mathcal{N}(0,1)$, et les noyaux symétriques

41

$$\int_0^{+\infty} x^{2r+1} e^{-\frac{1}{2}x^2} dx = 2^r r! \leq C r^{\frac{1}{2}} 2^{r-\frac{1}{2}} \Gamma(r + \frac{1}{2}) = C r^{\frac{1}{2}} \frac{(\frac{1}{2})^{r+1} (2r)! \sqrt{2\pi}}{r!}$$

$$\begin{aligned} \int_{-\infty}^{+\infty} |f^{(2n+1)}(x)| dx &\leq \frac{2(2n+1)!}{\sqrt{2\pi}} \sum_{r=0}^n \frac{2^n (n-r)!}{2^{2r} r! (2n-2r+1)!} \\ &\leq \frac{C(2n+1)!}{n!} \sum_{r=0}^n \frac{n!}{2^{nr} (n-r)! (2n-2r+1)!} \\ &\leq \frac{C(2n+1)!}{n!} \end{aligned}$$

considérons maintenant

$$\int_{-\infty}^{+\infty} f^{(2n-r)}(x) f^{(r)}(x) dx = (-1)^r \int_{-\infty}^{+\infty} f^{(2n)}(x) f(x) dx$$

et

$$\begin{aligned} \int_{-\infty}^{+\infty} f^{(2n)}(x) f(x) dx &= (2n)! \sum_{r=0}^n \frac{(-1)^r I_{n-r}}{2^r r! (2n-2r)!} \\ &= \frac{(\frac{1}{2})^{2n+1} (2n)!}{n! \sqrt{\pi}} \left(\sum_{r=0}^n C_n^r (-2)^r \right) \\ &= \frac{(-1)^n (\frac{1}{2})^{2n+1} (2n)!}{n! \sqrt{\pi}} \end{aligned}$$

on obtient alors par la formule de Stirling : asymptotiquement,

$$\left(\frac{1}{n!} \int_{-\infty}^{+\infty} |f^n(x)| dx \right)^{\frac{1}{n}} \leq \sim \sqrt{\frac{2e}{n}};$$

$$\left(\frac{1}{n!} \left(\int_{-\infty}^{+\infty} |f^n(x)|^2 dx \right)^{\frac{1}{2}} \right)^{\frac{1}{n}} \sim \sqrt{\frac{e}{n}} \quad (2.28)$$

2.2.4 Développement par les noyaux de Gram-Charlier

Le i ème polynôme d'Hermite $H_1(x)$ étant défini par 2.1. Les noyaux de Gram-Charlier sont de la forme :

$$K(y) = \sum_{r=0}^R c_r H_{2r}(y) \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$$

Ils peuvent être également définis sous une forme dissymétrique, faisant intervenir les polynômes d'Hermite impairs, mais présentant moins d'intérêt alors. On obtient immédiatement que $[[y^{2n+1}K]] = 0, \forall n \geq 1$; calculons $[[y^{2n}K]]$:

$$y^{2n} = \sum_{p=0}^n \left(\frac{(2n)!}{(2p)!2^{n-p}(n-p)!} \right) H_{2p}(y)$$

d'où

$$[[y^{2n}K]] = (2n)! \sum_{r=0}^{\inf(n,R)} c_r \left(\frac{1}{2^{n-r}(n-r)!} \right); \quad (2.29)$$

écrivons les premiers termes obtenus par cette formule, en posant $c_r = 0$ pour $r > R$:

$$[[K]] = c_0, [[y^2K]] = c_0 + 2c_1, [[y^4K]] = 3c_0 + 12c_1 + 24c_2$$

$$[[y^6K]] = 15c_0 + 90c_1 + 360c_2 + 720c_3,$$

$$[[y^8K]] = 105(c_0 + 8c_1 + 48c_2 + 192c_3 + 384c_4), \dots,$$

On constate que, si on pose $c_0 = 1$, pour que $[[K]] = 1$, il existe $\forall R, c_1, \dots, c_R$, tels que $[[y^nK]] = 0, \forall 1 \leq n \leq 2R$. Les premières valeurs de cette suite sont :

$$c_0 = 1, c_1 = -\frac{1}{2}, c_2 = 1/8, c_3 = -1/48, c_4 = 1/384, \dots$$

On obtient l'expression générale de c_n :

$$c_n = \frac{(-1)^n}{2^n n!} \quad (2.30)$$

On peut définir, formellement, la somme de la série suivante :

$$K(y) = \sum_{n=0}^{\infty} c_n H_{2n}(y) \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} = \sum_{n=0}^{\infty} (-1)^n e^{-\frac{1}{2}x^2} H_{2n}(y) \quad (2.31)$$

En posant :

$$g(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2},$$

on pourra écrire formellement :

$$K(y) = \sum_{n=0}^{\infty} \frac{\left(-\frac{1}{2}\right)^n}{n!} g^{2n}(y);$$

on en déduit formellement

$$[[K^2]] = \sum_{n=0}^{\infty} \frac{\left(\frac{1}{2}\right)^{3n+1} (2n)!}{(n!)^2} \left(\sum_{r=0}^n \frac{n!}{(n-r)!r!} \right) = \sum_{n=0}^{\infty} \left(\frac{(2n)!}{(n!)^2 2^{2n+1}} \right)$$

expression divergente, qu'on retrouve dans le développement :

$$\frac{1}{\sqrt{1-x}} = \sum_{n=0}^{\infty} \frac{(2n)! x^n}{(n!)^2 2^{2n}}, \quad \text{pour } |x| < 1. \quad (2.32)$$

2.2.5 Développements par les noyaux de Legendre

Le nième polynôme de Legendre est défini par :

$$\begin{aligned} P_n(x) &= \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n \\ &= \frac{1}{2^n n!} \sum_{p=0}^{\binom{n}{2}} C_n^p \frac{(-1)^p x^{n-2p} (2n-2p)!}{(n-2p)!} \end{aligned} \quad (2.33)$$

Les polynôme de Legendre vérifient :

$$\int_{-1}^{+1} P_m(x) P_n(x) dx = 0 \text{ si } m \neq n, \frac{2}{2n+1}, m = n \quad (2.34)$$

On définit en conséquence :

$$K(y) = \sum_{r=0}^R c_r P_{2r}(y), \text{ si } |y| < 1, K(y) = 0, \text{ si } |y| > 1$$

On obtient ainsi les noyaux de Legendre, qui ont comme cas particulier le noyau unité (n^o 1). On pourrait également faire intervenir Les polynômes impairs, mais ceci ne présenterait que peu d'intérêt.

On constate immédiatement que :

$$\int_{-1}^{+1} y^{2n+1} P_{2r}(y) dy = 0,$$

et que

$$\int_{-1}^{+1} y^{2n} P_{2r}(y) dy = 0, \text{ si } n < r.$$

Si $n \geq r$:

$$\begin{aligned} \int_{-1}^{+1} y^{2n} P_{2r}(y) dy &= \left(\frac{1}{2}\right)^{2r} C_{2n}^{2r} \int_{-1}^{+1} y^{2n-2r} (y^2 - 1)^{2r} dy \\ &= \left(\frac{1}{2}\right)^{2r} C_{2n}^{2r} \beta\left(n - r + \frac{1}{2}, 2r + 1\right) \\ &= \frac{2^{2r+1} (n+r)! (2n)!}{(n-r)! (2n+2r+1)!}, \end{aligned}$$

d'où :

$$[[y^{2n} K]] = \sum_{r=0}^{inf(n,R)} c_r \left(\frac{2^{2r+1} (n+1)! (2n)!}{(n-r)! (2n+2r+1)!} \right), \quad (2.35)$$

ce qui fournit en particulier :

$$[[K]] = 2c_0, \quad [[y^2 K]] = \frac{2}{3}(c_0 + \frac{2}{5}c_1), \quad [[y^4 K]] = \frac{2}{5}(c_0 + \frac{4}{7}c_1 + \frac{8}{63}c_2)$$

$$[[y^6 K]] = \frac{2}{7}(c_0 + \frac{2}{3}c_1 + \frac{8}{33}c_2 + \frac{16}{429}c_3),$$

$$[[y^8 K]] = \frac{2}{9}((c_0 + \frac{8}{11}c_1 + \frac{48}{143}c_2 + \frac{64}{715}c_3 + \frac{128}{12155}c_4)), \dots,$$

On constate, qu'en posant $c_0 = \frac{1}{2}$, pour que $[[K]] = 1$, il existe $\forall R, c_1, \dots, c_R$,

tels que $[[y^n K]] = 0, \forall 1 \leq n \leq 2R$. Les premières valeurs de cette suite sont :

$$c_0 = \frac{1}{2}, \quad c_1 = -\frac{5}{4}, \quad c_2 = 27/16, \quad c_3 = -65/32, \quad c_4 = 2025/256, \dots$$

Cette suite a une limite infinie on oscillant ; en posant $R = \infty$ est divergente.

Les premiers polynômes de Legendre sont donnés par :

$$\begin{aligned}
 P_0(x) &= 1, P_1(x) = x, P_2(x) = \frac{1}{2}(3x^2 - 1), \\
 P_3(x) &= \frac{1}{2}(5x^3 - 3x), P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3), \\
 P_5(x) &= \frac{1}{8}(63x^5 - 70x^3 + 15x), \\
 P_6(x) &= \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5), \\
 P_7(x) &= \frac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x), \\
 P_8(x) &= \frac{1}{128}(6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35),
 \end{aligned}$$

plutôt que par l'usage de 2.33, on construit plus aisément ces polynômes par la formule de récurrence :

$$(n + 1)P_{n+1}(x) - (2n + 1)xP_n(x) + nP_{n-1}(x) = 0 \quad (2.36)$$

Les propriétés suivantes sont vérifiées par les polynômes de Legendre :

$$P_n(1) = 1, \quad P_n(-1) = (-1)^n$$

pour $-1 \leq x \leq 1$, on a la majoration de Stieltjes :

$$|P_n(x)| \leq \frac{4\sqrt{2}}{\sqrt{\pi n}(1 - x^2)^{\frac{1}{4}}}$$

2.2.6 Développement par les noyaux de Laguerre

Le $n^{ième}$ polynôme de Laguerre est défini à l'ordre a , par : ($a \in \mathbb{R}, a > -1$)

$$\begin{aligned}
 P_{n,a}(x) &= \frac{x^{-a}e^x}{n!} \frac{d^n}{dx^n}(x^{n+a}e^{-x}) \\
 &= (-1)^n \sum_{m=0}^n \frac{\Gamma(a + n + 1)(-1)^m}{m!(n - m)!\Gamma(a + n - m + 1)} x^{n-m}, n \geq 1,
 \end{aligned}$$

$$P_{0,a}(x) = 1$$

Les polynômes de Laguerre vérifiant :

$$\begin{cases} \int_0^{+\infty} e^{-x} x^a P_{n,a}(x) P_{m,a}(x) dx = 0 & \text{si } m \neq n \\ \frac{\Gamma(n+a+1)}{n!} & \text{si } m = n \end{cases}$$

On définit les noyaux de Laguerre par :

$$K(y) = \sum_{r=0}^R c_r P_{r,a}(|y|) |y|^a e^{-|y|} \quad (2.37)$$

Remarquons qu'on obtient ainsi comme cas particulier le noyau de picard ($a = 0$, $R = 0$, $n^o 8$). Nous étudierons ici principalement les noyaux de Laguerre d'ordre 0, bien que les résultats puissent être étendus sans difficulté aux noyaux d'ordre a quelconque.

pour ces noyaux :

$$P_n(x) = P_{n,0}(x) = (-1)^n \sum_{m=0}^n C_n^m \frac{(-1)^m x^m}{m!},$$

et

$$[[y^{2n+1}K]] = 0, \forall n \geq 1$$

$$[[y^{2n}K]] = 2(2n!)^2 \sum_{r=0}^{\inf(R,2n)} \frac{(-1)^r}{r!(2n-r)!} c_r$$

On constate, qu'en posant $c_0 = \frac{1}{2}$, pour que $[[K]] = 1$, il existe $\forall R$, c_1, \dots, c_R , tels que $[[y^n K]] = 0$, $\forall 1 \leq n \leq \frac{1}{2}R$. La solution est la suite : $c_r = \frac{1}{2}$ on obtiendra donc :

$$K_n(y) = \frac{1}{2} \sum_{r=0}^{2n} P_r(|y|) e^{-|y|} \quad (2.38)$$

La série ainsi obtenue n'est pas convergente, comme le montre l'approximation d'Uspensky :

$$P_r(x) = \sqrt{\frac{2}{\pi}} \left[\frac{e^{\frac{1}{2}x}}{((4n+1)x)^{1/4}} \right] \cos\left(\sqrt{(4n+1)x} - \frac{\pi}{4}\right) + o\left(\frac{x^{3/2}}{\sqrt{4n+1}}\right)$$

uniformément en $n > 1$, sur (a,b) fixé.

2.2.7 Comportement asymptotique des estimations, critère du M.I.S.E

on suppose ici que les hypothèses de la proposition 2.2.3 sont vérifiées.

1) Noyau symétrique tel que $[[y^2 K]] > 0$:

$$B_1^2 \sim \frac{1}{4} \delta_n^4 [[y^2 K]]^2 \left(\int_I (f''(x))^2 dx \right),$$

$$B_2^2 \sim \frac{1}{n \delta_n} [[K^2]]$$

comme M^2 est de la forme $M^2 = A\delta^4 + \frac{B}{n\delta}$, on obtient la valeur minimale de M^2 en annulant la dérivée de cette expression, soit $\delta^5 = \frac{B}{4An}$; on obtient donc ici :

$$\delta_n = \left(\frac{[[K^2]]}{[[y^2 K]]^2 \int_I (f''(x))^2 dx} \right)^{\frac{1}{5}} n^{-1/5}$$

d'où

$$M^2 \sim \frac{5[[K^2]]}{4n\delta_n} = \frac{5}{4} [[K^2]]^{\frac{5}{4}} [[y^2 K]]^{\frac{2}{5}} \left(\int_I (f''(x))^2 dx \right)^{\frac{1}{5}} n^{-4/5} \quad (2.39)$$

Certains auteurs, notamment Rosenblatt [18], préfèrent un autre choix de δ que celui fourni en (2.39), et obtenu en écrivant que :

$B_1^2 \sim B_2^2$, ce qui fournit, avec les notations précédentes $\delta^5 = \frac{B}{An}$

$$\delta'_n = \left(\frac{4[[K^2]]}{[[y^2 K]]^2 \int_I (f''(x))^2 dx} \right)^{\frac{1}{5}}$$

d'où

$$M'^2 \sim \frac{2[[K^2]]}{n\delta'_n} = 2^{\frac{3}{5}} [[K^2]]^{\frac{4}{5}} [[y^2 K]]^{\frac{2}{5}} \left(\int_I (f''(x))^2 dx \right)^{\frac{1}{5}} n^{-4/5} \quad (2.40)$$

2) Noyau symétrique tel que $[[y^2 K]] = \dots = [[y^{2p-2} K]] = 0$, $[[y^{2p} K]] > 0$:

On obtient de tels noyaux par exemple par les noyaux de Legendre, Gram-Charlier, Laguerre, étudiés précédemment, on obtient :

$$B_1^2 \sim \frac{1}{((2p)!)^2} \delta_n^{4p} [[y^{2p} K]]^2 \left(\int_I (f^{(2p)}(x))^2 dx \right),$$

$$B_2^2 \sim \frac{1}{n \delta_n} [[K^2]]$$

Le même raisonnement que dans 2.39, fournit une valeur asymptotiquement optimale de la forme :

$$\delta_n = \left(\frac{((2p)!)^2 [[K^2]]}{4p [[y^{2p} K]]^2 \int_I (f^{(2p)}(x))^2 dx} \right)^{\frac{1}{4p+1}} n^{-1}/(p+1)$$

$$M^2 \sim \frac{(4p+1) [[K^2]]}{4pn \delta_n} = \frac{(4p+1)}{4p} [[K^2]]^{\frac{4p}{4p+1}} \left(\int_I (f^{(2p)}(x))^2 dx \right)^{\frac{1}{4p+1}} n^{-\frac{4p}{(4p+1)}} \quad (2.41)$$

Naturellement les noyaux qui vérifient ces propriétés ne peuvent être positifs partout.

Chapitre 3

Estimation fonctionnelle

Dans tout ce chapitre on suppose que les observations x_1, \dots, x_n sont des réalisations de variables aléatoires fonctionnelles X_1, \dots, X_n indépendantes et de même loi, de fonction de répartition F , et de densité f , si elle existe.

Dans le chapitre précédent, on s'est intéressé à l'estimation de quantités réelles caractéristiques de la loi de probabilité de l'échantillon et les moments. Aussi riches d'enseignement que soient ces quantités, elles ne suffisent pas à déterminer entièrement la loi de probabilité de l'échantillon.

C'est pourquoi nous allons maintenant nous intéresser à l'estimation de la fonction de répartition et, si elle existe, de la densité de l'échantillon. Par rapport au chapitre précédent, il s'agit maintenant d'estimer des fonctions, d'où le nom d'**estimation fonctionnelle**. De plus l'une comme l'autre de ces fonctions caractérisent entièrement la loi de probabilité de l'échantillon.

La fonction de répartition empirique est un estimateur simple et performant de la fonction de répartition de l'échantillon. Il est beaucoup plus difficile d'estimer une densité. On connaît déjà l'estimateur de base de la densité d'un échantillon, l'histogramme. Bien que très connu et très utilisé, il est de médiocre qualité.

Estimer une fonction g , c'est d'abord estimer $g(x)$ pour tout x donné. Il faut ensuite juger de la qualité de l'estimation de $g(x)$ pour chaque x , puis de l'estimation de g dans son ensemble.

Si $\hat{g}(x)$ est un estimateur de $g(x)$, la qualité de l'estimation pour un x

donné est usuellement mesurée par le biais, la variance et l'Erreur Quadratique Moyenne (ou risque quadratique), qu'on notera $EQM_x(\hat{g})$:

$$EQM_x(\hat{g}) = \mathbb{E}[(\hat{g}(x) - g(x))^2] = [\mathbb{E}(\hat{g}(x)) - g(x)]^2 + Var(\hat{g}(x))$$

On voit que l'erreur quadratique moyenne se décompose en un terme de biais et un terme de variance. Si $\hat{g}(x)$ est un estimateur sans biais de $g(x)$, l'erreur quadratique moyenne se réduit à la variance. On verra que, si on peut trouver facilement un estimateur sans biais pour la fonction de répartition en un point x , il n'en est pas de même pour la densité. Aussi utilisera-t-on l'erreur quadratique moyenne plutôt que la variance dans ce cas.

Pour juger de la qualité de l'estimation de g dans son ensemble, il faut utiliser des mesures de l'écart entre g et \hat{g} . Suivant les cas, on utilisera :

- l'**Erreur Quadratique Moyenne Intégrée** (EQMI) :

$$EQMI(\hat{g}) = \int_{-\infty}^{+\infty} EQM_x(\hat{g})dx = \int_{-\infty}^{+\infty} [\mathbb{E}(\hat{g}(x)) - g(x)]^2 dx + \int_{-\infty}^{+\infty} Var(\hat{g}(x))dx$$

- l'écart maximum entre les deux fonctions :

$$\sup\{|\hat{g}(x) - g(x)|; x \in \mathbb{R}\}.$$

3.1 Données fonctionnelles

L'analyse des données fonctionnelles (Ramsay et Silverman, 2005 [16]; Ferraty et Vieu, 2006 [7]; Ferraty et Romain, 2011[6]) a connu un intérêt croissant durant les vingt dernières années, intérêt motivé par de nombreuses applications. En effet, les progrès techniques récents permettent d'enregistrer des données sur des grilles de plus en plus fines. Typiquement, les données sont récoltées sous la forme suivante $(X_i(t_{i,1}); \dots; X_i(t_{i,p}))_{1 \leq i \leq n}$ où $(t_{i,1}; \dots; t_{i,p})$ est une suite ordonnée (par exemple une discrétisation temporelle) et n la taille de l'échantillon.

L'apport de l'analyse des données fonctionnelles consiste à traiter ce type de données comme une séquence de réalisations $(X_i)_{1 \leq i \leq n}$ d'une variable aléatoire à valeurs dans un espace de fonctions et observée à certains instants. L'intérêt de cette approche est double : d'une part, il est assez fréquent que p soit très grand ($p \gg n$) ce qui rend ce type de données très difficiles (voire impossibles) à traiter avec les méthodes classiques de la statistique multivariée. D'autre part, il arrive souvent en pratique que les données ne soient pas récoltées sur la même grille, par exemple lorsque certaines observations sont manquantes. L'analyse des données fonctionnelles offre un cadre théorique et numérique utile pour résoudre ce genre de problème.

3.2 Estimation de la fonction de répartition

3.2.1 Estimation ponctuelle

Rappelons que la fonction de répartition empirique \mathbb{F}_n de l'échantillon est définie par :

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} = \text{pourcentage d'observations inférieures à } x.$$

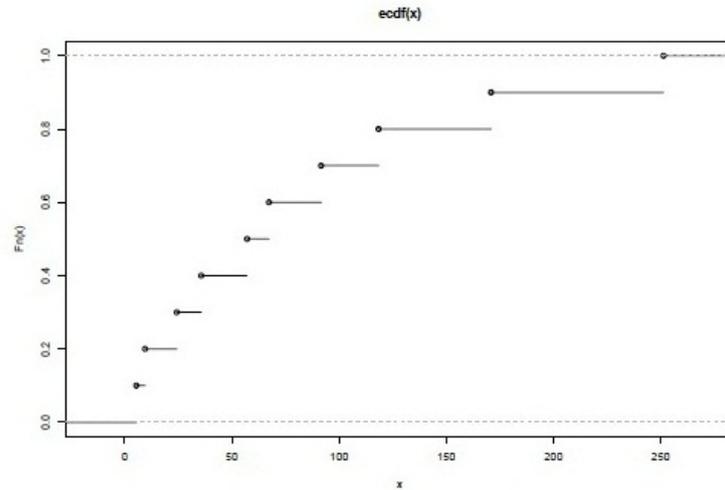


Figure 3.1-Fonction de répartition empirique

Il s'avère que \mathbb{F}_n est un excellent estimateur de F , ce que l'on peut montrer en plusieurs étapes.

Propriété 3.2.1. $\forall x \in \mathbb{R}$, $n\mathbb{F}_n(x)$ est de loi binomiale $\mathcal{B}(n, F(x))$.

Démonstration : $n\mathbb{F}_n(x) = \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}$ est une somme de n variables

aléatoires indépendantes et de même loi de Bernoulli de paramètre $P(X_i \leq x) = F(x)$, donc c'est une variable aléatoire de loi $\mathcal{B}(n, F(x))$.

On peut dire aussi que $n\mathbb{F}_n(x)$ est le nombre de X_i inférieurs à x , qui peut s'interpréter comme le nombre de fois où, en n expériences identiques et indépendantes, un évènement de probabilité $P(X_i \leq x) = F(x)$ s'est produit. Donc c'est une variable aléatoire de loi $\mathcal{B}(n, F(x))$.

On en déduit facilement les qualités de l'estimation de $F(x)$ par $\mathbb{F}_n(x)$.

Propriété 3.2.2. $\forall x \in \mathbb{R}$, $\mathbb{F}_n(x)$ est un estimateur sans biais et convergent en moyenne quadratique de $F(x)$.

Démonstration : $\mathbb{E}(\mathbb{F}_n(x)) = \frac{1}{n} \mathbb{E}(n\mathbb{F}_n(x)) = \frac{1}{n} nF(x) = F(x)$

$$\begin{aligned} \text{Var}(\mathbb{F}_n(x)) &= \frac{1}{n^2} \text{Var}(n\mathbb{F}_n(x)) = \frac{1}{n^2} nF(x)(1 - F(x)) \\ &= \frac{F(x)(1-F(x))}{n} \end{aligned}$$

qui tend vers 0 quand n tend vers l'infini.

En fait, la convergence est presque sûre :

Propriété 3.2.3. $\forall x \in \mathbb{R}$, $\mathbb{F}_n(x) \xrightarrow{PS} F(x)$

Démonstration : Il suffit d'appliquer la loi des grands nombres aux variables aléatoires de loi de Bernoulli $\mathbb{1}_{\{X_i \leq x\}}$:

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \xrightarrow{PS} \mathbb{E} \left(\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \right) = F(x)$$

Vue sous cet angle, la loi des grands nombres dit que la probabilité d'un évènement est la limite de la fréquence d'occurrence de cet évènement dans une suite d'expériences identiques et indépendantes. On en déduit que l'on peut estimer la probabilité que X soit inférieure à x , $F(x)$, par le pourcentage d'observations inférieures à x , $\mathbb{F}_n(x)$. Cette estimation est d'excellente qualité.

Pour juger de la qualité globale de l'estimation de F par $\mathbb{F}_n(x)$, on utilise le théorème de Glivenko-Cantelli, qui dit que \mathbb{F}_n est un estimateur convergent uniformément et presque sûrement de F :

Théorème 3.2.1. Théorème de Glivenko-Cantelli.

$$D_n = \sup\{|\mathbb{F}_n(x) - F(x)|; x \in \mathbb{R}\} \xrightarrow{PS} 0.$$

Par ailleurs, l'erreur quadratique moyenne intégrée est :

$$EQMI(\mathbb{F}_n) = \int_{-\infty}^{+\infty} \text{Var}(\mathbb{F}_n(x)) dx = \frac{1}{n} \int_{-\infty}^{+\infty} F(x)(1 - F(x)) dx.$$

On ne peut pas calculer explicitement cette erreur, mais on sait qu'elle tend vers 0 quand n tend vers l'infini à la vitesse $1/n$.

3.2.2 Intervalle de confiance

Soit x fixé. Un intervalle de confiance de seuil α pour $F(x)$ est un intervalle aléatoire I tel que $\mathbb{P}(F(x) \in I) = 1 - \alpha$. on va chercher un intervalle de confiance de la forme $I = [\mathbb{F}_n(x) - a_\alpha; \mathbb{F}_n(x) + a_\alpha]$, où a_α est déterminé en écrivant :

$$\begin{aligned} \mathbb{P}(F(x) \in I) &= \mathbb{P}(\mathbb{F}_n(x) - a_\alpha \leq F(x) \leq \mathbb{F}_n(x) + a_\alpha) \\ &= \mathbb{P}(F(x) - a_\alpha \leq \mathbb{F}_n(x) \leq F(x) + a_\alpha) \\ &= \mathbb{P}(n(F(x) - a_\alpha) \leq n\mathbb{F}_n(x) \leq n(F(x) + a_\alpha)) \\ &= \sum_{k=\lfloor n(F(x)-a_\alpha) \rfloor + 1}^{\lfloor n(F(x)+a_\alpha) \rfloor} C_n^k [F(x)]^k [1 - F(x)]^{n-k} \\ &= 1 - \alpha \end{aligned}$$

On ne peut pas déduire la valeur a_α de de cette expression car elle implique $F(x)$, qui est inconnue.

Chapitre 4

Exemples sur les développements des noyaux

4.1 Développements pour les noyaux symétriques

Exemple 1 : Moyenne quadratique intégrée, loi $\mathcal{N}(0,1)$, noyau unité :

$$K(y) = 1 \text{ pour } |y| < \frac{1}{2}, \quad [[y^{2p}K]] = \frac{(\frac{1}{2})^{2p}}{2^{p+1}}, \quad [[K^2]] = [[K]] = 1,$$

$$\begin{aligned} \sum_{p=1}^{r-1} C_{2r}^{2p} [[y^{2p}K]] [[y^{2r-2p}K]] &= \frac{(\frac{1}{2})^{2r}}{(2r+1)(2r+2)} \sum_{p=1}^{r-1} C_{2r+1}^{2p+1} \\ &= \frac{1}{2r+1} \left(\frac{1}{r+1} - \left(\frac{1}{2}\right)^{2r-1} \right) \end{aligned}$$

$$\sum_{p=0}^r C_{2r}^{2p} [[y^{2p}K]] [[y^{2r-2p}K]] = \frac{1}{(2r+1)(r+1)}$$

d'où

$$\begin{aligned} B_1^2 &= \sum_{r=2}^{\infty} \delta_n^{2r} \left(\frac{(-1)^r (\frac{1}{2})^{2r+1}}{r!(2r+1)\sqrt{\pi}} \right) \left(\frac{1}{r+1} - \left(\frac{1}{2}\right)^{2r-1} \right) \\ B_2^2 &= \frac{1}{n\delta_n} - \frac{1}{n\sqrt{\pi}} \sum_{r=0}^{\infty} \left(\delta_n^{2r} \frac{(-1)^r (\frac{1}{2})^{2r}}{r!(2r+1)(2r+2)} \right) \end{aligned} \quad (4.1)$$

Ces expressions sont valables pour tout $\delta_n > 0$.

Exemple 2 : Moyenne quadratique intégrée, loi $\mathcal{N}(0,1)$, noyau normal :

$$K(y) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2}, \quad [[y^{2p}K]] = \left(\frac{1}{2}\right)^p \frac{(2p)!}{p!}, \quad [[K^2]] = \frac{1}{2\sqrt{\pi}},$$

$$\sum_{p=1}^{r-1} C_{2r}^{2p} [[y^{2p}K]] [[y^{2r-2p}K]] = \frac{(2r)!}{2^r r!} \sum_{p=1}^{r-1} C_r^p = \frac{(2r)!}{r!} \left(1 - \left(\frac{1}{2}\right)^{r-1}\right)$$

$$\sum_{p=0}^r C_{2r}^{2p} [[y^{2p}K]] [[y^{2r-2p}K]] = \frac{(2r)!}{r!}$$

d'où

$$\begin{aligned} B_1^2 &= \sum_{r=2}^{\infty} \delta_n^{2r} \left(\frac{(-1)^r \left(\frac{1}{2}\right)^{2r+1} (2r)!}{(r!)^2 \sqrt{\pi}} \right) \left(1 - \left(\frac{1}{2}\right)^{r-1}\right) \\ &= \frac{1}{2\sqrt{\pi}} \left(\frac{1}{\sqrt{1+\delta_n^2}} - \frac{2}{\sqrt{1+\frac{1}{2}\delta_n^2}} + 1 \right) \\ B_2^2 &= \frac{1}{n\delta_n(2\sqrt{\pi})} \left(1 - \delta_n \sum_{r=0}^{\infty} \delta_n^{2r} \frac{(-1)^r (2r)!}{2^{2r} (r!)^2} \right) \\ &= \frac{1}{n\delta_n(2\sqrt{\pi})} \left(1 - \frac{\delta_n}{\sqrt{1+\delta_n^2}} \right), \quad \text{si } \delta_n > 0 \end{aligned} \tag{4.2}$$

Ces développements sont valables si $|\delta_n| < 1$. Une argumentation portant sur l'analyticité de B_1^2 et B_2^2 en fonction de δ_n montre que les formules sommées ont une validité étendue à $\delta_n > 0$.

Remarquons que dans ce cas, un calcul direct montre que :

$$\mathbb{E}(f_n(x)) = \frac{1}{\sqrt{\delta_n^2 + 1} \sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2 \frac{1}{\delta_n^2 + 1}\right). \tag{4.3}$$

L'espérance de f_n est une densité normale $\mathcal{N}(0, \delta_n^2 + 1)$.

Exemple 3 : Moyenne quadratique intégrée, loi $\mathcal{N}(0,1)$, noyau d'épa-

nechnikov :

$$K(y) = \frac{3}{4}(1 - y^2), |y| < 1; [[y^{2p}K]] = \frac{3}{(2p+1)(2p+3)}, [[K^2]] = \frac{3}{5},$$

$$\sum_{p=1}^{r-1} C_{2r}^{2p} [[y^{2p}K]] [[y^{2r-2p}K]] = \frac{6}{(2r+1)(2r+3)} \left(\frac{2^{2r}3}{(r+2)(r+3)} - 1 \right)$$

$$\sum_{p=0}^r C_{2r}^{2p} [[y^{2p}K]] [[y^{2r-2p}K]] = \frac{9 \cdot 2^{2r+1}}{(r+2)(r+3)(2r+1)(2r+3)},$$

$$B_1^2 = \sum_{r=2}^{\infty} \delta_n^{2r} \left(\frac{9(-1)^r}{r!(2r+1)(2r+3)\sqrt{\pi}} \right) \left(\frac{1}{(r+2)(r+3)} - \frac{1}{3} \left(\frac{1}{2} \right)^{2r} \right)$$

$$B_2^2 = \frac{1}{n\delta_n} \left(\frac{3}{5} - \frac{\delta_n}{\sqrt{\pi}} \sum_{r=0}^{\infty} \delta_n^{2r} \left(\frac{9(-1)^r}{r!(r+2)(r+3)(2r+1)(2r+3)} \right) \right) \quad (4.4)$$

Ces développements sont valables pour tout $\delta_n > 0$.

Exemple 4 : Moyenne quadratique intégrée, loi $\mathcal{N}(0,1)$, noyau triangulaire :

$$K(y) = 1 - |y|, |y| < 1; [[y^{2p}K]] = \frac{2}{(2p+1)(2p+2)}, [[K^2]] = \frac{2}{3},$$

$$\sum_{p=1}^{r-1} C_{2r}^{2p} [[y^{2p}K]] [[y^{2r-2p}K]] = \frac{4}{(2r+1)(2r+2)(2r+3)(2r+4)} (2^{2r+3} - (2r+4)(2r+3) - 2)$$

$$\sum_{p=0}^r C_{2r}^{2p} [[y^{2p}K]] [[y^{2r-2p}K]] = \frac{4}{(2r+1)(2r+2)(2r+3)(2r+4)} (2^{2r+3} - 2)$$

d'où

$$B_1^2 = \sum_{r=2}^{\infty} \delta_n^{2r} \left(\frac{4(-1)^r}{(r+2)!(2r+1)(2r+3)\sqrt{\pi}} \left(1 - \left(\frac{1}{2} \right)^{2r+3} (2r+3)(2r+4) - \left(\frac{1}{2} \right)^{2r+2} \right) \right)$$

$$B_2^2 = \frac{1}{n\delta_n} \left(\frac{2}{3} - \frac{\delta_n}{\sqrt{\pi}} \sum_{r=0}^{\infty} \delta_n^{2r} \left(\frac{4(-1)^r}{(r+2)!(2r+1)(2r+3)} \right) \left(1 - \left(\frac{1}{2} \right)^{2r+2} \right) \right) \quad (4.5)$$

Ces développements sont valables pour tout $\delta_n > 0$.

4.2 Développements pour les noyaux de Gram-charlier

Exemples : 1) Noyau de Gram-Charlier d'ordre 1 :

Ce noyau est donné par :

$$K(y) = \frac{1}{2}(3 - y^2)(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2}, \quad [[K]] = 1,$$

$$[[K^2]] = \frac{27}{32\sqrt{\pi}}, \quad [[y^{2n}K]] = -\frac{(2n)!}{2^n n!} (n-1) \quad (4.6)$$

2) Noyau de Gram-Charlier d'ordre 2 :

$$K(y) = \frac{1}{8}(y^4 - 10y^2 + 15) \frac{e^{-\frac{1}{2}y^2}}{\sqrt{2\pi}},$$

$$[[y^{2p}K]] = \left(\frac{1}{2} \right)^{p+1} \frac{(2p)!}{p!} (p-1)(p-2), \quad [[K^2]] = \frac{2265}{2048\sqrt{\pi}}; \quad (4.7)$$

Les propriétés asymptotiques de ces noyaux sont discutées plus loin [2.2.7](#).

4.3 Développements pour les noyaux de Legendre

Exemples : 1) Noyau de Legendre d'ordre 1 :

$$K(y) = \frac{3}{8}(3 - 5y^2) = \frac{1}{2}P_0(y) - \frac{5}{4}P_2(y), \quad |y| < 1 \quad (4.8)$$

$$[[K]] = 1, \quad [[y^{2n+1}K]] = 0, \quad \forall n \geq 1, \quad [[y^2K]] = 0, \quad [[K^2]] = 9/8$$

$$[[y^{2n}K]] = -\frac{3(n-1)}{(2n+1)(2n+3)} \quad \text{pour } n \geq 1;$$

asymptotiquement lorsque $\delta_n \rightarrow 0$,

$$B_1^2 \sim \frac{\delta_n^8}{(4!)^2} \frac{9}{1225} \left(\int_I (f^{(4)}(x))^2 dx \right), \quad B_2^2 \sim \frac{9}{8n\delta_n}$$

Ces formules sont rapprocher de celles qu'on obtient pour le noyau unité ; en utilisant la même échelle, c'est-à-dire ici $K'(y) = \frac{1}{2}$ si $|y| < 1$, on obtient le noyau de Legendre d'ordre 0, qui fournit asymptotiquement :

$$B_1^2 \sim \frac{\delta_n^4}{(2!)^2} \frac{1}{9} \left(\int_I (f''(x))^2 dx \right), \quad B_2^2 \sim \frac{1}{2n\delta_n}$$

Dans le premier cas, l'erreur quadratique minimale est obtenue pour :

$$\delta_n = \left(\frac{5^2 7^2 3^2}{\int_I (f^{(4)}(x))^2 dx} \right)^{\frac{1}{9}} n^{-1/9},$$

d'où

$$M^2 \sim \frac{81}{64n\delta_n} = \frac{81}{64} \left(\frac{\int_I (f^{(4)}(x))^2 dx}{5^2 7^2 3^2} \right)^{\frac{1}{9}} n^{-8/9},$$

Dans le deuxième cas

$$\delta'_n = \left(\frac{9}{2 \int_I (f''(x))^2 dx} \right)^{\frac{1}{5}} n^{-1/5},$$

d'où

$$M'^2 \sim \frac{5}{8n\delta'_n} = \frac{5}{8} \left(\frac{2 \int_I (f''(x))^2 dx}{9} \right)^{\frac{1}{5}} n^{-4/5},$$

Pour la loi normale $\mathcal{N}(0,1)$

$$M^2 = \frac{5}{8} \left(\frac{1}{12\sqrt{\pi}} \right)^{\frac{1}{5}} n^{-4/5}, \quad M^2 = \frac{81}{64} \left(\frac{1}{5.7.3.2^5\sqrt{\pi}} \right)^{\frac{1}{9}} n^{-8/9}$$

En considérant les formules asymptotiques précédentes et en négligeant les termes d'ordre supérieur. Cette remarque prouve pleinement l'intérêt des noyaux de ce type pour le critère du MISE; il faut cependant noter que, pour une distribution différente de la loi normale, ou pour l'estimation de la densité en un point, ces noyaux auront des performances variables suivant les cas.

2) Noyau de Legendre d'ordre 2 :

$$K(y) = \frac{15}{128}(15 - 70y^2 + 63y^4) = \frac{1}{2}P_0(y) - \frac{5}{4}P_2(y) + \frac{27}{16}P_4(y), \quad |y| < 1,$$

$$[[K]] = 1, \quad [[y^{2n+1}K]] = 0, \quad n \geq 1,$$

$$[[y^2K]] = 0, \quad [[y^4K]] = 0, \quad [[K^2]] = \frac{7425}{4096}$$

$$[[y^{2n}K]] = \frac{15(n-1)(n-2)}{2(2n+1)(2n+3)(2n+5)} \quad \text{pour } n \geq 1;$$

asymptotiquement lorsque $\delta_n \rightarrow 0$,

$$B_1^2 \sim \frac{\delta_n^{12}}{11^2 \cdot 7^2 \cdot 3^6 \cdot 2^8} \left(\int_I (f^{(6)}(x))^2 dx \right), \quad B_2^2 \sim \frac{11 \cdot 5^2 \cdot 3^3}{2^{12} n \delta_n}$$

Asymptotiquement l'erreur quadratique minimale est obtenue pour :

$$\delta_n = \left(\frac{11^3 \cdot 7^2 \cdot 5^2 \cdot 3^8}{2^6 \int_I (f^{(6)}(x))^2 dx} \right)^{\frac{1}{13}} n^{-\frac{1}{13}}$$

d'où

$$M^2 \sim \frac{13 \cdot 11 \cdot 5^2 \cdot 3^2}{2^{14}} \left(\frac{2^6}{11^3 7^2 5^2 3^8} \int_I (f^{(6)}(x))^2 dx \right)^{\frac{1}{13}} n^{-\frac{12}{13}}$$

pour la loi normale $\mathcal{N}(0,1)$

$$M^2 \sim \frac{13 \cdot 11 \cdot 5^2 \cdot 3^2}{2^{14}} \left(\frac{1}{11^2 \cdot 7 \cdot 5 \cdot 3^5 \cdot 2 \sqrt{\pi}} \right)^{\frac{1}{13}} n^{-\frac{12}{13}}$$

La même remarque que précédemment peut être appliquée à ce noyau.

Conclusion générale et perspectives

Conclusion générale

Nous nous sommes intéressés plus particulièrement dans ce travail un modèle non paramétrique qui traite le cas des variables réelles et fonctionnelles. L'objectif était l'estimation de la fonction de densité au moyen de l'histogramme généralisé.

La richesse de ce domaine de recherche de la statistique fonctionnelle offre de nombreuses perspectives tant du point de vue théorique que pratique. Dans ce qui suit, on va commenter quelques résultats déjà obtenus, avec comme souci majeur de mettre l'accent sur tous les problèmes ouverts dont certains sont en cours de développement.

Perspectives

Le travail développé dans ce mémoire offre de nombreuses perspectives à court et long termes. Concernant les perspectives à court terme :

- On peut envisager tout en faisant une adaptation des outils développés par Niang et Rhomari (2003) pour étudier la convergence en norme L^p de nos estimateurs dans le cas ergodique.
- Une autre perspective envisageable est d'obtenir des vitesses de convergence et la formule du paramètre de lissage à l'aide de l'erreur quadratique et l'erreur quadratique intégré dans le cas spatial.

- D'autres thèmes peuvent être abordés à long terme tels le conditionnement par p variables fonctionnelles ou une combinaison linéaire de ces p variables fonctionnelles.
- L'étude des propriétés asymptotiques de l'estimateur à noyau de la densité conditionnelle pour des variables réponse et explicatives fonctionnelle.

Bibliographie

- [1] S. Abou-Jaoudé. Sur une condition nécessaire et suffisante de l_1 convergence presque complète de l'estimateur de la partition fixe pour une densité. C.R. Acad. Paris Sér. A-B, 283(16 :AII) :A1107-A1110, 1976.
- [2] Bickel-Rosenblatt(1973) : On some global measures of the deviates of density function estimates. *Ann. statist.*, 1, p. 1071-95.
- [3] Bosq Denis et Lecoutre Jean-Pierre : Théorie de l'estimation fonctionnelle, Economica, Paris (1987).
- [4] Cencov (1962) : Estimation of an unknown distribution density from observations. *Sovet. Math.*, 3, p.1559-62.
- [5] Epanechnikov(1969) :Nonparametric estimates of a multivariate probability density, *Teor. Prob. Appl.*, 14, p. 153-80.
- [6] Ferraty, F. et Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford Handbooks in Mathematics. OUP Oxford (cf. p. 3).
- [7] Ferraty, F. et Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Theory and practice. New York : Springer (cf. p. 3, 79, 97).
- [8] J. Geoffrey. Sur l'estimation d'une densité dans un espace métrique. C.R. Acad. Paris Sér. A, 278 :1449-1452, 1974.
- [9] P. Hall and J. S. Marron. Estimation of integrated squared density derivatives. 6(2) :109-115, 1987.

-
- [10] M. C. Jones and S. J. Sheather. Using non-stochastic terms to advantage in kernelbased estimation of integrated squared density derivatives. 11(6) :511-514, 1991.
- [11] J. Lecoutre. *Contribution à l'estimation non paramétrique de la régression*. PhD thesis, Université de Pierre et Marie Curie-ParisVI-France, 1982.
- [12] R.Moché :Estimation de la densité par les histogrammes et les estimateurs à noyau (10 juillet 2006).
- [13] B.Nehme :Techniques non-additives d'estimation de la densité de probabilité (2010).
- [14] Parzen (1962) : On the estimation of a probability density function and the mode. *Ann. Math. Statist.*, 33, p. 1065-76.
- [15] Prakasa-Rao (1969) : Estimation of a unimodal density. *Sankhya*, Ser. A, 31, p. 26-36.
- [16] Ramsay, J. et Silverman, B. (2005). *Functional Data Analysis*. anglais. 2e éd. Springer Series in Statistics. Springer (cf. p. 3, 5, 6, 15, 16, 32, 40, 41, 78, 152, 156).
- [17] Robertson (1967) : On estimating a density which is measurable with respect to a lattice. *Ann. Math. statist.*, 38, p.482-93.
- [18] Rosenblatt (1956) :Remarks on some nonparametric estimates of a density. *Ann. Math. Statist.*, 33 p. 1065-76.
- [19] Schwarz-stuart (1967) : Estimation of a probability density by orthogonale series. *Ann. Math. statist.*, 38, p. 1261-5.
- [20] J. S. Simonoff. *Smoothing Methods in Statistics*. Springer-Verlag, 1996.
- [21] Van Ryzin (1966) : Bayes risk consistency of classification procedures using density estimation. *Sankhya*, Ser. A,28, p. 261-70.
- [22] Wahba (1971) : A polynomial algorithm for density estimation. *Ann. Math. statist.*, 42, p. 1496-8.

- [23] Watson (1969) : Density estimation by orthogonal series. *Ann. Math. Statist.*, 40, p.1496-8.
- [24] Wegman (1969) : A note on estimating a unimodal density.