

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderrahmane Mira de Béjaïa
Faculté des Sciences Exactes
Département Mathématiques



جامعة بجاية
Tasdawit n Bgayet
Université de Béjaïa

Mémoire de fin de cycle

En vue de l'obtention du diplôme de Master en Mathématiques
Option : Probabilités Statistique et Applications

Thème

Sur l'estimation non paramétrique de la densité et du mode.

Réalisé par :

M^r DEHIMI Aimen

Devant le Jury composé de :

M ^r SOUFIT	Massinissa	Président	M.C.B	Université A.Mira de Béjaïa
M ^{me} SAADI	Nora	Promotrice	M.C.A	Université A.Mira de Béjaïa
M ^{me} LAGHA	Karima	Examinatrice	M.C.A	Université A.Mira de Béjaïa

Année universitaire : 2020 / 2021

Remerciements

Au terme de mon modeste travail, je tiens à remercier tous d'abord Dieu, le tout puissant de m'avoir accordé la volonté, force, santé, détermination et le courage afin d'accomplir mon mémoire.

*Je tiens à exprimer ma profonde gratitude à ma promotrice M^{me}. **SAADI** pour son encadrement. Je suis également reconnaissante pour la confiance qu'elle m'a accordée, sa disponibilité et son engagement. Ses remarques constructives m'ont beaucoup apporté tout au long de ce travail.*

*Je tiens vivement à remercier M^{me}. **Lagha** pour l'immense honneur qu'elle ma fait en décidant de présider le jury de ce mémoire.*

*Mes remerciements chaleureux s'adresse également à M^{me}. **BARECHE** d'avoir accepté d'examiner et d'évaluer ce modeste travail.*

Je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste mémoire.

Table des matières

1	Estimation non paramétrique de la densité de probabilité	9
1.1	Introduction	9
1.2	Estimation non paramétrique par Histogramme	9
1.2.1	Propriétés statistiques de l'histogramme	11
1.2.2	Choix du paramètre h	12
1.3	Estimation non paramétrique de la densité par des séries orthogonales	12
1.3.1	Propriétés statistiques de l'estimateur	13
1.3.2	Choix pratique de la base	14
1.3.3	Exemples	14
1.3.4	Choix pratique du paramètre de lissage	17
1.3.5	Méthode de Kronmal-Tarter	17
1.3.6	Méthode de Bosq	18
1.4	Estimateur à noyau	19
1.4.1	Noyaux usuels symétriques	20
1.4.2	Etude du biais et de la variance	21
1.4.3	Risque quadratique ponctuel et risque quadratique intégré	24
1.4.4	Noyau d'ordre l	26
1.5	Les méthodes de sélection du paramètre de lissage	27
1.6	Méthodes plug-in	28
1.6.1	Rule of Thumb	28
1.6.2	Plug-in itéré	29
1.6.3	Méthodes cross validation (validation croisée)	30
1.7	Conclusion	32
2	Estimation non paramétrique du mode	33
2.1	Introduction	33
2.2	Le mode d'une variable aléatoire de loi de probabilité	33
2.2.1	Mode d'une variable aléatoire de loi continue usuelles	35
2.2.2	Mode d'une variable aléatoire de loi discrète usuelles	36
2.3	Estimation non paramétrique du mode	36
2.4	Méthode directe	37

2.4.1	Méthode indirecte	37
2.5	Conclusion	43
3	Simulation	44
3.1	Introduction	44
3.2	Méthode des séries orthogonales (Méthode indirecte)	44
3.2.1	Plan de simulation	44
3.2.2	Algorithme	45
3.2.3	Interprétation	46
3.3	Estimateur à noyau du mode	46
3.3.1	Plan de simulation	46
3.3.2	Algorithme	46
3.4	La méthode de Grenander (Méthode directe)	48
3.4.1	Algorithme	48
3.5	Étude comparative	49
	Conclusion générale et perspectives	51
	Bibliographie	53

Principales notations

Nous utilisons les notations suivantes.

\mathbb{R} : Ensemble des nombres réels ;

X : Une variable aléatoire ;

$\mathbb{E}[X]$: L'espérance de variable aléatoire X ;

$\mathbb{V}[X]$: Variance de la variable aléatoire X ;

$\mathbf{1}_A$: Fonction indicatrice de l'ensemble A ;

f : Densité de probabilité de X ;

F : Fonction de répartition associée à la densité f ;

F_n : La fonction de répartition empirique.

f_h : Estimateur de la densité f par la méthode du noyau ;

\hat{f}_{d_n} : Estimateur de la densité f par la méthode des séries orthogonales ;

MSE : L'erreur quadratique moyenne ;

$MISE$: L'erreur quadratique moyenne intégrée ;

$AMISE$: L'erreur quadratique moyenne intégrée asymptotique.

Liste des tableaux

2.1	Mode d'une variable aléatoire de loi Continue.	36
2.2	Mode d'une variable aléatoire de loi discrète.	36
3.1	Le paramètre de lissage optimal et la Variation du $\mathbf{MSE}(\hat{x}_0)$ en fonction de n pour la Loi Beta $\beta(2, 8)$, en utilisant la base saadi-Adjabi et la base d'Efromovich.	46
3.2	Le paramètre de lissage optimal et la Variation du $\mathbf{MSE}(\hat{x}_0)$ optimal en fonction de n pour la Loi normale $\mathcal{N}(0, 1)$, en utilisant le noyau triangulaire et le noyau cosine.	48
3.3	Variation du $\mathbf{MSE}(\hat{x}_0)$ optimal en fonction de n pour la Loi $\beta(2, 8)$, en utilisant l'estimateur de Grenander.	49
3.4	Variation du $\mathbf{MSE}(\hat{x}_0)$ pour la loi $\beta(2, 8)$, en utilisant l'estimateur de Grenander et l'estimateur de saadi-adjabi.	50
3.5	Variation du $\mathbf{MSE}(\hat{x}_0)$ pour la loi $G(2; 4)$, en utilisant l'estimateur de Grenander et l'estimateur à noyau (gaussien).	50

Table des figures

1.1	Histogramme construit avec une largeur $h=0.2$	11
1.2	Noyaux usuels	21
2.1	Détermination graphique du mode de loi Normal $N(0, 1)$	34
2.2	Détermination graphique du mode de loi Gamma $G(2, 1)$	35

Introduction

Un des plus vieux problèmes de la statistique non paramétrique consiste à estimer la densité de probabilité à partir d'un échantillon observé. Il s'agit d'un problème fondamental qui a connu, durant ces dernières années, des développements théoriques et pratiques à la fois rapides et nombreux.

Une des premières analyses consacrées à ce sujet est due au biométricien Karl Pearson, il y a. Dans son article de 1902, une approche paramétrique consiste à supposer que la densité de probabilité f appartient à une famille de densités qui peuvent être décrites par un petit nombre (connu) de paramètres réels, l'estimation de la densité se réduit alors à un problème d'estimation de paramètres.

le champ d'application d'un modèle paramétrique n'est satisfaisant que lorsque l'inflation du nombre de paramètres est telle que les méthodes d'estimation du modèle deviennent tout à fait inefficaces. Pour pallier les insuffisances et les défauts des familles paramétriques, une seconde approche dite non paramétrique propose d'estimer la densité à partir de l'information disponible. On dit souvent que dans cette approche les données parlent d'elles mêmes. L'avantage principal de l'estimation non paramétrique de la densité est de ne pas nécessiter d'hypothèses à priori sur l'appartenance de cette densité à une famille de lois connues. L'outil d'estimation non paramétrique nous est fourni par l'histogramme : une fois les données regroupées en classes de valeurs, les fréquences empiriques sont représentées par des aires rectangulaires dont les bases correspondent aux classes elles mêmes.

Il existe d'autres méthodes non paramétriques plus robustes que la méthode par histogramme : la méthode d'estimation par des séries orthogonales et la méthode du noyau.

Dans le premier chapitre, nous nous intéressons à l'estimation non paramétrique de la densité de probabilité par des séries orthogonales introduite par Cencov (1962) . Nous nous sommes placés, dans le cas où on dispose de réalisations de variables (X_1, \dots, X_n) aléatoires indépendantes et identiquement distribuées (i.i.d.) . Cencov a considéré la famille des estimateurs de la densité f de la forme :

$$\hat{f}_{d_n}(x) = \frac{1}{n} \sum_{k=0}^{d_n} \sum_{i=1}^n e_k(X_i) e_k(x), x \in \mathbb{R}.$$

avec $(d_n) \rightarrow n$ lorsque $n \rightarrow \infty$ et où $\{e_k\}_{k \in \mathbb{N}}$ est un système orthogonal dans $L^2(\mathbb{R})$. Cette méthode est très utilisée dans de nombreux domaines (analyse harmonique, traitement du signal, compression d'images, statistique fonctionnelle...). A la suite de Cencov,

de nombreux auteurs ont contribué à améliorer cette méthode. Les principales études ont concerné le choix de la base orthogonale et celui du paramètre de lissage (nombre de termes de la série orthogonale). Sur le premier point, Efromovich (1997) s'est basé sur les travaux de Krylov (1955) concernant les systèmes orthonormaux. On trouvera de nombreux autres exemples de bases possibles dans les travaux de Sanson (1959) ou Kolmogorov et Fomin (1957). Le choix du paramètre de lissage fut souvent choisi de manière à minimiser un estimateur de l'espérance quadratique intégrée Hart (1985), Diggle et Hall (1986) ou Tarter et Lock (1993)). On peut également citer les travaux de Devroye et Györfi (1985) ou Devroye (1987) qui s'intéressent à un autre critère d'erreur basé sur la norme L_1 . Pour compléter cette bibliographie, on peut se référer aux ouvrages de Prakasa (1983), Nadaraya (1989), Hart (1997), Thompson et Tapia (1990) ou plus récemment Bosq (2002).

La troisième partie de ce chapitre s'intéresse à l'estimation non paramétrique de la densité de probabilité par la méthode du noyau introduite par Rosenblatt en 1956, suivi de Parzen en 1962, qui ont proposé une classe d'estimateurs à noyau d'une densité univariée. Les estimateurs à noyau sont fonction de deux paramètres K , appelé noyau, et h dit paramètre de lissage (largeur de fenêtre). Les propriétés de convergence de l'estimateur à noyau ont été établies par Parzen (1962) et Nadaraya (1964). Devroye en 1983 a fait une étude complète sur la convergence L_1 . Les théorèmes relatifs à l'erreur quadratique moyenne et l'erreur quadratique intégrée moyenne ont été obtenus sous forme élémentaire par Parzen (1962). Enfin, c'est Epanechnikov en (1969) qui s'est rendu compte de l'existence d'un noyau asymptotiquement optimal K_E .

Estimer le mode est souvent une conséquence directe de l'estimation de la densité. Son importance est due au fait que c'est une mesure naturelle de tendance centrale, qui n'est pas influencée par les queues des distributions. Le mode est la valeur la plus probable : pour une densité de probabilité f , c'est la valeur pour laquelle f admet un maximum. Pour une distribution symétrique, il coïncide avec deux autres paramètres de position, la moyenne et la médiane. On peut distinguer plusieurs estimateurs non paramétriques du mode : on en citera quelques uns qui sont définis selon deux approches :

- La Méthode indirecte : consiste à obtenir dans un premier temps une estimation de la densité et de prendre pour mode de valeur de x pour laquelle $f(x)$ est maximum. Parzen (1962) a été l'un des premiers à s'intéresser au problème de l'estimation du mode x_0 dans le cas d'une densité univariée. Il définit un estimateur comme la variable aléatoire qui maximise l'estimateur à noyau f_n de la densité :

$$x_0 = \arg \max_{x \in \mathbb{R}} f_n(x).$$

Il démontre que cet estimateur est uniformément convergent (en probabilité), asymptotiquement normal et donne une évaluation de l'erreur quadratique moyenne (MSE). Nadaraya (1965) en établit la consistance forte. D'autres études ont affaibli les conditions suffisantes de normalité asymptotique (Eddy (1980), (1982), Romano (1988)), ou établi la convergence dans L_p (Devroye Wagner (1980), Grund Hall (1995)). La liste des travaux est longue, on citera ceux de Van Ryzin (1969), Mokkadem Pelletier (2003), Herrmann

Ziegler (2004) parmi d'autres. Plus récemment, Shi al (2009) améliorent le taux de convergence de l'estimateur du mode, exprimé en fonction de la fenêtre h . D'autres auteurs se sont penchés sur le sujet, parmi lesquels on peut citer :Vieu (1996) qui propose l'étude de quatre estimateurs à noyau du mode, basés sur des estimateurs à noyau de la densité et de sa dérivée.

L'estimateur à noyau du mode introduit par Parzen à été largement étudié, mais, très peu de résultats existent sur l'estimateur basé sur des séries orthogonales. À notre connaissance, les propriétés statistiques et asymptotiques de cet estimateur n'ont pas été étudiées.

- La méthode directe : on procède d'entrée à l'estimateur du mode x_0 en considérant que dans l'échantillon, on doit observer un groupement de valeurs dans le voisinage du mode (Chernoff (1964), Grenander (1965)) : Chernoff (1964) présente l'estimateur noté \hat{x}_{a_n} comme le milieu d'un intervalle de longueur $2a_n$ contenant le maximum d'observations x_1, x_2, \dots, x_n , $(a_n)_n$ étant une suite de réels positifs décroissant lentement vers zéro. Wegman (1971) montre la consistance forte de cet estimateur. Grenander (1965) définit un estimateur à partir d'un échantillon d'observations $(x_1, x_2, , x_n)$ ordonnées par ordre croissant. Bickel (2002) propose deux estimateurs plus robustes appelés : HSM (Half-Sample Mode) obtenu à partir d'algorithmes basés sur des demi-échantillons successifs. HRM (Half-Range Mode), obtenu en cherchant un "intervalle modal" contenu dans d'autres intervalles modaux (le mode sera alors l'étendue du petit intervalle modal).

- Le premier objectif de ce mémoire est de faire le point sur les méthodes d'estimation non paramétrique de la densité de probabilité et d'étudier les différentes méthodes d'estimation du mode, à savoir les méthodes directes et les méthodes indirectes.
- Le second objectif de ce travail est d'appliquer la méthode des fonctions orthogonales pour estimer le mode de la densité.
- Le dernier objectif de ce mémoire est d'appliquer et de comparer les différentes méthodes sur plusieurs distributions connues.

Nous avons structuré notre travail en trois parties principales. Le premier chapitre est consacré à une description des méthodes d'estimation non paramétrique de la densité de probabilité. En particulier, nous nous intéressons à l'estimateur par la méthode du noyau et la méthode des fonctions orthogonales et l'estimation par Histogramme en donnant les propriétés de convergence de ces estimateurs.

Dans le second chapitre, nous nous intéressons à l'estimation non paramétrique du mode. Enfin, dans le chapitre 3, nous présentons les résultats de simulations conduites à partir de plusieurs densités cibles connues : loi normale loi Beta et la loi Gamma.

Ce mémoire se termine par une conclusion générale et quelques propositions d'axes de recherche.

Chapitre 1

Estimation non paramétrique de la densité de probabilité

1.1 Introduction

Dans ce chapitre, nous nous intéressons au problème d'estimation non paramétrique de la densité de probabilité. Il se divise en cinq sections. Dans la première et la deuxième section, nous définissons et nous donnons la forme de l'estimateur de la densité par Histogramme et par projection introduit par Cencov (1962). Nous donnons aussi les propriétés globales de l'estimateur. La troisième section illustre théoriquement certains aspects de l'estimateur à noyau. Cette section propose des critères de choix du paramètre de lissage. Enfin, nous donnons une conclusion dans la dernière section.

1.2 Estimation non paramétrique par Histogramme

Etant données des observations x_1, \dots, x_n qui sont les réalisations des variables aléatoires réelles indépendantes et identiquement distribuées X_1, X_2, \dots, X_n de densité f inconnu sur l'intervalle $[a, b]$. Construire un histogramme consiste à partitionner l'intervalle $[a, b]$ en $p \in \mathbb{N}^*$ classes A_k , $k \in \{1, \dots, p\}$, et à compter le nombre d'observations appartenant à chaque classe A_k . Si toutes les classes de l'histogramme ont la même largeur, on dit que l'histogramme est régulier. On note $h \in \mathbb{R}$, la largeur des classes qui est alors appelée le paramètre de lissage. Le nombre d'observations appartenant à chaque classe A_k est appelé accumulateur de la classe A_k est noté $A_{cck} = \sum_{i=1}^n 1_{A_k}(x_i)$, la probabilité de A_k (basée sur les observations x_i), notée $\mathbb{P}(A_k)$, est donnée par :

$$\mathbb{P}(A_k) = \frac{A_{cck}}{n}. \quad (1.1)$$

En émettant l'hypothèse, que les observations se répartissent uniformément dans la cellule A_k (de largeur h), on peut alors construire un estimateur de la densité f , pour tout $x \in [a, b]$,

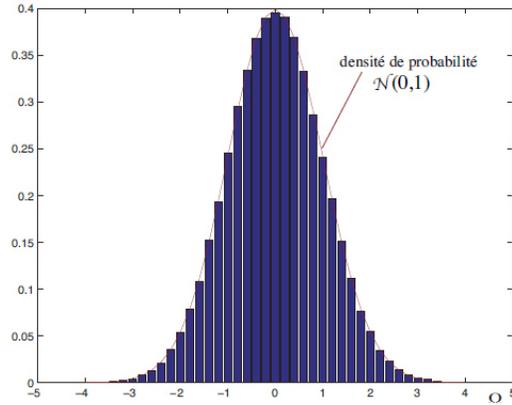
par :

$$f_h(x) = \frac{1}{h} \sum_{k=1}^p \mathbb{P}(A_k) 1_{A_k}(x) = \frac{1}{nh} \sum_{k=1}^p A_k 1_{A_k}(x). \quad (1.2)$$

Dans la suite, nous émettons l'hypothèse, que les classes $A_k \forall k \in \{1, \dots, p\}$ forment une partition de $[a, b]$ et définissons pour chaque classe A_k , son centre a_k telles que

$$\forall k \in \{1, \dots, p\}, A_k = [a_k - \frac{h}{2}, a_k + \frac{h}{2}] \text{ et } \forall k \in 1, \dots, p, a_{k+1} = a_k + h.$$

Remarque 1.2.1. 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 [1.1] 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 $A = [-5, 5]$. La largeur de l'histogramme est $h = 0.2$.

FIGURE 1.1 – Histogramme construit avec une largeur $h=0.2$.

1.2.1 Propriétés statistiques de l'histogramme

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 est donné pour tout $x \in [a_k, a_{k+1}]$ par :

$$\text{Biais}(f_h(x)) = \mathbb{E}(f_h(x)) - f(x) = \frac{1}{2}f'(x)(h - 2(x - a_k)) + O(h^2), \quad (1.3)$$

ou O est un terme résiduel et f' est la dérivée de f . f doit être une fonction de $L^2([a, b])$ absolument continue et carrée intégrable .

- La variance de l'estimateur est donnée pour tout $x \in [a_k, a_{k+1}]$ par :

$$\text{Var}(f_h(x)) = \mathbb{E}(f_h^2(x)) - \mathbb{E}^2(f_h(x)) = \frac{f(x)}{nh} + o(n^{-1}). \quad (1.4)$$

Discussion du comportement du biais et de la variance

- * Le biais décroît si h diminue mais la variance augmente.
- * Pour que la variance tende vers 0, il faut que $nh \rightarrow \infty$.
- * La variance diminue si h augmente mais le biais augmente.

Il s'ensuit que :

$$\text{MSE}(f_h(x)) = \frac{f(x)}{nh} + \frac{f'(x)^2}{4}(h - 2(x - a_k))^2 + O(h^3) + O(n^{-1}), \quad (1.5)$$

$$\text{MISE}(f_h(x)) = \frac{1}{nh} + \frac{h^2 \int f'(t)^2 dt}{12} + O(h^3) + O(n^{-1}). \quad (1.6)$$

Remarque 1.2.2. Dans 1.6, on voit que : un petit h donne un histogramme peu biaisé, tandis qu'un grand h et un grand n déterminent un histogramme moins variable.

1.2.2 Choix du paramètre h

Définition 1.1. (Règle de Scott.) La valeur qui minimise l'erreur quadratique moyenne intégrée, $MISE$ est :

$$h_{opt} = \left[\frac{6}{\int f'(t)^2 dt} \right]^{\frac{1}{3}} n^{-\frac{1}{3}}.$$

Si f la densité de loi normale $\mathcal{N}(\mu, \sigma)$, alors

$$h_{opt} = 3.491\sigma n^{-\frac{1}{3}}.$$

En estimant σ par l'écart-type S de l'échantillon, on obtient ainsi la règle de Scott.

$$h_{opt} = 3.491Sn^{-\frac{1}{3}}.$$

Définition 1.2. (Règle de Sturges.) Prendre le nombre k de classes égal à $1 + \log_2 n$. En pratique, cela revient à prendre $h = \frac{x_{(k)} - x_{(1)}}{k}$, $x_{(i)}$ sont les valeurs d'observations d'un échantillon ordonné par ordre croissant. La règle de Sturges a tendance à produire des histogrammes trop lisses.

Remarque 1.2.3. La plus grande qualité de l'histogramme est sa simplicité. Parmi ses inconvénients importants, citons celui d'être trop peu sensible aux propriétés locales de f . En outre, alors que la plupart des fonctions de densité ne sont pas des fonctions en escalier, l'histogramme est un estimateur toujours de cette forme. L'application de certaines opérations sur l'estimé, comme par exemple une dérivée ou une intégration, devient alors impossible ou très difficile à effectuer. De plus, si on a un paramètre de lissage h trop petit, cela conduit à un histogramme plus découpé, tandis qu'à un paramètre de lissage h trop grand résulte un histogramme plus lissé.

Il existe d'autres méthodes non paramétriques plus robustes que la méthode par histogramme : la méthode d'estimation par des séries orthogonales et la méthode du noyau.

1.3 Estimation non paramétrique de la densité par des séries orthogonales

Définition 1.3. Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes et identiquement distribuées de densité de probabilité f par rapport à la mesure de Lebesgue sur \mathbb{R} , il s'agit d'estimer f à partir des observations x_1, \dots, x_n . Pour cela on suppose que :

1. L'espace de Hilbert \mathbb{L}^2 est de dimension infinie ;

2. $\{e_k, k \in \mathbb{N}\}$ un système orthogornal dans \mathbb{L}^2 ;

3. $f \in \mathbb{L}^2$ tel que :

$$f(x) = \sum_{k=0}^{\infty} a_k e_k(x), k = 0, \dots, x \in \mathbb{R}; \quad (1.7)$$

4. Avec $\{a_k, k \in \mathbb{N}\}$ sont les coefficients de Fourier associés à f donnés par :

$$a_k = \int_{\mathbb{R}} e_k(x) f(x) dx = \mathbb{E}[e_k(X)], k = 0, 1, \dots \quad (1.8)$$

5. Considérant un sous espace vectorile G_{d_n} de \mathbb{L}^2 de dimension finie d_n . Le développement à l'ordre d_n de $f(x)$ dans G_{d_n} est donné par :

$$f_{d_n}(x) = \sum_{k=0}^{d_n} a_k e_k(x), k = 0, \dots, x \in \mathbb{R}. \quad (1.9)$$

Pour estimer $f(x)$ dans \mathbb{L}^2 on se propose de constuire un estimateur sans biais de sa projection orthogonale $f_{d_n}(x)$ dans G_{d_n} . Par la méthode des moments, on peut estimer les coefficients $\{a_k, k \in \mathbb{N}\}$ par :

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^n e_k(X_i), k = 0, \dots \quad (1.10)$$

Ainsi, $f(x)$ peut être estimée par :

$$\hat{f}_{d_n}(x) = \sum_{k=0}^{d_n} \hat{a}_k e_k(x). \quad (1.11)$$

”L’expression (1.11) explique pourquoi $\hat{f}_{d_n}(x)$ est souvent désigné comme l’estimateur de la densité de probabilité par la méthode des séries orthogonales”.

1.3.1 Propriétés statistiques de l’estimateur

a. Les coefficients $(\hat{a}_k)_{k=0, \dots, d_n}$ sont des estimateurs sans biais de $(a_k)_{k=0, \dots, d_n}$.

En effet,

$$\mathbb{E}(\hat{a}_k) = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n e_k(X_i)\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[e_k(X_i)] = \mathbb{E}[e_k(X)] \quad (1.12)$$

$$= a_k. \quad (1.13)$$

b. Le biais de $\hat{f}_{d_n}(x)$ est par définition :

$$\text{Biais}(\hat{f}_{d_n}(x)) = \mathbb{E}(\hat{f}_{d_n}(x)) - h(x) = \sum_{k=0}^{d_n} a_k e_k(x) - f(x), \quad (1.14)$$

ce qui implique que $\hat{f}_{d_n}(x)$ est un estimateur biaisé de $f(x)$.

c. L'erreur quadratique moyenne intégrée de l'estimateur est donnée par le théorème suivant :

Théorème 1.3.1. (*Kronmal-Tarter*)

Si $\int f^2(x)dx < \infty$, alors :

$$MISE(\hat{f}_{d_n}(x)) = \int f^2(x)dx - \sum_{k=0}^{d_n} a_k^2 + \sum_{k=0}^{d_n} \text{Var}(\hat{a}_k). \quad (1.15)$$

1.3.2 Choix pratique de la base

Le choix de la base dépend d'abord du support de la densité à estimer. Si le support de f est un intervalle compact, on pourra choisir les fonctions trigonométriques ou les fonctions de Legendre. Sur \mathbb{R}_+ , on pourra utiliser les fonctions de Laguerre ou les fonctions d'Hermite. Quand on ne possède aucune information sur le support de f on peut utiliser les fonctions d'Hermite. Les fonctions d'Hermite donnent de bons résultats au voisinage de la loi normale réduite puisque le premier élément de la base $e_0(x) = \pi^{-\frac{1}{2}} \exp(-\frac{x^2}{2})$ est la densité d'une variable aléatoire de loi normale centrée réduite. Au voisinage d'une loi normale quelconque on peut considérer des fonctions d'Hermite modifiées données par

$$e_j^1(x) = e_j\left(\frac{x - \bar{X}}{S_n}\right), j \in \mathbb{N}, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n = \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right]^{\frac{1}{2}}. \quad (1.16)$$

"On voit qu'il n'y a pas de solution évidente qui se dégage. En effet, les systèmes orthonormaux sont très variés et il n'existe pas de théorèmes qui permettrait de conseiller un système particulier".

1.3.3 Exemples

Dans cette section, nous allons présenter quelques cas particuliers d'estimateurs basés sur des systèmes orthonormaux trigonométriques (Dirichlet et Fejer) et ceux associés aux fonctions d'Hermite, de Laguerre et de Legendre.

a. L'estimateur associé aux fonctions d'Hermite

Les fonctions d'Hermite sont données par les formules suivantes :

$$e_j(x) = (2^j j! \pi^{\frac{1}{2}})^{-\frac{1}{2}} Q_j(x) \exp\left(-\frac{x^2}{2}\right), x \in \mathbb{R}; j = 0, \dots$$

où $Q_j(x)$ est le $j^{\text{ième}}$ polynôme d'Hermite défini par :

$$Q_j(x) = (-1)^j \exp(-x^2) \frac{d^j}{dx^j} \exp(x^2); x \in \mathbb{R}, \quad j = 0, \dots$$

L'estimateur associé est donné par :

$$\hat{f}_{d_n}(x) = \frac{d_n + 1}{2n} \sum_{i=1}^n \left[\frac{Q_{j+1}(X_i)Q_j(x) - Q_j(X_i)Q_{j+1}(x)}{X_i - x} \right]. \quad x \in \mathbb{R}, \quad (1.17)$$

b. L'estimateur associé aux fonctions de Laguerre

Les fonctions de Laguerre sont données par les formules :

$$L_j(x) = \left[\frac{\Gamma(d_n + 1)}{\Gamma(d_n + 1 + \alpha)} x^{-\alpha} \exp(x) \right]^{\frac{1}{2}} \frac{1}{i!} \frac{d^i}{dx^i} (x^{i+\alpha} \exp(-x)), \quad i \geq 0, \alpha > 0. \quad x \in \mathbb{R}^+$$

L'estimateur de Laguerre associé est alors défini par :

$$\hat{f}_{d_n}(x) = \frac{\Gamma(d_n + 1)}{n\Gamma(d_n + 1 + \alpha)} \sum_{i=1}^n \left[\frac{l_{j+1}(X_i)l_j(x) - l_j(X_i)l_{j+1}(x)}{X_i - x} \right]. \quad x \in \mathbb{R}^+$$

c. L'estimateur associé aux fonctions de Legendre

Les fonctions de Legendre sont définies par :

$$p_i(x) = \sqrt{\frac{2i+1}{2}} \frac{1}{2^i i!} \frac{d^i}{dx^i} ((x^2 - 1)^i), \quad x \in [-1, 1], i \geq 0.$$

L'estimateur associé est alors défini par :

$$\hat{f}_{d_n}(x) = \frac{d_n + 1}{n\sqrt{2d_n + 1}\sqrt{2d_n + 3}} \sum_{i=1}^n \frac{p_{d_n}(X_i)p_{d_n+1}(x) - p_{d_n+1}(X_i)p_{d_n}(x)}{x - X_i}, \quad x \in [-1, 1].$$

d. L'estimateur de Saadi-Adjabi

La base est donnée par :

$$e_k(x) = \frac{1}{\sqrt{2\pi}} (\cos(kx) + \sin(kx)) 1_{[-\pi, \pi]}(x), \quad k = 0, \dots \quad (1.18)$$

L'estimateur de la densité associé à cette base trigonométrique est alors de la forme :

$$\hat{f}_{d_n}(x) = \frac{1}{4\pi n} \sum_{i=1}^n \left[\frac{\sin\left[\frac{(2d_n+1)(X_i-x)}{2}\right]}{\sin\left[\frac{X_i-x}{2}\right]} + \frac{\sin\left[\frac{(2d_n+1)\left(\frac{\pi}{2}-(X_i+x)\right)}{2}\right]}{\sin\left[\frac{\frac{\pi}{2}-(X_i+x)}{2}\right]} \right], \quad x \in [-\pi, \pi]. \quad (1.19)$$

d. L'estimateur associé à la base de Dirichlet

Nous supposons que $I = [-\pi, \pi]$ un intervalle de \mathbb{R} muni de la mesure de Lebesgue et la base orthonormale est définie par :

$$e_0(x) = \frac{1}{\sqrt{2\pi}}, e_{2k}(x) = \frac{\cos(kx)}{\sqrt{\pi}}, e_{2k+1}(x) = \frac{\sin(kx)}{\sqrt{\pi}}; k = 1, \dots$$

pour d_n impair, l'estimateur de Dirichlet est donné par :

$$\hat{f}_{d_n, D}(X_i, x) = \begin{cases} \frac{1}{2\pi n} \sum_{i=1}^n \frac{\sin[d_n \frac{(X_i - x)}{2}]}{\sin[\frac{X_i - x}{2}]} & \text{pour } X_i \neq x, \quad x \in [-\pi, \pi] \\ \frac{d_n}{2\pi} & \text{sinon.} \end{cases}$$

e. L'estimateur de Fejer

Considérons les systèmes suivants dans $[a, b]$:

$$\left\{ \cos k\pi \left[\frac{x-a}{b-a} \right], k = 0, 1, \dots \right\}, \quad \left\{ \sin k\pi \left[\frac{x-a}{b-a} \right], k = 0, 1, \dots \right\}, \quad \text{et} \quad \left\{ \cos k\pi \left[\frac{x-a}{b-a} \right], \sin k\pi \left[\frac{x-a}{b-a} \right], k = 0, 1, \dots \right\} \quad (1.20)$$

Kronmal et Tarter partir d'un développement en série de Fourier. Ils proposent trois estimateurs

$$\hat{f}_{d_n}(x) = \frac{c_0}{2} + \sum_{k=1}^{d_n} \bar{c}_k \cos k\pi \left[\frac{x-a}{b-a} \right], \quad \hat{f}_{d_n}^1(x) = \frac{c_0}{2} + \sum_{k=1}^{d_n} \bar{s}_k \sin k\pi \left[\frac{x-a}{b-a} \right], \quad \text{et} \quad \hat{f}_{d_n}^2(x) = \frac{1}{2} (\hat{f}_{d_n}(x) + \hat{f}_{d_n}^1(x)) \quad (1.21)$$

Dans ces formules \bar{c}_k et \bar{s}_k représentent les moments trigonométriques de l'échantillon, c'est-à-dire :

$$\bar{c}_k = \frac{c_0}{n} \sum_{i=1}^n \cos k\pi \left[\frac{X_i - a}{b-a} \right], \quad \bar{s}_k = \frac{c_0}{n} \sum_{i=1}^n \sin k\pi \left[\frac{X_i - a}{b-a} \right] \quad \text{et} \quad c_0 = \frac{2}{b-a}. \quad (1.22)$$

L'estimateur considéré est alors défini par :

$$g_{d_n}(x) = \frac{c_0}{2n(d_n + 1)} \sum_{i=1}^n \left[\frac{\sin((d_n + 1) \frac{\pi(X_i - a)}{b-a})}{\sin(\frac{\pi(X_i - a)}{b-a})} \right]^2, \quad (1.23)$$

dont l'erreur quadratique moyenne intégrée est donnée par

$$MISE(g_{d_n}(x)) = c_0 \int_a^b g^2(x) dx - \frac{c_0^2}{2} + \sum_{k=1}^{d_n} \left(1 - \frac{k}{d_n + 1}\right)^2 [\text{Var}(\bar{c}_k) + \text{Var}(\bar{s}_k)] - \frac{d_n + 1 + k}{d_n + 1 - k} (\mathbb{E}^2(\hat{c}_k) + \mathbb{E}^2(\hat{s}_k)) \quad (1.24)$$

f. Estimateur associé à la base d'Efremovich

La base cosinus est donnée par :

$$e_k(x) = \begin{cases} 1 & \text{si } k = 0 \\ \sqrt{2} \cos(\pi k x) & \text{si } k = 1, 2, \dots, \end{cases} \quad x \in [0, 1] \quad (1.25)$$

L'estimateur de la densité associé à cette base est donné par :

$$\hat{f}_{d_n}(x) = \frac{1}{2n} \sum_{i=1}^n \left[\frac{\sin\left(\frac{\pi(2d_n+1)(X_i+x)}{2}\right)}{\sin\left(\frac{\pi(X_i+x)}{2}\right)} + \frac{\sin\left(\frac{\pi(2d_n+1)(X_i-x)}{2}\right)}{\sin\left(\frac{\pi(X_i-x)}{2}\right)} \right], \quad X_i \neq x, \quad x \in [0, 1]. \quad (1.26)$$

1.3.4 Choix pratique du paramètre de lissage

La base étant supposée fixée, il reste à choisir le paramètre de lissage d_n . pour cela, on cherche à minimiser l'erreur quadratique moyenne intégrée $MISE(\hat{f}_{d_n}(x))$. Il existe plusieurs méthodes pour le choix du paramètre de lissage : La méthode de Kronmal-Tarter et la méthode de Bosq

1.3.5 Méthode de Kronmal-Tarter

L'emploi de (1.11) pour estimer $f(x)$ n'est possible qu'après avoir déterminé le nombre optimum de terme d_n de la somme. Il est naturel de choisir d_n de sorte que l'erreur quadratique moyenne intégrée $MISE(\hat{f}_{d_n}(x))$ soit minimum. La règle adoptée pour déterminer la valeur optimum d_n repose sur l'algorithme suivant : A partir de $d_n = 1$ on augmente la valeur de d_n d'une unité jusqu'à ce que $MISE(\hat{f}_{d_n}(x))$ augmente on donne alors à d_n la valeur qui précède juste l'augmentation de $MISE(\hat{f}_{d_n}(x))$. On ajoutera donc à la somme (1.11) le $d_n^{\text{ième}}$ terme si et seulement si

$$\Delta_{d_n} = MISE(\hat{f}_{d_n}(x)) - MISE(\hat{f}_{d_n-1}(x)) \leq 0. \quad (1.27)$$

En tenant compte de (1.3.1), Δ_{d_n} se met sous la forme :

$$\begin{aligned}
\Delta_{d_n} &= MISE(\hat{f}_{d_n}(x)) - MISE(\hat{f}_{d_n-1}(x)) \\
&= \int f^2(x)dx + \sum_{k=0}^{d_n} [\text{Var}(\hat{a}_k) - a_k^2] - \int h^2(x)dx + \sum_{k=0}^{d_n-1} [\text{Var}(\hat{a}_k) - a_k^2] \\
&= \text{Var}(\hat{a}_{d_n}) - a_{d_n}^2 \\
&= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n e_{d_n}(X_i)\right) - a_{d_n}^2 \\
&= \frac{1}{n} \text{Var}(e_{d_n}(X)) - a_{d_n}^2 \\
&= \frac{1}{n} \int e_{d_n}(x)h(x)dx - \frac{1}{n}a_{d_n}^2 - a_{d_n}^2 \\
&= \frac{1}{n} \left[\int e_{d_n}(x)h(x)dx - (n+1)a_{d_n}^2 \right] \\
&= \frac{n+1}{n} \text{Var}(e_{d_n}(X)) - \mathbb{E}(e_{d_n}(X))^2.
\end{aligned}$$

Posons alors

$$\theta_i = e_{d_n}(X_i), i = 1, \dots, n, \quad \bar{\theta} = \frac{1}{n} \sum_{i=1}^n \theta_i.$$

On peut alors définir un estimateur symétrique sans biais de Δ_{d_n} donné par :

$$\hat{\Delta}_{d_n} = \frac{1}{n} \left[\frac{n+1}{n-1} \sum_{i=1}^n (\theta_i - \bar{\theta})^2 - \sum_{i=1}^n \theta_i^2 \right]. \quad (1.28)$$

On se fixe maintenant un entier positif D , l'optimum d_n^* est alors de la forme :

$$d_n^* = \begin{cases} \inf\{d_n, 1 \leq d_n \leq D\} & \hat{\Delta}_d > 0 \\ D & \text{sinon.} \end{cases}$$

1.3.6 Méthode de Bosq

Bosq a proposé un nouveau estimateur de paramètre de lissage donné par :

$$\hat{d}_n = \max\{j : 0 \leq j \leq d_n, |\hat{a}_j| \geq \gamma_n\}, \quad (1.29)$$

avec

$$\gamma_n = c \sqrt{\frac{\log n}{n}}, c > 0. \quad (1.30)$$

Théorème 1.3.2.

1. Si $\frac{d_n}{n} \rightarrow 0$ on a

$$MISE(\hat{h}_{d_n}(x)) \rightarrow 0. \quad (1.31)$$

2. Si $\sum_n |a_j| < \infty$ et $\sum_n d_n \exp[-\frac{n}{d_n^2} a] < \infty, a > 0$,
alors

$$\sup_{x \in E} |\hat{f}_{d_n}(x) - f(x)| \xrightarrow{p.s.} 0. \quad (1.32)$$

1.4 Estimateur à noyau

Définition 1.4. Soit (x_1, \dots, x_n) un échantillon de loi $f(x)$ sur \mathbb{R} , de fonction de répartition $F(x) = \int_{-\infty}^x f(t)dt$. On appelle fonction de répartition empirique associée à (x_1, \dots, x_n) , la fonction aléatoire $F_n : \mathbb{R} \rightarrow [0, 1]$ définie pour tout $x \in \mathbb{R}$ par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{]-\infty, x[}(x_i).$$

La densité est la dérivée de la fonction de répartition, ce qui permet d'écrire pour tout x :

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

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

$$f_h(x) = \lim_{h \rightarrow 0} \frac{F_n(x+h) - F_n(x-h)}{2h} = \frac{1}{n} \sum_{i=1}^n 1_{(x-h, x+h]}(x_i) = \frac{1}{n} \sum_{i=1}^n w\left(\frac{x - x_i}{h}\right),$$

avec

$$w(u) = \begin{cases} \frac{1}{2} & \text{si } -1 \leq u \leq 1, \\ 0, & \text{sinon} \end{cases}$$

Cet estimateur appelé estimateur de **Rosenblatt**, est le premier exemple d'estimateur à noyau construit à l'aide du noyau uniforme $K(u) = \frac{1}{2}1_{-1 \leq u \leq 1}$. Parzen a étudié une classe générale d'estimateurs. la méthode de Parzen consiste à utiliser la formule ci-dessus pour tout $x \in \mathbb{R}$ et pas seulement pour la classe $[-1, 1]$. Cette généralisation est certes utile, car elle conduit vers un estimateur qui est constant par morceaux comme les histogrammes, mais a l'avantage d'avoir des plateaux de longueurs variables. On remarque aisément que la discontinuité de l'estimateur défini ci-dessus est une conséquence de la discontinuité de la fonction indicatrice. Par conséquent, en remplaçant $1(z) \leq \frac{1}{2}$ par une fonction K quelconque, on obtient l'estimateur suivant :

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right), \quad (1.33)$$

qui est continu et même l -fois continûment différentiable du moment où la fonction K l'est. h est un paramètre qui est fonction de n , appelé paramètre de lissage. K est une fonction définie sur \mathbb{R} appelée noyau.

Lemme 1.4.1. Si le noyau K est une fonction positive et $\int_{-\infty}^{+\infty} K(\mu)d\mu = 1$, alors $f_h(x)$ est une densité de probabilité.

1.4.1 Noyaux usuels symétriques

Les noyaux les plus couramment utilisés en pratique sont :

- Le noyau uniforme(rectangulaire)

$$K(u) = \frac{1}{2}, \quad |u| \leq 1,$$

- Le noyau Triangulaire

$$K(u) = (1 - |u|), \quad |u| \leq 1,$$

- Le noyau Gaussien

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), \quad u \in \mathbb{R},$$

- Le noyau d'Epanechnikov

$$K(u) = \frac{3}{4}(1 - u^2), \quad |u| \leq 1,$$

- Le noyau Cosine

$$K(u) = \frac{\pi}{4} \cos\left(\frac{\pi u}{2}\right), \quad |u| \leq 1.$$

Les courbes de ces noyaux sont présentées ci-dessous :

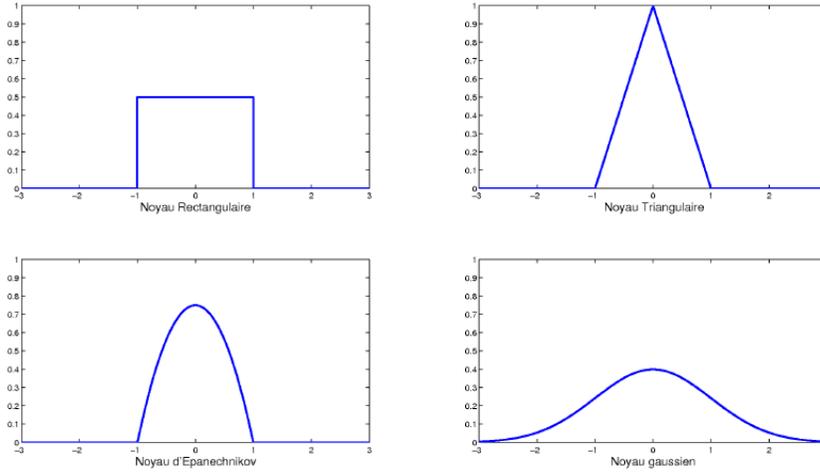


FIGURE 1.2 – Noyaux usuels .

1.4.2 Étude du biais et de la variance

Lorsqu'on définit un estimateur à noyau, on a non seulement le choix de la fenêtre $h > 0$ mais aussi celui du noyau K . Il y a un certain nombre de conditions qui sont considérées comme usuelles pour les noyaux et qui permettent d'analyser le risque de l'estimateur à noyau qui en résulte. On suppose que le noyau K vérifié les 4 conditions suivantes :

1. $\int_{-\infty}^{+\infty} K(\mu)d\mu = 1$, K est une densité de probabilité,
2. $\int_{-\infty}^{+\infty} \mu K(\mu)d\mu = 0$, K est symétrique,
3. $\int_{-\infty}^{+\infty} \mu^2 K(\mu)d\mu = \sigma_K^2 < +\infty$,
4. $\int_{-\infty}^{+\infty} K^2(\mu)d\mu < +\infty$.

Proposition 1.4.2. *Si les trois premières conditions sont remplies, alors*

$$\mathbb{Biais}(f_h(x)) = \mathbb{E}(f_h(x)) - f(x) = \frac{h^2}{2!} f''(x) \mu_2(K) + o(h^2), \quad \mu_2(K) = \int_{-\infty}^{+\infty} y^2 K(y) dy. \quad (1.34)$$

Si, de plus la condition 4 est satisfaite, alors

$$\mathbb{V}(f_h(x)) = \frac{f(x)}{nh} \int K^2(u) du + O\left(\frac{1}{n}\right).$$

Démonstration. L'espérance mathématique de $f_h(x)$ est :

$$\begin{aligned}\mathbb{E}(f_h(x)) &= \frac{1}{nh} \mathbb{E} \left(\sum_{i=1}^n K \left(\frac{x_i - x}{h} \right) \right) \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} K \left(\frac{u - x}{h} \right) f(u) du\end{aligned}$$

En posant $y = \frac{u-x}{h} \Rightarrow dy = \frac{du}{h}$

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

En effectuant un développement de Taylor à l'ordre 2 au point x de la fonction $f(x + hy)$, il vient

$$\begin{aligned}\mathbb{E}(f_h(x)) &= \int_{-\infty}^{+\infty} K(y) \left[f(x) + (yh)f'(x) + \frac{(yh)^2}{2} f''(x) \right] dy + o(h^2) \\ &= f(x) \int_{-\infty}^{+\infty} K(y) dy + hf'(x) \int_{-\infty}^{+\infty} yK(y) dy + \frac{h^2 f''(x)}{2} \int_{-\infty}^{+\infty} y^2 K(y) dy + o(h^2).\end{aligned}$$

Il en résulte que

$$\text{Biais}(f_h(x)) = \mathbb{E}(f_h(x)) - f(x) = \frac{h^2}{2!} f''(x) \mu_2(k) + o(h^2), \quad \mu_2(k) = \int_{-\infty}^{\infty} y^2 K(y) dy, \quad (1.35)$$

Pour prouver la seconde assertion, on utilise le fait que les variables aléatoires sont i.i.d. et que la variance de la somme de variables indépendantes coïncide avec la somme des variances :

$$\begin{aligned}\mathbb{V}(f_h(x)) &= \mathbb{V} \left(\sum_{i=1}^n \frac{1}{nh} K \left(\frac{x_i - x}{h} \right) \right) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n \mathbb{V} \left(K \left(\frac{x_i - x}{h} \right) \right) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n \left[\mathbb{E} \left(K^2 \left(\frac{x_i - x}{h} \right) \right) \right] - \frac{1}{n^2 h^2} \sum_{i=1}^n \left[\mathbb{E} \left(K \left(\frac{x_i - x}{h} \right) \right) \right]^2 \\ &= \frac{f(x)}{nh} \int_{-\infty}^{\infty} K^2(y) dy - \frac{f'(x)}{n} \int_{-\infty}^{\infty} y K^2(y) dy - \frac{1}{n} (f(x) + \text{Biais}(f_n(x)))^2,\end{aligned}$$

ce qui nous donne :

$$\text{Var}(f_h(x)) = \frac{f(x)}{nh} \int_{-\infty}^{\infty} K^2(y) dy + O \left(\frac{1}{nh} \right). \quad (1.36)$$

Proposition 1.4.3. *Si les trois premières conditions sont remplies et f est une densité bornée dont la dérivée seconde est bornée, alors*

$$|\text{Biais}(f_h(x))| \leq C_1 h^2, \quad (1.37)$$

où $C_1 = \frac{1}{2} \sup_{x \in \mathbb{R}} |f''(x)| \int \mu^2 |K(\mu)| d\mu$.

Si, de plus la condition 4 est satisfaite, alors

$$\mathbb{V}(f_h(x)) \leq \frac{C_2}{nh},$$

avec, $C_2 = \sup_{x \in \mathbb{R}} |f(x)| \int K^2(\mu) d\mu$.

Démonstration. Supposons f de classe C^2 et telle que f'' soit bornée.

$$\begin{aligned} \text{Biais}(f_h(x)) &= \mathbb{E}(f_h(x)) - f(x) = \frac{1}{nh} \mathbb{E} \left(\sum_{i=1}^n K \left(\frac{x_i - x}{h} \right) \right) - f(x) \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} K \left(\frac{u - x}{h} \right) f(u) du - f(x) \end{aligned}$$

En posant $y = \frac{u-x}{h} \Rightarrow dy = \frac{du}{h}$

$$\text{Biais}(f_h(x)) = \int_{-\infty}^{+\infty} K(y) [f(x + hy) - f(x)] dy.$$

Puisque f est supposée de classe C^2 , on peut appliquer la formule de Tylor à l'ordre 2, ce qui nous donne :

$$\text{Biais}(f_h(x)) = \frac{h^2}{2!} f''(x) \mu_2(k) + o(h^2). \quad (1.38)$$

Ainsi, en ayant supposé de plus que le noyau k est symétrique et f'' est bornée,

$$|\text{Biais}(f_h(x))| \leq \frac{h^2}{2} \sup_{x \in \mathbb{R}} |f''(x)| \int_{-\infty}^{\infty} y^2 |K(y)| dy. \quad (1.39)$$

Pour prouver la seconde assertion, on utilise le fait que les variables aléatoires sont i.i.d. et que la variance de la somme de variables indépendantes coïncide avec la somme des variances :

$$\begin{aligned} \mathbb{V}(f_h(x)) &= \mathbb{V} \left(\sum_{i=1}^n \frac{1}{nh} K \left(\frac{x_i - x}{h} \right) \right) \\ &= \frac{1}{n^2 h^2} \sum_{i=1}^n \mathbb{V} \left(K \left(\frac{x_i - x}{h} \right) \right) \\ &\leq \frac{1}{nh^2} \left[\mathbb{E} \left(K^2 \left(\frac{X - x}{h} \right) \right) \right] = \frac{1}{nh^2} \int_{-\infty}^{\infty} K^2 \left(\frac{u - x}{h} \right) f(u) du. \end{aligned}$$

On en déduit la majoration :

$$\mathbb{V}(f_h(x)) \leq \frac{\|f\|_\infty}{nh} \int_{-\infty}^{\infty} K^2(y) dy.$$

Discussion du comportement du biais et de la variance :

- * Le biais décroît si h diminue mais la variance augmente.
- * Pour que la variance tende vers 0, il faut que $nh \rightarrow \infty$.
- * La variance diminue si h augmente mais le biais augmente.

1.4.3 Risque quadratique ponctuel et risque quadratique intégré

Risque quadratique ponctuel :

$$\begin{aligned} MSE(f(x), f_h(x)) &= \mathbb{E}(f(x) - f_h(x))^2 \\ &= [\mathbb{E}(f_h(x) - f(x))]^2 + \mathbb{E}(f_h^2(x)) - [\mathbb{E}(f_h(x))]^2. \end{aligned}$$

$$MSE(f(x), f_h(x)) = [\text{Biais}(f_h(x))]^2 + \mathbb{V}(f_h(x)). \quad (1.40)$$

En remplaçant les expressions finales des deux termes, le biais et la variance dans l'équation (1.40) on obtient :

$$MSE(f(x), f_h(x)) = \frac{f(x)}{nh} \int_{-\infty}^{\infty} K^2(y) dy + \frac{1}{4} h^4 (f''(x))^2 \left(\int_{-\infty}^{\infty} y^2 K(y) dy \right)^2 + O\left(\frac{1}{nh} + h^5\right). \quad (1.41)$$

Risque quadratique intégré :

$$\begin{aligned} MISE(f, f_h) &= \int_{-\infty}^{\infty} MSE(f(x), f_h(x)) dx \\ &= \int \mathbb{E}(f(x) - f_h(x))^2 dx \end{aligned}$$

$$MISE(f, f_h) = \int [(\text{Biais}(f_h(x)))^2 + \mathbb{V}(f_h(x))] dx. \quad (1.42)$$

En remplaçant les expressions finales des deux termes, le biais et la variance dans l'équation (1.42) on obtient :

$$MISE(f(x), f_h(x)) = \frac{h^4}{4} \sigma_k^4 \int (f''(x))^2 dx + \frac{1}{nh} \int K^2(u) du + O\left(h^5 + \frac{1}{n}\right). \quad (1.43)$$

L'erreur quadratique moyenne intégrée asymptotique AMISE :

$$AMISE = MISE(f(x), f_h(x)) - O\left(h^5 + \frac{1}{n}\right),$$

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

avec

$$R(s) = \int s^2(x) dx.$$

Théorème 1.4.4. *Si les 4 conditions sont remplies, alors Le paramètre de lissage h^* qui minimise l'erreur quadratique moyenne intégrée asymptotique est de la forme :*

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

La valeur du AMISE optimale $AMISE^* = AMISE(h^*)$ est alors de forme

$$AMISE^* = \frac{5}{4} [\sigma_K R^4(K) R(f'')]^{1/5} n^{-4/5}.$$

Démonstration. le paramètre de lissage h qui minimise l'erreur quadratique moyenne intégrée asymptotique est :

$$\begin{aligned} \frac{dAMIS}{dh} &= h^3 \sigma_K^2 R(f'') - \frac{R(K)}{nh^2} = 0 \\ nh^5 \sigma_K^4 - R(K) &= 0 \Rightarrow h^5 = \frac{R(K)}{n \sigma_K^4 R(f'')}. \end{aligned}$$

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

$$\begin{aligned} \frac{d^2 AMIS}{dh^2} &= 3h^2 \sigma_K^4 R(f'') + \frac{R(K)}{nh^3} > 0 \Rightarrow h^* \text{ minimise AMISE} \\ nh^5 \sigma_K^4 - R(K) &= 0 \Rightarrow h^5 = \frac{R(K)}{n \sigma_K^4 R(f'')}. \end{aligned}$$

La valeur du AMISE optimale $AMISE^* = AMISE(h^*)$ est donnée par :

$$AMISE^* = \frac{5}{4} [\sigma_K^4 R^4(K) R(f'')]^{1/5} n^{-4/5}. \quad (1.44)$$

Théorème 1.4.5. (Majoration du risque quadratique ponctuel). *Si les 4 conditions sont remplies et f est une densité bornée dont la dérivée seconde est bornée, alors le risque quadratique admet la majoration suivante :*

$$MSE(f(x), f_h(x)) \leq \frac{C^2 h^4}{4} + \frac{C_2}{nh}.$$

Discussion : Sur-lissage et sous-lissage

Lorsque la fenêtre h est très petit, le biais de l'estimateur à noyau est très petit face à sa variance et c'est cette dernière qui détermine la vitesse de convergence du risque quadratique. Dans ce type de situation, l'estimateur est très volatile et on parle de **sous-lissage** (under-smoothing, en anglais). En revanche, lorsque h grandit, la variance devient petite et c'est le biais qui devient dominant. L'estimateur est alors très peu variable et est de moins à moins influencé par les données. On parle alors d'un effet de **sur-lissage** (over-smoothing en anglais). En pratique, il est primordial de trouver la bonne dose de lissage qui permet d'éviter le sous-lissage et le **sur-lissage**.

1.4.4 Noyau d'ordre l

Définition 1.5. On dit que $K : \mathbb{R} \rightarrow \mathbb{R}$ est un noyau d'ordre l si les fonctions

$$\begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ u \rightarrow u^j K(u) \end{cases}$$

sont intégrables pour $j = 0, 1, \dots, l$ et vérifient $\int_{-\infty}^{+\infty} K(u) du = 1$, ainsi que

$$\int_{-\infty}^{+\infty} u^j K(u) du = 0, j = 1, \dots, l.$$

Définition 1.6. Soient T un intervalle de \mathbb{R} , et deux réels $\beta, L > 0$. La classe de Hölder $\Sigma(\beta, L)$ sur T est définie comme l'ensemble des fonctions $g : T \rightarrow \mathbb{R}$ telles que g est $l = \lfloor \beta \rfloor$ fois dérivable et telle que $|g^l(x) - g^l(y)| \leq |x - y|^{\beta-l}, \forall x, y \in T$.

Théorème 1.4.6. Soit f une densité et K un noyau tels que : f est bornée et dans une classe de Hölder $\Sigma(\beta, L)$ sur \mathbb{R} , K est un noyau d'ordre $l = \lfloor \beta \rfloor$ de carré intégrable et tel que $\int_{-\infty}^{+\infty} |\mu|^\beta |K(\mu)| d\mu < \infty$, alors

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

Pour $h_{opt} = bn^{-\frac{1}{2\beta+1}}$, avec $b > 0$, il existe $C > 0$ tel que

$$MSE(\hat{f}_h(x)) \leq Cn^{-\frac{2\beta}{2\beta+1}}.$$

Définition 1.7. Soient deux réels $\beta, L > 0$ et soit $p = \lfloor \beta \rfloor$. On appelle espace de **Niolski** l'ensemble, noté $\mathcal{H}(\beta, L)$, des fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$ tel que f est p fois dérivable et tel que

$$\left(\int_{-\infty}^{\infty} (f^{(p)}(x+t) - f^{(p)}(x))^2 dx \right)^{\frac{1}{2}} \leq L |t|^{\beta-p}, \forall t \in \mathbb{R}.$$

Théorème 1.4.7. Soit $f \in \mathcal{H}(\beta, L)$ et K est un noyau d'ordre $l = \lfloor \beta \rfloor$, tel que $\int_{-\infty}^{+\infty} |\mu|^\beta |K(\mu)| d\mu < \infty$, alors

$$MISE \left(\hat{f}_h(x) \right) \leq \frac{\int K^2}{nh} + c_2 h^{2\beta},$$

avec

$$c_2 = \frac{1}{p!} \int_{-\infty}^{+\infty} |\mu|^\beta |K(\mu)| d\mu.$$

Pour $h_{opt} = bn^{-\frac{1}{2\beta+1}}$, avec $b > 0$, il existe $C > 0$ tel que

$$MISE \left(\hat{f}_h(x) \right) \leq Cn^{-\frac{2\beta}{2\beta+1}}.$$

1.5 Les méthodes de sélection du paramètre de lissage

Le choix judicieux du paramètre de lissage permet une bonne utilisation de cette méthode dans la pratique. Dans ce chapitre, nous présenterons les méthodes suivantes :

- la méthode de "plug-in" (Rul of Thumb) ;
- la méthode de "plug-in" itéré ;
- la méthode de la validation croisée.

Définition 1.8. (Choix théorique optimal de h).

Le paramètre de lissage h est un réel positif dont le choix est dominant par rapport au choix du noyau K . Le paramètre de lissage h_{opt} qui minimise l'erreur quadratique moyenne intégrée asymptotique AMISE est de la forme :

$$h_{opt} = \left[\frac{R(K)}{\sigma_K^4 R(f'')} \right]^{1/5} n^{-1/5} = \psi(K) \varphi(f'') n^{-1/5} \quad (1.45)$$

avec

$$\psi(K) = \left[\frac{R(K)}{\sigma_K^4} \right]^{1/5} \quad \text{et} \quad \varphi(f'') = \left[\frac{1}{R(f'')} \right]^{1/5}, \varphi(f'') \neq 0.$$

Définition 1.9. La valeur du AMISE optimale $AMISE_{opt} = AMISE(h_{opt})$ est donnée par :

$$AMISE_{opt} = \frac{5}{4} \left[\sigma_K^4 R^4(K) R(f'') \right]^{1/5} n^{-4/5}. \quad (1.46)$$

Remarque 1.5.1. La convergence de l'estimateur à noyau est plus rapide que pour l'histogramme, étant d'ordre $n^{-4/5}$ au lieu de $n^{-2/3}$, mais ce résultat dépend de la quantité inconnue f'' , c'est pourquoi on considère des méthodes pratiques pour le choix du paramètre de lissage.

1.6 Méthodes plug-in

1.6.1 Rule of Thumb

L'idée de base de la méthode Rule of Thumb pour le choix du paramètre h , est d'estimer dans l'expression de h_{opt} (1.45), la quantité inconnue $R(f'')$. Supposant que f appartient à une famille de distributions normales $\mathcal{N}(\mu; \sigma^2)$, de moyenne μ et variance σ^2 inconnues. Sous cette hypothèse :

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

la densité de probabilité normale centrée réduite et

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

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 $dv = \frac{1}{\sigma} dx$

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

mais

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

ce qui implique

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

1.6.2 Plug-in itéré

Définition 1.10. En adoptant le critère de l'erreur quadratique intégrée moyenne (*MISE*), Scott, Tapia et Thompson choisissent d'estimer le paramètre $R(f'')$ à l'aide de l'estimateur naturel $\hat{R}_h(f'')$ défini comme suit :

$$\hat{R}_h(f'') = R(f_h''), \quad (1.47)$$

où, f_h'' désigne la dérivée seconde de l'estimateur à noyau f_h . Avec un noyau K deux fois dérivable, on voit que :

$$f_h''(x) = \frac{1}{nh^3} \sum_{i=1}^n K'' \left(\frac{x - x_i}{h} \right). \quad (1.48)$$

En choisissant par exemple le classique noyau gaussien

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{u^2}{2} \right), u \in \mathbb{R}.$$

L'estimateur $\hat{R}_h(f'')$ s'écrit comme suit :

$$\hat{R}_h(f'') = \frac{3}{8\sqrt{\pi}n^2h^9} \sum_{i=1}^n \sum_{j=1}^n [h^4 - (x_i - x_j)h^2 + \frac{1}{12}(x_i - x_j)^4] \exp \left[-\frac{(x_i - x_j)^2}{4h^2} \right] \quad (1.49)$$

Remarque 1.6.1. Il est important de noter que la largeur de fenêtre h contrôlant l'estimateur $\hat{R}_h(f'')$ de $R(f'')$ a été choisie identique à la largeur de fenêtre intervenant dans l'estimateur f_h de f . En supposant que la quantité $\hat{R}_h(f'')$ devrait être robuste par rapport à une erreur de spécification sur f , Scott, Tapia et Thompson proposent finalement d'injecter l'estimateur $\hat{R}_h(f'')$ dans l'expression (1.45). Cette approche amène à considérer l'équation numérique suivante en h :

$$h = \psi(K)\varphi(f_h)n^{-\frac{1}{5}}, \quad \text{où } \varphi(f_h) = \left[\frac{1}{R(f_h'')} \right]^{\frac{1}{5}}. \quad (1.50)$$

Algorithme (S.T.T)

La méthode de sélection suggérée par Scott, Tapia et Thompson revient à examiner les éventuels points fixes du système dynamique discret défini sur \mathbb{R}^+ de la façon suivante :

$$h_{i+1} = \phi(h_i), \quad (1.51)$$

où,

$$\phi(f_{h_i}) = \psi(K)\varphi(f_{h_i})n^{-\frac{1}{5}} = \psi(K) \left[\frac{1}{R(f_{h_i}'')} \right]^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

Les étapes de ce processus itératif, appelé algorithme (S.T.T) sont :

- h_0 : solution initiale, prenant par exemple l'étendue de l'échantillon.
- $h_{i+1} = \psi(K)\varphi(f_h)n^{-\frac{1}{5}}$.
- Le critère d'arrêt est donné par la formule suivante :

$$\left| \frac{h_{i+1} - h_i}{h_i} \right| \leq \epsilon, \quad (1.52)$$

où ϵ est une précision petite donnée.

1.6.3 Méthodes cross validation (validation croisée)

L'idée de base des méthodes validation croisée consiste à trouver une fonction de score $CV(h)$ ayant la même structure que le $MISE(h)$ et dont le calcul soit plus simple.

Validation croisée non biaisée

Cette méthode a été proposée par Rudemo et Bowman en 1984. Le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de :

$$ISE(h) = \int_{\mathbb{R}} [f_h(x) - f(x)]^2 dx = \int_{\mathbb{R}} f_h^2(x) dx - 2 \int_{\mathbb{R}} f_h(x) f(x) dx + \int_{\mathbb{R}} f^2(x) dx.$$

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

$$UCV(h) = ISE(h) - \int_{\mathbb{R}} f^2(x) dx = \int_{\mathbb{R}} f_h^2(x) dx - 2 \int_{\mathbb{R}} f_h(x) f(x) dx.$$

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

$$\int_{\mathbb{R}} f_h(x) f(x) dx = \mathbb{E}[f_h(x)].$$

L'estimateur empirique de $\int_{\mathbb{R}} f_h(x) f(x) dx$, est alors $\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i)$. Le critère à optimiser est alors :

$$UCV(h) = \int_{\mathbb{R}} f_h^2(x) dx - \frac{2}{n} \sum_{i=1}^n f_{h,i}(x_i). \quad (1.53)$$

Où $f_{h,i}(x) = \frac{1}{(n-1)h} \sum_{1 \leq i \leq n, i \neq j} k\left(\frac{x-x_j}{h}\right)$ est l'estimateur de la densité construit à partir de l'ensemble de points sauf le point x_i .

En utilisant l'équation (1.53), le critère $UCV(h)$ devient :

$$UCV(h) = \frac{R(K)}{nh} + \sum_{i=1}^n \sum_{i \neq j, j=1}^n \left[\int \frac{1}{n^2 h^2} K\left(\frac{x-x_i}{h}\right) K\left(\frac{x-x_j}{h}\right) dx - \frac{2}{n(n-1)h} K\left(\frac{x_i-x_j}{h}\right) \right]. \quad (1.54)$$

avec

$$R(K) = \int K^2(u) du.$$

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

Validation croisée biaisée

La méthode de validation croisée biaisée, a été introduit par Scott et Terrell pour remédier aux problèmes de validation croisée non biaisée. Il s'agit d'introduire un biais dans le UCV afin de réduire sa variance.

Lemme 1.6.1. (Scott et Terrell)

Supposant que le noyau k satisfait aux conditions suivantes :

$$\int K''(u)du = 0, \quad u_1(K'') = \int uK''(u)du = 0, \quad u_2(K'') = \int u^2K''(u)du = 2.$$

On obtient le développement asymptotique :

$$\mathbb{E}[R(f_h'')] = R(f_h'') + \frac{R(K'')}{nh^5} + o(h^2).$$

avec :

$$R(f_h''(x)) = \int_{\mathbb{R}} (f_h''(x))^2 dx, \quad R(K'') = \int_{\mathbb{R}} (K''(u))^2 du.$$

L'estimateur du $AMISE$:

$$BCV(h) = \frac{h^4}{4} \sigma_k^4 \left[R(f_h'') - \frac{R(K'')}{nh^5} \right] + \frac{R(K)}{nh}. \quad (1.55)$$

Validation croisée de la vraisemblance

Définition 1.11. Pour un estimateur à noyau f_h , la sélection par validation croisée de la vraisemblance consiste à maximiser par rapport à h la vraisemblance pour l'échantillon $(x_1)_{\leq i \leq n}$ définie par :

$$LCV(h) = \prod_{i=1}^n f_{h,i}(x_i),$$

où ;

$$f_{h,i}(x_i) = \frac{1}{n(n-1)} \sum_{1 \leq j \leq n, i \neq j} K\left(\frac{x-x_i}{h}\right),$$

est l'estimateur à noyau basé sur les $(n-1)$ observations différentes de x_i . La vraisemblance est alors :

$$LCV(h) = \prod_{i=1}^n \frac{1}{n(n-1)} \sum_{1 \leq j \leq n, i \neq j} K\left(\frac{x-x_i}{h}\right).$$

Avantages et Inconvénients de diverses méthodes de sélection du paramètre de lissage

- Pour les méthodes plug-in, plusieurs points importants peuvent être impérativement soulignés. D'abord, cette technique est très satisfaisante théoriquement puisque l'expression de minimisant le MISE(h) qui est d'ordre $O(n^{-1/5})$ est de la forme :

$$h_{opt} = \left[\frac{R(K)}{\sigma_K^4 R(f'')} \right]^{1/5} n^{-1/5} = \psi(K) \varphi(f'') n^{-1/5}.$$

- Les difficultés importantes se posent en pratique : pour calculer h_{opt} il faut, en effet, estimer $\varphi(f'')$ et donc les dérivées de la fonction de distribution f qui s'avère techniquement délicat.
- Le principal inconvénient des méthodes validation croisée est que la largeur de fenêtre estimée par cette technique présente une grande variabilité, c'est-à-dire que pour deux échantillons distincts issus de la même distribution, les fenêtres obtenues seront très différentes.
- Cette méthode présente cependant de nombreux avantages : outre le fait qu'elle ne demande pour être applicable, que des hypothèses faibles sur le degré de différentiabilité de f .
- La méthode de Validation croisée de la vraisemblance révèle un certain nombre de faiblesses pour les estimateurs non paramétriques tels que les estimateurs à noyau. Plusieurs études ont mis en avant la mauvaise robustesse de cette méthode ainsi que le risque qu'elle conduise à une estimation non consistante quand elle est appliquée à des observations dont la distribution présente des queues.

1.7 Conclusion

Nous avons exposé dans ce chapitre les différentes méthodes non paramétriques d'estimation de la densité de probabilité. Nous avons choisi d'étudier plus particulièrement la méthode du noyau et la méthode des fonctions orthogonales, car ces méthodes sont les plus développées à ce jour et les propriétés statistiques des deux méthodes sont établies. Nous avons donné les différents résultats théoriques concernant la convergence et la consistance de deux méthodes. L'estimateur obtenu par la méthode du noyau dépend de deux paramètres : la fonction K appelée noyau et h la largeur de fenêtre ou paramètre de lissage. Le choix judicieux de ces deux paramètres permet une bonne utilisation de cette méthode dans la pratique. Dans l'estimation de la densité par les fonctions orthogonales, il n'existe pas de système (base) optimal, le choix de la base dépend du support de la densité à estimer. Par contre le choix du paramètre de lissage est important. Il peut être fait sans aucune information sur la densité inconnue. Il existe deux méthodes pour le choix du paramètre de lissage : la méthode de Kronmal-Tarter basée sur la minimisation l'erreur quadratique intégrée et la méthode de Bosq qui détermine le paramètre de lissage de telle façon que l'erreur quadratique moyenne intégrée tend vers 0 quand n tend vers l'infini.

Chapitre 2

Estimation non paramétrique du mode

2.1 Introduction

Le mode est la valeur la plus probable : pour une densité de probabilité f , c'est la valeur pour laquelle f admet un maximum. Pour une distribution symétrique, il coïncide avec deux autres paramètres de position, la moyenne et la médiane. Dans ce chapitre nous allons exposer les différents estimateurs du mode.

2.2 Le mode d'une variable aléatoire de loi de probabilité

Définition 2.1. Soit X_1, \dots, X_n un n -échantillon de variables aléatoires indépendantes et identiquement distribuées de densité de probabilité f . le mode de f est la valeur x_0 vérifiant :

$$f(x_0) = \max_{x \in \mathbb{R}} f(x) \quad (2.1)$$

x_0 est la solution de l'équation :

$$\begin{cases} f'(x) = 0 \\ f''(x) < 0 \end{cases} \quad (2.2)$$

Il s'agit d'une caractéristique de f qui joue un rôle important tant en probabilité qu'en statistique.

Exemple 2.2.1. (loi Normale). Soient $\mu > 0, \sigma > 0$. On dit qu'une variable aléatoire X suit une loi normale si sa fonction de densité f est donnée par :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathbf{1}_{\mathbb{R}}(x).$$

pour $\mu = 0$ et $\sigma = 1$, On a

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathbf{1}_{\mathbb{R}}(x).$$

D'après (2.1) et (2.2), On obtient :

$$f'(x) = \frac{-x}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathbf{1}_{\mathbb{R}}(x).$$

$$f'(x) = 0 \Rightarrow x = 0.$$

$$f''(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} [-1 + x^2]$$

$$f''(0) = -\frac{1}{\sqrt{2\pi}} < 0.$$

Alors le mode de $f(x)$ est défini par : $x_0 = \mu = 0$.

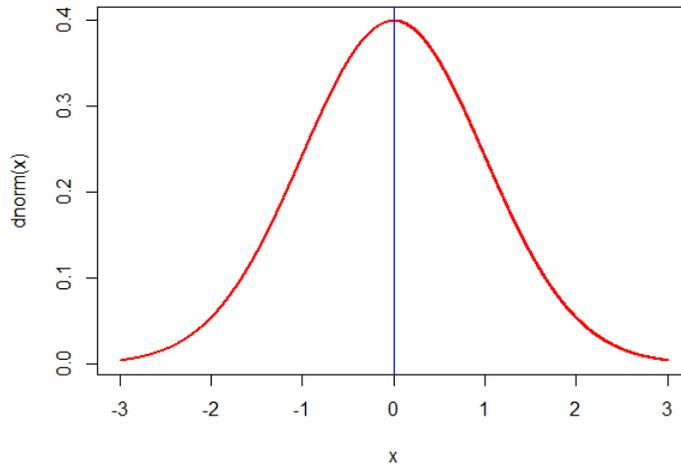


FIGURE 2.1 – Détermination graphique du mode de loi Normal $N(0, 1)$

Exemple 2.2.2. (loi Gamma). Soient $\alpha > 0, \beta > 0$. On dit qu'une variable aléatoire X suit une loi Gamma si sa fonction de densité f est donnée par :

$$f(x) = \frac{\beta}{\Gamma(\alpha)} (\beta x)^{\alpha-1} e^{-\beta x} \mathbf{1}_{\mathbb{R}_+^*}(x).$$

pour $\alpha = 2, \beta = 1$, On a

$$f(x) = \frac{1}{\Gamma(2)} x e^{-x} \mathbf{1}_{\mathbb{R}_+^*}(x).$$

D'après 2.1 et 2.2, On obtient :

$$f'(x) = \frac{1}{\Gamma(2)} (1-x)e^{-x}.$$

$$f'(x) = 0 \Rightarrow x = 1.$$

$$f''(x) = \frac{1}{\Gamma(2)} [x-2]e^{-x}.$$

$$f''(1) < 0.$$

Alors le mode est donné par : $x_0 = \frac{\alpha-1}{\beta} = 1$.

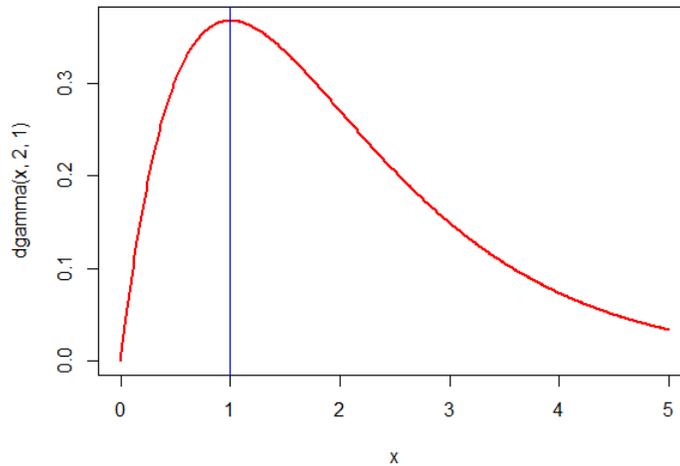


FIGURE 2.2 – Détermination graphique du mode de loi Gamma $G(2, 1)$.

2.2.1 Mode d'une variable aléatoire de loi continue usuelles

Supposons que $[a, b] \subset \mathbb{R}$, $m \in \mathbb{R}$, $\sigma \in \mathbb{R}_+^*$, $\lambda \in \mathbb{R}_+^*$, $\alpha \in \mathbb{R}_+^*$, $\beta \in \mathbb{R}_+^*$, $b \in \mathbb{R}_+^*$. Dans le tableau suivant, nous présentons le mode de certaines variables aléatoires de loi continue :

La loi	$f_X(x)$	le mode x_0
Uniforme $U[a, b]$	$\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$	toute valeur dans $[a, b]$
Normale $N(m, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \mathbf{1}_{\mathbb{R}}(x)$	m
Exponentielle $\zeta(\lambda)$	$\lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$	0
Gamma $\Gamma(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} (\beta x)^{\alpha-1} e^{-\beta x} \mathbf{1}_{\mathbb{R}_+^*}(x)$	$\frac{\alpha-1}{\beta}, \beta > 0$
Cauchy	$\frac{1}{\pi(1+x^2)} \mathbf{1}_{\mathbb{R}}(x)$	0
Beta (α, b)	$\frac{x^{\alpha-1}(1-x)^{b-1}}{\beta(\alpha, b)} \mathbf{1}_{[0,1]}(x)$	$\frac{\alpha-1}{\alpha+b-2}$, pour $\alpha > 1$ et $b > 1$

TABLE 2.1 – Mode d’une variable aléatoire de loi Continue.

2.2.2 Mode d’une variable aléatoire de loi discrète usuelles

Dans le tableau ci-dessous, nous présentons le mode de certaines variables aléatoires de loi discrètes.

On suppose que $n \in \mathbb{N}^*$, $p, q \in]0, 1[$, $\lambda \in \mathbb{R}_+^*$:

La loi	Probabilités $p(X = k)$	Le mode x_0
Bernoulli $\beta(p)$	$\begin{cases} p(X = 0) = 1 - p = q \\ p(X = 1) = p \end{cases}$	$\begin{cases} 0 & \text{si } p < q \\ 0.1 & \text{si } p = q \\ 1 & \text{si } p > q. \end{cases}$
Binomiale $\beta(n, p)$	$p(X = k) = C_k^n p^k (1-p)^{n-k} \mathbf{1}_{\{0, \dots, n\}}(k)$	$\lfloor (n+1)p \rfloor$
Poisson $\rho(\lambda)$	$p(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \mathbf{1}_{\mathbb{N}}(k)$	$\begin{cases} \lfloor \lambda \rfloor & \text{si } \lambda \in \mathbb{R} \\ \lambda & \text{et } \lambda - 1 & \text{si } \lambda \in \mathbb{N} \end{cases}$
Géométrique $g(p)$	$p(X = k) = p(1-p)^{k-1} \mathbf{1}_{\mathbb{N}^*}(k)$	1

TABLE 2.2 – Mode d’une variable aléatoire de loi discrète.

2.3 Estimation non paramétrique du mode

L’estimation non paramétrique du mode x_0 de $f(x)$ est la seule façon de procéder lorsque on ne dispose d’aucune information à priori sur la densité elle même. On peut distinguer deux méthodes d’estimation non paramétrique :

- La première méthode (méthode directe) procède d’emblée à l’estimation de x_0 en se basant sur le fait que, dans l’échantillon, on doit observer un groupement des valeurs dans le voisinage du mode.
- La seconde méthode (méthode indirecte) consiste à obtenir dans un premier temps une estimation $\hat{f}(x)$ de la densité et à prendre pour mode de valeur de x pour laquelle $\hat{f}(x)$ est maximum.

2.4 Méthode directe

Nous présentons dans cette section l'estimation directe du mode x_0 . En particulier, Nous nous intéressons aux estimateurs suivants :

- a. Estimateur de Grenander.
- b. Estimateur de Chernouff.

Estimateur de Grenander

Supposons que

- 1) $f(x)$ est positive et a une dérivée continue .
- 2) $f(x)$ a un unique maximum, le mode x_0 , de sorte que $f(x) < f(x_0)$, avec égalité seulement pour $x = x_0$.
- 3) $f(x)$ monotone pour les grandes valeurs de $|x|$. Nous ne supposons aucune représentation paramétrique de $f(x)$, et nous ne limitons pas non plus le choix de $f(x)$ à l'une des propriétés de symétrie

Définition 2.1. (Grenander (1965)). Si les conditions 1),2) et 3) sont remplies, alors l'estimateur de Grenander est donné par :

$$M_{p,k} = \left[\frac{1}{2} \sum_{i=1}^{n-k} \frac{x_{i+k} + x_i}{(x_{i+k} + x_i)^p} \right] \left[\sum_{i=1}^{n-k} \frac{1}{(x_{i+k} + x_i)^p} \right]^{-1}, 1 < p < k, \quad (2.3)$$

pour un échantillon d'observations x_1, \dots, x_n ordonnées par ordre croissant, $x_1 \leq x_2 \leq \dots \leq x_n$.

Estimateur de Chernouff (1964)

Chernouff (1964) présente l'estimateur du mode noté \hat{x}_{a_n} comme le milieu d'un intervalle de longueur $2a_n$ contenant le maximum d'observations x_1, x_2, \dots, x_n , $(a_n)_n$ étant une suite de réels positifs décroissant lentement vers zéro.

2.4.1 Méthode indirecte

Cette méthode consiste à obtenir dans un premier temps une estimation $\hat{f}(x)$ de la densité et à prendre pour mode de valeur de x pour laquelle $\hat{f}(x)$ est maximum. Dans cette section, nous présentons deux méthodes d'estimation, à savoir l'estimation à noyau introduite par Parzen (1964) et la méthode d'estimation par des séries orthogonales.

Estimation du mode par la méthode du noyau

Parzen (1962) a été l'un des premiers à s'intéresser au problème de l'estimation du mode x_0 dans le cas d'une densité univariée. Il définit un estimateur comme la variable aléatoire qui maximise l'estimateur à noyau f_h de la densité :

$$\hat{x}_0 = \arg \max_{x \in \mathbb{R}} f_h(x).$$

Il démontre que cet estimateur est uniformément convergent (en probabilité), asymptotiquement normal et donne une évaluation de l'erreur quadratique moyenne (MSE). Nadaraya (1965) en établit la consistance forte. D'autres études ont affaibli les conditions suffisantes de normalité asymptotique (Eddy (1980),(1982), Romano (1988)), ou établi la convergence dans L_p (Devroye et Wagner (1980), Grund et Hall (1995)). La liste des travaux est longue, on citera ceux de Van Ryzin(1969), Mokkadem et Pelletier (2003), Herrmann et Ziegler (2004) parmi d'autres. Plus récemment, Shi et al (2009) améliorent le taux de convergence de l'estimateur du mode, exprimé en fonction de la fenêtre h .

Théorème 2.4.1. (Parzen (1962)). *Si h est fonction de n satisfaisant $\lim_{n \rightarrow \infty} nh^2 = +\infty$, et si la densité de probabilité $f(x)$ est uniformément continue, alors pour chaque $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sup_{x \in \mathbb{R}} |f_h(x) - f(x)| \leq \epsilon \right] = 1. \quad (2.4)$$

et pour $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} [|\hat{x}_0 - x_0| < \epsilon] = 1. \quad (2.5)$$

Normalité asymptotique

Théorème 2.4.2. (Parzen (1962)).

Soient x_1, x_2, \dots, x_n un échantillon de variables aléatoires indépendantes et identiquement distribuées de densité f (inconnue), φ est la fonction caractéristique de la variable aléatoire X et $f_h = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)$ l'estimateur à noyau de f . Supposons qu'il existe δ , $0 \leq \delta \leq 1$ et que h est une fonction de n , telle que

$$\lim_{n \rightarrow \infty} nh^6 = \infty, \quad \lim_{n \rightarrow \infty} nh^{5+2\delta} = 0.$$

et

$$\int_{-\infty}^{+\infty} u^{2+\delta} |\varphi(u)| du \leq \infty.$$

Alors,

$$\mathbb{E}[\sup_{x \in \mathbb{R}} |f_h''(x) - f''(x)|] \rightarrow 0,$$

et $\sqrt{nh^3}(\hat{x}_0 - x_0) \rightarrow N(0, f(x_0)/[f''(x_0)]^2 J)$, lorsque $n \rightarrow \infty$.

Avec

$$J = \int_{-\infty}^{+\infty} K'^2(y) dy = (2\pi)^{-1} \int_{-\infty}^{+\infty} u^2 K^2(u) du.$$

Vitesse de convergence

Théorème 2.4.3. (Shi (2004)). Si les conditions :

- 1) le mode x_0 est unique.
- 2) $f^{(k)}(\cdot)$ satisfait la condition de Lipschitz.
- 3) $f^{(j)}(x_0) = 0, j = 1, \dots, p-1, f^{(p)}(x_0) \neq 0$ et $f^{(p)}(x)$ est continue au voisinage de x_0 .
- 4) La fonction à noyau K est bornée et K dans l'espace de Hölder. Il remplit également les conditions $|u|K(u) \rightarrow 0$ et $|u|K^{(k)}(u) \rightarrow 0$ quand $u \rightarrow +\infty$.
- 5) K est un noyau d'ordre k et K' est un noyau d'ordre $(k-1)$.
- 6) $\frac{n}{\log(n)} h_n^{2k+3} \rightarrow 0$ et $\frac{n}{\log(n)} h_n^{(3+2(k+1)(p-1)/p)} \rightarrow 0$, lorsque $n \rightarrow \infty$

sont remplies, alors

$$\hat{x}_0 - x_0 = O\left(\left(\frac{\log(n)}{nh_n^3}\right)^{\frac{1}{(2(p-1))}}\right).$$

Estimation du mode par la méthode des séries orthogonales

Soit X_1, \dots, X_n une suite de variables aléatoire indépendante et identiquement distribuées de densité de probabilité f et de mode x_0 inconnu sur $\Omega \in \mathbb{R}$. Il s'agit d'estimer x_0 à partir des observation X_1, \dots, X_n . Supposons que le mode est unique et qu'il appartient à Ω , alors

$$f(x_0) = \max_{x \in \Omega} f(x) \quad (2.6)$$

Pour estimer le mode, nous utilisons l'estimateur de la densité par des séries orthogonales $\hat{f}_{d_n}(x)$ (1.11) et posons donc

$$\hat{f}_{d_n}(\hat{x}_0) = \max_x \hat{f}_{d_n}(x) \quad (2.7)$$

où \hat{x}_0 représente le mode associé à l'échantillon par l'estimateur \hat{f}_{d_n} . Pour établir les propriétés statistiques et asymptotiques de l'estimateur, nous avons besoin des hypothèses suivantes :

- H_1 : la densité de probabilité $f(\cdot)$ est uniformément continue sur Ω ,
- H_2 : f est 2-fois différentiable et $f''(x_0) < 0$,
- H_3 : $f'(x_0) = 0$,
- H_4 : $\hat{f}'_{d_n}(\hat{x}_0) = 0$ et $\hat{f}''_{d_n}(\hat{x}_0) < 0$.

Convergence en probabilité

Théorème 2.4.4. (Saadi(2016)). si

1. $M = \sup_{x,j} |e_j(x)| < \infty$

2. $E(\hat{f}_{d_n}(x)) \rightarrow f(x)$ lorsque $n \rightarrow \infty$,
alors,

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\hat{x}_0 - x_0| \geq \epsilon] = 0, \forall \epsilon > 0. \quad (2.8)$$

Démonstration. $f(x)$ étant uniformément continue sur Ω et ϵ étant un nombre positif donne, il existe un nombre $\eta > 0$, tel que, si $|x_0 - x| < \eta$ il en résulte toujours

$$|f(x_0) - f(x)| < \epsilon \quad (2.9)$$

Le mode x_0 étant supposé unique, pour prouver que $\hat{x}_0 \rightarrow x_0$ en probabilité, il suffira donc de montrer que

$f(\hat{x}_0) \rightarrow f(x_0)$ en probabilité lorsque $n \rightarrow \infty$.

Or,

$$|f(\hat{x}_0) - f(x_0)| \leq |f(\hat{x}_0) - \hat{f}_{d_n}(\hat{x}_0)| + |\hat{f}_{d_n}(\hat{x}_0) - f(x_0)|.$$

Par définition du mode x_0 :

$$|f(\hat{x}_0) - \hat{f}_{d_n}(\hat{x}_0)| \leq |f(x_0) - \hat{f}_{d_n}(\hat{x}_0)|.$$

D'autre part,

$$|\hat{f}_{d_n}(\hat{x}_0) - f(x_0)| \leq \sup_x |\hat{f}_{d_n}(x) - f(x)|. \quad (2.10)$$

Donc,

$$|f(\hat{x}_0) - f(x_0)| \leq 2 \sup_x |\hat{f}_{d_n}(x) - f(x)|.$$

Ainsi pour montrer que $f(\hat{x}_0) \rightarrow f(x_0)$ en probabilité, il suffira donc de montrer que

$$\sup_x |\hat{f}_{d_n}(x) - f(x)| \rightarrow 0.$$

Nous avons

$$|\hat{f}_{d_n}(x) - \mathbb{E}(\hat{f}_{d_n}(x))| = \left| \sum_{k=0}^{d_n} \hat{a}_k e_k(x) - \sum_{k=1}^{d_n} a_k e_k(x) \right| \leq \sum_{k=1}^{d_n} |(\hat{a}_k - a_k) e_k(x)|.$$

Il s'ensuit que :

$$\sup_x |\hat{f}_{d_n}(x) - \mathbb{E}(\hat{f}_{d_n}(x))| \leq \sup_{x,k} |e_k(x)| \sum_{k=1}^{d_n} |\hat{a}_k - a_k|.$$

Or,

$$\left(\sup_x |\hat{f}_{d_n}(x) - \mathbb{E}(\hat{f}_{d_n}(x))| \right)^2 \leq \left[M \sum_{k=1}^{d_n} |\hat{a}_k - a_k| \right]^2.$$

Il vient donc,

$$\mathbb{E} \left[\sup_x |\hat{f}_{d_n}(x) - \mathbb{E}(\hat{f}_{d_n}(x))| \right]^2 \leq M^2 \mathbb{E} \left[\sum_{k=1}^{d_n} |\hat{a}_k - a_k| \right]^2 = M^2 \left[\sum_{k=1}^{d_n} \text{Var}(\hat{a}_k) + \sum_{k=1}^{d_n} \sum_{j=1, j \neq k}^{d_n} \text{Cov}(\hat{a}_k, \hat{a}_j) \right].$$

Nous avons :

$$\mathbb{C}ov(\hat{a}_k, \hat{a}_j) = \mathbb{E}(\hat{a}_k \hat{a}_j) - \mathbb{E}(\hat{a}_k) \mathbb{E}(\hat{a}_j),$$

et

$$\mathbb{E}(\hat{a}_k \hat{a}_j) = \mathbb{E}\left[\frac{1}{n^2} \sum_{i=1}^n e_k(X_i) \sum_{i=1}^n e_j(X_i)\right] = \frac{1}{n^2} \mathbb{E}\left[\sum_{i=1}^n e_k(X_i) e_j(X_i) + \sum_{i=1}^n \sum_{l=1, l \neq i}^n e_k(X_i) e_j(X_l)\right].$$

Qui peut être développée comme suit :

$$\mathbb{E}(\hat{a}_k \hat{a}_j) = \frac{1}{n} \mathbb{E}[e_k(X) e_j(X)] + \frac{n-1}{n} a_k a_j.$$

On en déduit que :

$$\mathbb{C}ov(\hat{a}_k, \hat{a}_j) = \frac{1}{n} \psi_{k,j} - \frac{1}{n} a_k a_j.$$

Avec

$$\psi_{k,j} = \mathbb{E}[e_k(X) e_j(X)].$$

Nous avons :

$$\mathbb{V}ar(\hat{a}_k) = \mathbb{V}ar\left[\frac{1}{n} \sum_{i=1}^n e_k(X_i)\right] \leq \frac{1}{n} \mathbb{E}(e_k(X))^2 \leq \frac{M^2}{n} \int_{\Omega} h(x) dx = \frac{M^2}{n}.$$

Par conséquent,

$$\lim_{n \rightarrow \infty} \mathbb{V}ar(\hat{a}_k) = \lim_{n \rightarrow \infty} \mathbb{C}ov(\hat{a}_k, \hat{a}_j) = 0,$$

Alors,

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[\left(\sup_x |\hat{f}_{d_n}(x) - \mathbb{E}(\hat{f}_{d_n}(x))|\right)^2\right]^{\frac{1}{2}} = 0, \quad (2.11)$$

Mais on a aussi :

$$\mathbb{E}(\hat{f}_{d_n}(x)) \longrightarrow f(x) \text{ lorsque } n \longrightarrow \infty. \quad (2.12)$$

Donc,

$$\lim_{n \rightarrow \infty} \left[\mathbb{E}\left[\left(\sup_x |\hat{f}_{d_n}(x) - f(x)|\right)^2\right]\right]^{\frac{1}{2}} = 0, \quad (2.13)$$

Finalement,

$$\lim_{n \rightarrow \infty} P\left[\sup_x |\hat{f}_{d_n}(x) - f(x)| \leq \varepsilon\right] = 1, \forall \varepsilon > 0. \quad (2.14)$$

□

Ceci complète la preuve du théorème.

Propriétés de l'estimateur

Biais de l'estimateur

Théorème 2.4.5. (Saadi(2016)). *si*

1. $\sup_{x,k} |e_k(x)| = M < \infty$, $\sup_{x,k} |e'_k(x)| = M' < \infty$ et $\sup_{x,k} |e''_k(x)| = M'' < \infty$,
2. $d_n = o(\sqrt{n})$,
3. $\mathbb{E}(\hat{f}'_{d_n}(x_0)) \rightarrow f'(x_0)$ et $\mathbb{E}(\hat{f}''_{d_n}(x_0)) \rightarrow f''(x_0)$ lorsque $n \rightarrow \infty$,
alors l'estimateur \hat{x}_0 est asymptotiquement convergent.

Variance de l'estimateur

Théorème 2.4.6. (Saadi(2016)). *si*

1. $\sup_{x,k} |e_k(x)| = M < \infty$, $\sup_{x,k} |e'_k(x)| = M' < \infty$ et $\sup_{x,k} |e''_k(x)| = M'' < \infty$,
2. $d_n = o(\sqrt{n})$, alors

$$\text{Var}(\hat{x}_0) = \frac{\text{Var}[\hat{h}'_{d_n}(x_0)]}{[\mathbb{E}(\hat{F}''_{d_n}(x_0))]^2} + \frac{[\mathbb{E}(\hat{f}'_{d_n}(x_0))]^2 \text{Var}[\hat{F}''_{d_n}(x_0)]}{[\mathbb{E}(\hat{F}''_{d_n}(x_0))]^4} - \frac{2\mathbb{E}(\hat{f}'_{d_n}(x_0))\text{Cov}(\hat{f}'_{d_n}(x_0), \hat{F}''_{d_n}(x_0))}{[\mathbb{E}(\hat{F}''_{d_n}(x_0))]^3}.$$

Théorème 2.4.7. *si*

1. $\sup_{x,k} |e_k(x)| = M < \infty$, $\sup_{x,k} |e'_k(x)| = M' < \infty$ et $\sup_{x,k} |e''_k(x)| = M'' < \infty$,
2. $d_n = o(\sqrt{n})$, alors

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{x}_0) = 0. \quad (2.15)$$

Erreur quadratique moyenne

Théorème 2.4.8. (Saadi(2016)). *si*

1. $\sup_{x,k} |e_k(x)| = M < \infty$, $\sup_{x,k} |e'_k(x)| = M' < \infty$ et $\sup_{x,k} |e''_k(x)| = M'' < \infty$,
2. $d_n = o(\sqrt{n})$, alors

$$\begin{aligned} \mathbb{E}|\hat{x}_0 - x_0|^2 &= \frac{\text{Var}[\hat{f}'_{d_n}(x_0)]}{\mathbb{E}[\hat{F}''_{d_n}(x_0)]^2} + \frac{[\mathbb{E}(\hat{f}'_{d_n}(x_0))]^2 \text{Var}[\hat{F}''_{d_n}(x_0)]}{\mathbb{E}[\hat{F}''_{d_n}(x_0)]^4} - \frac{2\mathbb{E}(\hat{f}'_{d_n}(x_0))\text{Cov}(\hat{f}'_{d_n}(x_0), \hat{F}''_{d_n}(x_0))}{\mathbb{E}[\hat{F}''_{d_n}(x_0)]^3} \\ &\quad + \frac{\mathbb{E}^2(\hat{f}'_{d_n}(x_0))}{\mathbb{E}^2(\hat{f}''_{d_n}(x_0))} \end{aligned}$$

Théorème 2.4.9. (Saadi(2016)). *si*

1. $\sup_{x,k} |e_k(x)| = M < \infty$, $\sup_{x,k} |e'_k(x)| = M' < \infty$ et $\sup_{x,k} |e''_k(x)| = M'' < \infty$,
2. $d_n = o(\sqrt{n})$,
3. $\mathbb{E}(\hat{f}'_{d_n}(x_0)) \rightarrow f'(x_0)$ et $\mathbb{E}(\hat{f}''_{d_n}(x_0)) \rightarrow f''(x_0)$ lorsque $n \rightarrow \infty$, alors

$$\lim_{n \rightarrow \infty} \mathbb{E}|\hat{x}_0 - x_0|^2 = 0. \quad (2.16)$$

2.5 Conclusion

Dans ce chapitre, nous avons donné les différentes méthodes d'estimation non paramétrique du mode, à savoir les méthodes directes et les méthodes indirectes. Dans le chapitre suivant, nous allons présenter le travail de simulation effectué pour justifier de la qualité des estimateurs étudiés.

Chapitre 3

Simulation

3.1 Introduction

Nous présentons dans ce chapitre le travail de simulation effectué pour étayer les différents aspects théoriques abordés dans notre étude. Les simulations ont été réalisées à l'aide du logiciel R. Nous avons utilisé la version 2.3.0 pour la programmation.

R est un système d'analyse statistique et graphique créé par Ross Ihaka et Robert Gentleman. Il est à la fois un langage et un logiciel qui comporte de nombreuses fonctions pour les analyses statistiques et les graphiques. Le langage R permet, par exemple, de programmer différentes fonctions statistiques pour réaliser des analyses plus complexes.

3.2 Méthode des séries orthogonales (Méthode indirecte)

Nous présentons dans ce paragraphe le travail de simulation effectué pour justifier de la qualité des estimateurs par des séries orthogonales. L'expérimentation numérique nous servira en particulier à :

- Comparer les performances de l'estimateur du mode construit à partir de la base de saadi-adjabi à celui d'Efromovich, et ceci en comparant les erreurs quadratiques des deux estimateurs ;
- Étudier l'influence de la taille de l'échantillon sur ces différents algorithmes.

3.2.1 Plan de simulation

Nous allons faire des simulations et observer le comportement des deux estimateurs en question, calculés à partir d'échantillons simulés, censés représenter une loi connue, dont la densité de probabilité est f .

Nous utilisons pour les simulations des échantillons de loi $\beta(2, 8)$ de taille de plus en plus

grande(100,500, 700,1000,2000,5000,10000) .

Pour la densité choisie et pour chaque taille d'échantillon 100 répliques d'expérience ont été conduites.

La densité cible choisie pour tirer des échantillons est la loi beta $\beta(2, 8)$.

3.2.2 Algorithme

L'algorithme de simulation que nous avons utilisé comporte les étapes suivantes :

- ◇ Simuler un échantillon de taille n ;
- ◇ Calculer le paramètre de lissage optimal pour la base de saadi-adjabi et la base d'Efromovich en utilisant la méthode de Kronmal-Tarter .
- ◇ Calculer l'estimateur de la densité par la méthode des séries orthogonal (base saadi-adjabi et la base d'Efromovich).
- ◇ Calculer l'estimateur du mode par la méthode des séries orthogonal (base saadi et adjabi, base d'efromovich).
- ◇ Calculer le **MSE** associé à chaque estimateur du mode .

Résultats de simulation

Les résultats de la simulation sont donnés sous forme du tableau.

Le tableau contient les notations suivantes :

- d_{E-kt}^* paramètre de lissage optimal associé à l'estimateur d'Efromovich en utilisant la méthode de Kronmal-Tarter .
- d_{SA-kt}^* paramètre de lissage optimal associé à l'estimateur de saadi-Adjabi en utilisant la méthode de Kronmal-Tarter.
- MSE_{E-kt} : L'erreur quadratique moyenne du mode associée à la base d'Efromovich.
- MSE_{SA-kt} : L'erreur quadratique moyenne du mode associée à la base de saadi-adjabi.

n	d_{E-kt}^*	d_{SA-kt}^*	$\text{MSE}_{E-kt}(\hat{x}_0)$	$\text{MSE}_{SA-kt}(\hat{x}_0)$
100	4	13	0.015625	0.000484
500	9	15	0.000081	0.0000324
700	10	15	0.000529	0.000064
1000	12	16	0.000484	0.000324
2000	16	16	0.000400	0.000256
5000	16	17	0.000361	0.000256
10000	19	17	0.000256	0.000100

TABLE 3.1 – Le paramètre de lissage optimal et la Variation du $\text{MSE}(\hat{x}_0)$ en fonction de n pour la Loi Beta $\beta(2, 8)$, en utilisant la base saadi-Adjabi et la base d'Efromovich.

3.2.3 Interprétation

D'après les résultats de Tableau (3.1), On remarque que :

- * le **MSE** des deux estimateurs diminue, quand on augmente la taille de l'échantillon.
- * L'estimation du mode en utilisant la base de saadi-adjabi donne des meilleurs résultats que celui basé sur la base d'Efromovich pour la loi Beta $\beta(2, 8)$.

3.3 Estimateur à noyau du mode

L'expérimentation numérique nous servira en particulier à :

- Comparer les performances de l'estimateur du mode construit à partir du noyau triangulaire à celui construit à partir du noyau Cosine, et ceci en comparant les erreurs quadratiques des deux estimateurs ;
- Étudier l'influence de la taille d'échantillon sur ces différents algorithmes.

3.3.1 Plan de simulation

Nous allons faire des simulations et observer le comportement des deux estimateurs en question, calculés à partir d'échantillons simulés, censés représenter une loi connue, dont la densité de probabilité est f .

Nous utilisons pour les simulations des échantillons de loi Normal $\mathcal{N}(0, 1)$ de taille de plus en plus grande (100, 500, 700, 1000, 2000, 5000, 10000).

Pour la densité choisie et pour chaque taille d'échantillon 100 répétitions d'expérience ont été conduites.

La densité cible choisie pour tirer des échantillons est la loi Normal $\mathcal{N}(0, 1)$.

3.3.2 ALgorithme

L'algorithme de simulation que nous avons utilisé comporte les étapes suivantes :

- ◇ Simuler un échantillon de taille n ;
- ◇ Calculer le paramètre de lissage optimal pour le noyau Cosine et le noyau Triangulaire en utilisant la méthode de validation croisée non biaisée.
- ◇ Calculer l'estimateur de la densité par la méthode du noyau (Cosine et Triangulaire).
- ◇ Calculer l'estimateur du mode par la méthode de noyau (Cosine et Triangulaire).
- ◇ Calculer le MSE associé à chaque estimateurs.

Résultats de simulation

Les résultats de la simulation sont donnés sous forme du tableau qui contient les notations suivantes :

- $h_{Tri-ucv}^*$ paramètre de lissage optimal associé au noyau triangulaire en utilisant la méthode de validation croisée non biaisée.
- $h_{Cos-ucv}^*$ paramètre de lissage optimal associé au noyau Cosine en utilisant la méthode de validation croisée non biaisée.
- MSE_{Tri} L'erreur quadratique moyenne du mode associée au noyau Triangulaire.
- MSE_{Cos} L'erreur quadratique moyenne du mode associée au noyau Cosine.

n	$h_{Tri-ucv}^*$	$h_{Cos-ucv}^*$	$\mathbf{MSE}_{Tri}(\hat{x}_0)$	$\mathbf{MSE}_{Cos}(\hat{x}_0)$
100	0.3868399	0.4185191	0.079579248	0.073533739
500	0.3277147	0.3080413	0.045463490	0.037438643
700	0.2845050	0.3061369	0.028298696	0.024759523
1000	0.2386967	0.2782037	0.027179805	0.024759523
2000	0.2044525	0.2114239	0.016809905	0.012954034
5000	0.1561501	0.1939811	0.012394880	0.009589964
10000	0.1082199	0.1744787	0.009321495	0.007844148

TABLE 3.2 – Le paramètre de lissage optimal et la Variation du $\mathbf{MSE}(\hat{x}_0)$ optimal en fonction de n pour la Loi normale $\mathcal{N}(0, 1)$, en utilisant le noyau triangulaire et le noyau cosine.

Interprétation

D'après les résultats du Tableau (3.2), On remarque que :

- * le \mathbf{MSE} des deux estimateurs diminue, quand on augmente la taille d'échantillon.
- * L'estimation du mode en utilisant le noyau Cosine donne des meilleurs résultats que celui basé sur le noyau triangulaire pour la loi Normal $\mathcal{N}(0, 1)$.

3.4 La méthode de Grenander (Méthode directe)

3.4.1 Algorithme

L'algorithme de simulation que nous avons utilisé comporte les étapes suivantes :

- ◇ Simuler un échantillon (x_1, \dots, x_n) par ordre croissant, $0 \leq x_i \leq 1$, de loi $\beta(2, 8)$.
- ◇ Calculer le mode théorique .
- ◇ La taille des échantillon est fixé à (100, 500, 700, 1000, 2000, 5000, 10000).
- ◇ k est égal la partie entière de $\log(n)$.
- ◇ p est égal la partie entière de $\frac{k}{2}$.
- ◇ Calculer le mode estimé par l'estimateur de Grenander : $M_{p,k}^* = \frac{\frac{1}{2} \sum_{i=1}^{n-k} (x_{i+k} + x_i) / (x_{i+k} - x_i)^p}{\sum_{i=1}^{n-k} 1 / (x_{i+k} - x_i)^p}$.
- ◇ Calculer le \mathbf{MSE} associé à l'estimateur du mode.

Résultats de simulation

Les résultats de la simulation sont donnés sous forme d'un tableau.

n	x_0	\hat{x}_0	$\mathbf{MSE}(\hat{x}_0)$
100	0,125	0,1546184	0.00087725
500	0,125	0,1476627	0.000513598
700	0,125	0,1472451	0.000494844
1000	0,125	0,1471121	0.000488945
2000	0,125	0,1470034	0.00048415
5000	0,125	0,1431822	0.000330592
10000	0,125	0,1431322	0.000328777

TABLE 3.3 – Variation du $\mathbf{MSE}(\hat{x}_0)$ optimal en fonction de n pour la Loi $\beta(2, 8)$, en utilisant l'estimateur de Grenander.

Interprétation

D'après les résultats du Tableau (3.3), On remarque que :

- * le \mathbf{MSE} de l'estimateur de Grenander diminue, quand on augmente la taille d'échantillon.
- * les valeurs du mode estimé par la méthode de Grenander sont au voisinage du mode théorique.

3.5 Étude comparative

Les tableaux (3.4) et (3.5) nous permettent de comparer les résultats obtenus par la méthode directe (Grenander) et les méthodes indirectes (la méthode du noyau et la méthode des fonctions orthogonales), pour les différentes densités tests :

n	x_0	$\hat{x}_0(G)$	$\mathbf{MSE}_G(\hat{x}_0)$	d_{SA-kt}^*	$\mathbf{MSE}_{SA}(\hat{x}_0)$
100	0,125	0,1546184	0.00087725	13	0.000484
500	0,125	0,1476627	0.000513598	15	0.0000324
700	0,125	0,1472451	0.000494844	15	0.000064
1000	0,125	0,1471121	0.000488945	16	0.000324
2000	0,125	0,1470034	0.00048415	16	0.000256
5000	0,125	0,1431822	0.000330592	17	0.000256
10000	0,125	0,1431322	0.000328777	17	0.000100

TABLE 3.4 – Variation du $\mathbf{MSE}(\hat{x}_0)$ pour la loi $\beta(2, 8)$, en utilisant l'estimateur de Grenander et l'estimateur de saadi-adjabi.

n	x_0	$\hat{x}_0(G)$	$MSE_{Grenander}(\hat{x}_0)$	h_{ucv}	$MSE_{Noyau\ normal}(\hat{x}_0)$
100	0.25	0.3417184	0.008412	0.467450	0.0738366
500	0.25	0.3169453	0.0044817	0.2825067	0.05312320
700	0.25	0.3154436	0.0042829	0.2438352	0.0446732
1000	0.25	0.3132834	0.0040048	0.2161307	0.0242696
1000	0.25	0.3128587	0.0039512	0.2069449	0.0168327
2000	0.25	0.3011245	0.0026137	0.1857220	0.0109126
10000	0.25	0.3003501	0.0025351	0.1686533	0.0084183

TABLE 3.5 – Variation du $\mathbf{MSE}(\hat{x}_0)$ pour la loi $G(2; 4)$, en utilisant l'estimateur de Grenander et l'estimateur à noyau (gaussien).

Interprétation

D'après les résultats des Tableaux (3.4) et (3.5), On remarque que :

- * le \mathbf{MSE} des deux estimateurs diminue, quand on augmente la taille d'échantillon.
- * L'estimation du mode par des séries orthogonales donne des meilleurs résultats que celui de Grenander pour la loi $\beta(2, 8)$.
- * L'estimation du mode en utilisant l'estimateur de Grenander donne des meilleurs résultats que celui basé sur le noyau gaussien pour la loi $G(2; 4)$.

Conclusion générale et perspectives

Ce travail est une contribution au problème d'estimation de la densité de probabilité et du mode d'une densité.

Dans une première partie, nous avons exposé les différentes méthodes d'estimation de la densité de probabilité, à savoir l'estimation par des séries orthogonales, l'estimation par histogramme et l'estimation par la méthode du noyau. L'estimateur à noyau est fonction de deux paramètres : la fonction K appelée noyau et h appelé paramètre de lissage ou fenêtre. Si le choix du noyau K n'est pas un problème dans l'estimation de la densité de probabilité, il n'en est pas de même pour le choix du paramètre de lissage h qui ne dépend que de la taille de l'échantillon. Dans ce chapitre, nous avons exposé les différentes méthodes de sélection du paramètre de lissage, à savoir Les méthodes reposant sur la validation croisée et l'autre classe de méthode dite plug-in (ré-injection).

Nous avons également donné le principe d'estimation par la méthode des fonctions orthogonales et nous avons indiqué les principaux résultats relatifs à la convergence de l'estimateur suivant divers modes stochastiques ainsi que la méthode de sélection du paramètre de lissage.

Dans Le deuxième chapitre, nous avons exposé les différentes méthodes d'estimation du mode, à savoir la méthode directe et la méthode indirecte.

Enfin, afin de tester l'applicabilité des différentes méthodes d'estimation du mode, nous avons simulé des densités de probabilité tests : la loi normale centrée réduite $\mathcal{N}(0, 1)$ et la loi $\beta(2, 8)$. L'algorithme de simulation que nous avons utilisé a permis :

- De calculer l'estimateur du mode de la densité de probabilité par la méthode du noyau et la méthode des fonctions orthogonales.
- De calculer l'estimateur du mode par la méthode directe en particulier par la méthode de Grender.

Les résultats numériques obtenus montrent que :

- Les performances de l'estimateur associées à la base saadi sont meilleures que celles de l'estimateur d'Efromovich.
- Les performances de l'estimateur associées à l'estimateur en utilisant le noyau cosinus

sont meilleures que celles de l'estimateur d'Epanichnikov.

- Les performances des estimateurs s'améliorent lorsque la taille de l'échantillon augmente.

Perspectives de recherche

Enfin, il serait judicieux d'explorer certaines pistes dans des recherches à venir :

- Il nous semble possible également d'étendre les résultats de ce travail à l'étude du mode conditionnel.
- On peut aussi étudier le prolongement de nos résultats au cas de données fonctionnelles.
- Il est toujours possibles d'affaiblir les hypothèses, comme par exemple celles sur la dépendance des observations et celles où l'on peut supposer que les bases orthogonales sont simplement bornées.

Bibliographie

- [1] Abattouy N., Valero A., Benajiba M.H., Lozano J., Martín-Sánchez J., 2011. *Anisakis simplex*.l. parasitization in mackerel (*Scomber japonicus*) caught in the North of Morocco - Prevalence and analysis of risk factors. *International Journal of Food Microbiology*. 150, 136-139.
- [2] Asselin de Beauville, J. P. (1978). Estimation non paramétrique de la densité de probabilité et du mode. Exemple de la distribution gamma. *Revue de Statistique Appliquée* 26 :47-70.
- [3] A.W. Bowman.(1984). An alternative method of cross-validation for the smoothing density estimates. *Biometrika*, 71 :353-360 .
- [4] Bickel,D.R.(2002).Robust estimators of the mode and skewness of continuous data.*Computational Statistics and Data Analysis* 39 :153-163.
- [5] Bosq, D., Lecoutre, J. (1987). *Théorie de l'estimation fonctionnelle*. Economica Edition.
- [6] Bosq, D. (2002). Estimation localement suroptimale et adaptative de la densité. *Comptes Rendus de l'Académie des Sciences* 334 :591-595.
- [7] Bosq, D. (2005), Estimation suroptimale de la densité par projection. *Can. J. Statist.*, Vol. 433 No. 1, 21-37.
- [8] Bickel,D.R.(2002).Robust estimators of the mode and skewness of continuous data.*Computational Statistics and Data Analysis* 39 :153-163.
- [9] Cencov.(1962). Evaluation of unknow distribution density from observation. *Sov-Maths*, (3) :1559-1562 .

-
- [10] Chernof,H.(1964).Estimation of the mode.Ann.Instit.Statist.Math.16 :31-41.
- [11] D.W. Scott and G.R. Terrell. (1987). Biased and unbiased cross-validation in density estimation. Journal of the American Statistical Association, 82 :1131-1146.
- [12] Diggle, P. J., Hall, P. (1986). The selection of terms in an orthogonal series density estimator. Journal of the American Statistical Association, 81(393), 230-233.
- [13] L. Devroye and L. Györfi. (1984). Nonparametric Density Estimation. Wiley Series in Probability and Mathematical Statistics, Montreal, Budapest.
- [14] Devroye.L. (1983). The equivalence of weak, strong and complete convergence in L^1 for kernel density estimates. The Annals of Statistics, 11 :896-904.
- [15] Devroye. L. (1987). A course in density estimation. Birkhauser Boston Inc..
- [16] Deheuvels,P.(1977). Estimation non paramétrique de la densité par histogrammes généralisés. Revue de Statistique Appliquée 25, 5-42.
- [17] Devroye, L. P., Wagner, T. J. (1980, March). The strong uniform consistency of kernel density estimates. In Multivariate Analysis V : Proceedings of the fifth International Symposium on Multivariate Analysis (Vol. 5, pp. 59-77).
- [18] E. Parzen.[1962] On estimation of a probability density function and mode. Annals of Mathematical Statistics, 33 :1065-1076 .
- [19] Epanechnikov .V. A. (1969). Nonparametric estimation of a multidimensional probability density. Theoretical probability, applications, 14 :153-158.
- [20] Eddy, W. F. (1980). Optimum kernel estimators of the mode. The Annals of Statistics, 8(4), 870-882.
- [21] Grenander,U.(1965).Some direct estimates of the mode.Ann.Math.Statist.36 :131-138.
- [22] Grund, B. and Hall, P. (1995). On the minimisation of L^p error in mode estimation. The Annals of Statistics, 23(6), 2264-2284.
- [23] Hall, P. (1986), On the rate of the convergence of orthogonal series density estimation. J.Roy. Statist.Soc., Vol. 48, 115-122.

-
- [24] Hart JD.(1997). Nonparametric Smoothing and Lack-of-Fit Tests. New York : Springer .
- [25] Herrmann, E. and Ziegler, K. (2004). Rates of consistency for nonparametric estimation of the mode in absence of smoothness assumptions. *Statistics probability letters*, 68(4), 359-368.
- [26] J.S.Simonoff.(1996). Smoothing Methods in Statistics. Springer-Verlag.
- [27] K.Lagha . S.Adjabi (2016) Nonparametric sequential estimation of the probability density function by orthogonal series, *Communications in Statistics - Theory and Methods*, DOI : 10.1080/03610926.2015.1115075.
- [28] Krylov. A.(1955). Collected works.Izdat. Akad. Nauk SSSR, Moscow.,XII, part one.
- [29] Kolmogorov. A and Fomin. S.(1957). Elements of the theory of functions and functional analysis.Graylock Press. Rochester, N. Y.
- [30] Romano, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *The Annals of Statistics*, 629-647.
- [31] M. Rosenblatt.(1956). Remarks in some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27 :832-837.
- [32] M. Rudemo.(1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9 :65-78.
- [33] N. Saadi, S. Adjabi. (2009). On the estimation of the probability density by trigonometric series.*Communications in Statistics-Theory and Methods*, 38(3583-3595).
- [34] Nadaraya, Elizbar Akakevi. (1989). Nonparametric estimation of probability densities and regression curves. Dordrecht : Kluwer Academic Publishers.
- [35] Nadaraya. E. (1964). Some new estimates for distribution function. *Theory of Probability and its Application*,9, 497-500.
- [36] Nadaraya,E.(1965). On nonparametric estimation density function and regression.*Theory Probab P.P.L*, 10,186-190 .

- [37] P. Hall.(1982). Cross-validation in density estimation. *Biometrika*, 69 :383-390.
- [38] Prakasa,Rao, B.L.S. (1983). *Nonparametric Functional Estimation*. Academic Press,New York.
- [39] R. Kronmal and M. Tarter. (1968) The estimation of probability densities and cumulatives by fourier series methods. *J. Amer. Statist. Assoc*, (63) : 925-952.
- [40] S.Efromovich.(2010)orthogonal series density estimation. *Interdisciplinary Review, Computational Statistics*, 2,467-476.
- [41] S.Efromovich .(1997) Density estimation for the case of supersmooth measurement error. *J Am Stat Assoc* ,92 :526-535.
- [42] Sanson. A.(1959). *Orthogonal functions*.Pure and Applied Math. Interscience Publ.New York.
- [43] Shi, X., Wu, Y., and Miao, B. (2009). A note on the convergence rate of the kernel density estimator of the mode. *Statistics and Probability Letters*, 79(17), 1866-1871.
- [44] S.Julian.(2008).An Assessment of Hermite Function Based Approximation of Mutuel Infor-mation Applied to Independant Component Analysis.10.745-756.
- [45] Tarter, M. E, and Lock, M. D. (1993). *Model-free curve estimation* (Vol. 56). CRC Press.
- [46] Thompson, James R. et Tapia, Richard A.'1990). *Nonparametric function estimation, modeling, and simulation*. Society for industrial and applied mathematics .
- [47] Van Ryzin, J. (1969). On strong consistency of density estimates. *The Annals of Mathematical Statistics*, 40(5), 1765-1772.
- [48] Vieu, P. (1996). A note on density mode estimation. *Statistics probability letters*, 26(4), 297-307.
- [49] Wegman,E.J.(1971).A note on the estimation of the mode.*Ann.Math.Statist*.42 :1909-1915.

Résumé

Ce travail est une contribution au problème d'estimation de la densité de probabilité et du mode d'une densité. Dans une première partie, nous avons exposé les différentes méthodes d'estimation de la densité, à savoir l'estimation par les séries orthogonales, l'estimation par histogramme et l'estimation par la méthode du noyau. L'estimateur à noyau est fonction de deux paramètres : la fonction K appelée noyau et h appelé paramètre de lissage ou fenêtre. Si le choix du noyau K n'est pas un problème dans l'estimation de la densité de probabilité, il n'en est pas de même pour le choix du paramètre de lissage h qui ne dépend que de la taille de l'échantillon. Dans ce mémoire, nous avons exposé les différentes méthodes de sélection du paramètre de lissage, à savoir Les méthodes reposant sur la validation croisée et l'autre classe de méthode dite plug-in (ré-injection). Nous avons également donné le principe d'estimation par la méthode des fonctions orthogonales et nous avons indiqué les principaux résultats relatifs à la convergence de l'estimateur suivant divers modes stochastiques ainsi que la méthode de sélection du paramètre de lissage. La deuxième partie de ce mémoire est consacrée à l'étude de l'estimation non paramétrique du mode. Dans ce but, nous avons exposé les différentes méthodes d'estimation, à savoir les méthodes directes et les méthodes indirectes. Des comparaisons des différentes méthodes sont exposées et validées par des simulations numériques.

Mots-clés : Estimation non paramétrique de la densité , estimation du mode, séries orthogonales, méthode de noyau.

Abstract

This work is a contribution to the problem of estimating the probability density and the mode of a density. In a first part, we exposed the different methods of density estimation, namely estimation by orthogonal series , histogram estimation and kernel method estimation. The kernel estimator is a function of two parameters : the function K called the kernel and h called the smoothing parameter. If the choice of the kernel K is not a problem in the estimation of the probability density, it is not the same for the choice of the smoothing parameter h which does not depends only on the sample size. In this work, we have exposed the different methods of selecting the smoothing parameter, namely the methods based on cross validation and the other class of method called plug-in (re-injection). We have also given the principle of estimation by the method of orthogonal functions and we have indicated the main results relating to the convergence of the estimator according to various stochastic modes as well as the method of selection of the smoothing parameter. The second part is devoted to the study of the nonparametric estimation of the mode. For this purpose, we have exposed the different estimation methods, namely direct methods and indirect methods. Comparisons of the different methods are exposed and validated by numerical simulations.

Keywords : Nonparametric density estimation, mode estimation, orthogonal series, kernel.