N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBÈS

THESE DE DOCTORAT

Présentée par

DJERFI Kouider

Spécialité : Mathématiques

Option : Probabilités et Statistique

Intitulée

Généralités sur l'estimateur de type Stein de l'opérateur de régression pour des données fonctionnelles

Soutenue le 05 Mars 2020

Devant le jury composé de :

Président :

Mr RABHI Abbès Professeur à l'université de S.B.A.

Examinateurs :

Mr MECHAB Boubaker M.C.A. à l'université de S.B.A.

Mr ALLAM Abdelazziz M.C.A. à l'université de Tlemcen.

Directeur de thèse :

Mr MADANI Fethi M.C.A . à l'université de Saida.

Année universitaire : 2019/2020

Remerciements

Je remercie, en premier lieu, notre Dieu qui nous a donné de la force pour effectuer le présent travail. En second lieu, je tiens à adresser mes vifs remerciements à mon directeur de thèse **Dr. Madani Fethi** pour son aide, ses conseils et tout le temps qu'il a consacré pour mener à bien ce travail. Je salut en lui son savoir faire, sa patiences et ses connaissances dont il ma transmet. Je tiens aussi à remercier vivement le **Professeur Abbes Rabhi** d'accepter de présider le jury de soutenance de cette thèse. Je remercie également les membres de jury examinateurs Mrs : le **Dr. Boubaker Mechab** et le **Dr. Allam Abdelazziz** de la confiance qu'ils m'accordent et de l'intérêt qu'ils témoignent pour ce travail en acceptant de faire partie du jury. Je tiens aussi à remercier les membres du laboratoire de Modèles stochastiques, Statistique et Applications et spécialement son directeur Professeur Abdeldjebbar Kandouci pour les conditions favorables aux quelles j'ai profité pour achever ce travail. Je remercie également le Professeur Idir Ouassou pour ses précieux conseils et sa disponibilité pour discuter du sujet de cette thèse à l'occasion du séjour scientifique que j'ai effectue à l'université Cadi Ayyad à Marrakech, Maroc. Merci enfin à mon cher collègue et ami le Professeur Djellouli Ghaouti pour son aide à la rédaction de cette thèse.

Table des matières

Résumé

La question principale évoquée dans cette thèse est l'amélioration par rétrécissement de l'estimateur usuel pour un modèle de régression non-paramétrique fonctionnelle (fixed-design), sous une fonction de perte non-quadratique.
Notre contribution à ce sujet est la proposition de certaines classes de fonctions de perte généralisant celle proposée par Zellner dans [65] (Balanced loss function). Par conséquent, on montre que l'estimateur usuel (l'éstimateur à noyau dans notre cas) est :

1. Inadmissible relativement à certaines classes de ces fonctions de perte équilibrées, d'où la possibilité de l'améliorer par rétrécissement. Cela se réalise dans le cas des fonctions de perte : Quadratic balanced loss , Weighted quadratic Balanced loss, Linex balanced loss.
2. Admissible dans certains cas, notamment dans le cas des fonctions de perte : Logistic balanced loss, Reflected normal balanced loss, Absolute balanced loss avec $1/2 \leq \omega < 1$. Cela est un phénomène qui se réalise indépendamment à la dimension, contrairement à ce qui a été reconnu sur le phénomène de Stein.

English summary

The main question raised in this thesis is the improvement of the usual estimator for a nonparametric functional regression model (fixed design), under a non-quadratic loss function. Our contribution on this subject is the proposition of some classes of loss functions generalizing that proposed by Zellner in [65] (balanced loss function). Therefore, we show that the usual estimator (the kernel estimator in this case) is :

1. Inadmissible for some classes of these weighted loss functions, hence the possibility to improve it by retraction. This is done in the case of loss functions : quadratic balanced loss, balanced weighted quadratic loss, Linex balanced loss.
2. Admissible in some cases, particularly for loss functions : balanced logistic loss, weighted normal loss reflected, absolute balanced loss with $1/2 \leq \omega < 1$. This is a phenomenon that occurs independently of the dimension, contrary to what was recognized on the Stein phenomenon.

Chapitre 1

Introduction générale

1.1 État de l'art de la statistique fonctionnelle

La statistique fonctionnelle a connu un grand potentiel en terme d'applications dans des domaines variés : la biométrie, la météorologie, la Biochimie, la Médecine et dans de nombreux autres domaines, en raison du développement des outils informatiques et leurs capacités de stockage.
Historiquement, les premiers ouvrages considérant la statistique fonctionnelle dans un contexte paramétrique sont les monographies de Ramsay et Silverman (cf. [51], [52]), ils ont étudié des méthodes statistiques pour des variables fonctionnelles, ainsi le livre de Bosq [5] dans le cadre de séries chronologiques. Dans le contexte non-paramétrique, les premiers ouvrages faisant référence en la matière sont ceux de Ferraty et Vieu [22], ils ont considéré un modèle non-paramétrique fonctionnelle dans des espaces vectoriels semi- normé, étendue par la suite pour la prévision en séries chronologiques en [23].

En (2004) les mêmes auteurs [24] ont construit un estimateur à noyau pour la régression et obtiennent des vitesses de convergence pour cet estimateur ainsi une solution au problème de fléau de la dimension, ce phénomène bien connu en statistique non-paramétrique concerne la dégradation considérable de la qualité de l'estimation lorsque la dimension augmente, ainsi il rend les vitesses de convergence très faibles. La solution de ce problème est la mesure de probabilité des petites boules ou la propriété de concentration qui intervient dans les vitesses de convergence.
La monographie de Ferraty et Vieu [25] offre un large choix d'étude concernant le comportement asymptotique des estimateurs à noyau pour la fonction de répartition conditionnelle, la densité conditionnelle et ses dérivées, le mode (avec ou sans conditionnement), la médiane (avec ou sans conditionnement) ainsi que le quantile conditionnel et la régression .cet ouvrage est devenu une référence en statistique non-paramétrique tant sur le plan pratique que sur le plan théorique.

1.2 L'estimateur de type Stein : Motivation et contexte bibliographique

En 1956, Charles Stein [57] découvre un phénomène statistique, auquel la communauté statisticienne donnera éventuellement son nom en parlant d'effet Stein ou de paradoxe de Stein. Ce phénomène consiste en la non admissibilité, sous coût quadratique, de l'estimateur du maximum de vraisemblance de la moyenne d'une loi gaussienne p-dimensionnelle lorsque la dimension p est supérieure ou égale à 3. Le caractère "phénoménal" fait lien entre l'inefficacité de l'estimateur usuel et à la coupure de dimension ainsi mise en évidence. En effet, si $p = 1$ ou si $p = 2$, le maximum de vraisemblance est admissible sous coût quadratique. En revanche, cette admissibilité est perdue pour tout $p > 2$ (cf. [40] et [59]).

Ce phénomène signifie que, disposant d'un p-uplet de moyennes, estimer celles-ci individuellement (en les considérant l'une après l'autre) ou les estimer globalement (en les envisageant comme formant un vecteur) n'est pas équivalent. Une différence est ainsi mise en évidence selon que $p < 3$ ou que $p > 3$. Le "paradoxe" tient à ce que le regroupement des p estimateurs admissibles peut conduire à un estimateur inadmissible alors même que les observations sous-jacentes sont des quantités indépendantes (pouvant, en pratique, n'avoir aucun lien entre elles).

Le résultat de Stein a des conséquences fondamentales : Il met en évidence que le caractère sans biais d'un estimateur n'est pas (ou n'est plus) la propriété utile qu'il faille rechercher. Le phénomène souligne le fait que la théorie de la décision apporte des critères qui, via la fonction de coût, peuvent s'avérer incompatibles avec les critères statistiques classiques. L'objectif de recherche d'un éventuel estimateur meilleur que tous les autres paraît alors sans objet.

Le paradoxe de Stein a ouvert la voie à de nombreuses recherches en estimation. Ainsi la littérature sur les estimateurs à rétrécissement est abondante.

La démarche conduisant à ces estimateurs repose tout d'abord sur le choix d'un estimateur standard (pour le problème d'estimation considéré), on entend par là que cet estimateur possède des propriétés telles que son usage s'impose naturellement (soit, par exemple, la minimaxité qui est présente dans le cas gaussien). Le but est alors de déterminer des estimateurs qui améliorent l'estimateur standard, le phénomène étant illustré par une coupure de dimension.

Dans le cas général, notons que Stein [57] et Brown [9] ont montré que le phénomène de Stein survient pour plusieurs lois d'échantillonnage, pour plusieurs fonctions de coût et pour la plupart des problèmes d'estimation ; l'estimation de coût est aussi concernée par ce phénomène (cf. [41], [26] et [27]). Ces derniers auteurs ont aussi exhibé des classes d'estimateurs contenant des estimateurs améliorés pour le même problème dans un cadre distributionnel plus large. Car, en dehors du cas gaussien, moins de choses ont été développées pour construire des estimateurs minimax explicites améliorant le meilleur estimateur équivariant.

Les développements dans le domaine de l'amélioration des estimateurs, qui ont suivi les résultats de Stein sur l'estimation multivariée. Le problème d'admissibilité et da minimaxité (à la fois) a été introduit par Johnstone [41], qui fait remarquer, que même si on choisit les estimateurs en fonction de leurs risques fréquentistes (en moyennant sur toutes les valeurs possibles des observations), il montre En utilisant des bases orthonormales d'ondelettes à support compact, qu'on peut développer une méthode non linéaire qui fonctionne dans le domaine des ondelettes par simple rétrécissements non linéaire des coefficients d'ondelettes empiriques, Le rétrécissement peut-être réglé pour être presque minimax sur n'importe quel membre d'une large classe d'estimateurs admissibles.

Zellner [65] a introduit la notion de fonction de perte équilibrée dans le contexte d'un modèle linéaire général afin de refléter à la fois la qualité de l'ajustement et la précision de l'estimation. Nous étudions cette notion dans la perspective d'unifier une variété de résultats à la fois fréquentistes et bayésiens. Nous montrons en général que les résultats fréquentistes et bayésiens pour une perte équilibrée découlent et impliquent également des résultats connexes pour des fonctions de perte quadratiques ne reflétant que la précision de l'estimation.

Brandwein et al. discutent dans [8] les développements récents et avancées en matière d'estimation à rétrécissement (de type Stein) pour les distributions à symétrie sphérique.Une certaine importance est accordée au développement de l'intuition quant aux raisons pour lesquelles l'amélioration par rétrécissement devrait fonctionner dans les problèmes de localisation, que la population sous-jacente soit normale ou non.

Une attention considérable est consacrée à la généralisation du "lemme de Stein" qui occupe une grande partie du développement théorique de l'estimation minimax améliorée pour les distributions à symétrie sphérique. L'accent est mis principalement sur les résultats de robustesse distributionnelle dans les cas où un vecteur résiduel est disponible pour estimer un paramètre d'échelle inconnu et, en particulier, sur la recherche d'estimateurs simultanément généralisés de Bayes et de minimax sur de grandes classes de distributions à symétrie sphériques.

Fait intéressant, en 1995, dans le bulletin de nouvelles de la Royal Statistical Society, Efron a prédit que le rétrécissement et les méthodes bayésiennes empiriques constitueraient un domaine de recherche important au début du 21e siècle. La stratégie de rétrécissement continue d'offrir des outils utiles pour une estimation efficace.

1.3 Contribution de la thèse

Notre travail consiste en la détermination des conditions nécessaires ou suffisantes pour les quelles un estimateur de type Stein est admissible, ce dernier est obtenu a partir d'un estimateur cible (l'estimateur à noyau) et le risque est calculé relativement à une certaine classes de fonctions de perte équilibrées (balanced loss functions).

On considère donc le modèle régression non-paramétrique fonctionnelle $Y = r(X) + \epsilon$, où Y est une variable aléatoire réelle et X est une variable fonctionnelle à valeurs dans un espace semi-normé. Le but est de trouver les conditions d'admissibilité des estimateurs de type Stein de ce modèle relativement à une classe de fonctions de perte équilibrées.

Notre méthode consiste à comparer le risque avec celui obtenu dans le cas d'une perte quadratique. L'originalité dans ce travail se résume en deux volets :

1. La proposition d'une nouvelle classe de fonctions de perte généralisant celle proposée par Zellner dans [65]. Cette fonction de perte de Zellner a été proposée dans le but de relier entre les deux critères : la précision de l'estimateur et la qualité d'ajustement. Or cette fonction de perte est basée uniquement sur la perte quadratique, la classe qu'on propose ici peut couvrir toute les fonctions de perte connues dans la littérature : Weighted quadratic loss, Logistic loss, Reflected normal, Linex balanced loss,...
2. Démontrer la possibilité d'améliorer l'estimateur à noyau de la régression pour un modèle fonctionnel par rétrécissement (shrinkage me-

thods) dans certains cas, tandis que pour d'autres cas on montre que cet estimateur est admissible.

1.4 Le plan de la thèse

Cette thèse est constituée de cinq chapitres :

Le premier chapitre est une introduction générale ou on présente des notes historiques et bibliographiques des deux thèmes principales abordés à cette thèse :

- La régression non-paramétrique fonctionnelle.
- Le phénomène de Stein et les estimateurs de type Stein.

Dans le deuxième chapitre on présente un aperçu sur la régression non-paramétrique. On définit l'estimateur à noyau pour le modèle de régression $Y = r(X) + \varepsilon$, où la variable aléatoire explicative X est réelle, vectorielle ou fonctionnelle. Dans les trois cas on présente quelques résultats sur les propriétés asymptotiques de l'estimateur.

Le troisième chapitre est consacré au phénomène de Stein. On évoque la question d'inadmissibilité de l'estimateur de maximum de vraisemblance relativement à la fonction de perte quadratique. On explique ce phénomène sous deux modèles classique : un modèle paramétrique et un modèle non-paramétrique. La résolution du problème conduit à la définition d'une famille d'estimateurs de types Stein et James-Stein, et motive la naissance d'une méthode d'amélioration des estimateurs par rétrécissement (contraction) et aussi des estimateurs pseudo-orthogonals (ridge). Par conséquent, les stratégies d'estimation de rétrécissement de type Stein ont bénéficié d'une grande attention de la part des chercheurs. La plupart des stratégies d'estimation par régularisation étendent puissamment les procédures de Stein. De nombreux travaux de recherche sont en cours dans le domaine des données de grande dimension, où le nombre de variables est supérieur à celui des observations.

Dans le quatrième chapitre on présente la problématique liée au phénomène de Stein dans un contexte non-paramétrique fonctionnel, et on explique notre contribution à la résolution de la problématique posée, contribution qui fait l'objet d'un article intitulé : "Admissibility results under some balanced loss functions for a functional regression model" dans le journal Comptes rendus - Mathematics.

Enfin, dans le dernier chapitre, nous récapitulons les résultats obtenus lors de ces travaux afin de faire le point sur les sujets en question. Nous ajoutons aussi une discussion sur les problèmes à aborder comme une suite naturelle de ce travail.

Chapitre 2

Sur la régression non-paramétrique

L'estimation non-paramétrique, et en particulier l'estimation de la régression, constitue un champ de recherche important de la Statistique depuis trois ou quatre décennies. Le lecteur pourra consulter les revues bibliographiques de Collomb (cf. [16], [17]) qui dès le début des années 80 font déja état de développements nombreux et variés sur ce thème, puis les ouverages de Hardle (cf. [36], [37]) et enfin pour terminer l'ouvrage collectif de Schimek [56] qui adresse un peu le bilan actuel des diverses connaissances en la matière. D'autres ouvrages généraux sur le thème incluent ceux de Bosq et Lecoutre [4] et Wand et Jones [62].

Nous nous plaçons le long de ce chapitre dans le cadre de l'estimation d'une fonction de régression, que nous noterons r. La variable aléatoire à expliquer sera notée Y, tandis que la variable aléatoire explicative sera notée X. Toutes les variables considérées seront supposées être définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$. Les divers modèles de régression que nous abordons peuvent tous être écrits sous la forme :

$$Y = r(X) + \varepsilon, \tag{2.1}$$

où, ε est une variable aléatoire centrée et independante de X. Dans tout ce qui suit, nous nous limiterons a une variable aléatoire Y (variable à expliquer) réelle, et nos diférents modèles se distingueront selon la nature de la variable X et selon la nature de la relation fonctionnelle r liant Y à X. La caractérisation d'un modèle de régression 2.1 se fait par une hypothèse de type

$$r \in \mathcal{C}, \tag{2.2}$$

où, $\mathcal{C}$ est une classe quelconque de fonctions. Sous cette hypothèse, la différence entre le cas paramétrique et non-paramétrique est gérer par la définition suivante.

Définition 2.1. *Nous dirons que le modèle défini par 2.1 et 2.2 est un modèle paramétrique lorsque la classe $\mathcal{C}$ est indexable par un nombre fini de paramètres réels. Par opposition nous parlerons de modèle non-paramétrique lorsque $\mathcal{C}$ est un espace de dimension infinie.*

2.1 Régression non-paramétrique réelle

2.1.1 Le modèle

Nous nous plaçons dans le cadre du modèle de régression défini par 2.1 et 2.2. On suppose que les deux variables Y et X sont réelles et que nous disposons pour cette estimation d'un échantillon $\{(X_i, Y_i), i = 1, ..., n\}$ de couples indépendants et ayant chacun même loi que (X, Y).
Ce modèle est un modèle non-paramétrique dans le sens que la seule condition que nous ferons sur la fonction r est une condition de régularité

$$r \ est \ f \ sont \ k \ fois \ continûment \ dérivable, \tag{2.3}$$

k étant un entier positif ou nul, et f désigne la densité de la variable explicative X relativement à la mesure de Lebesgue sur $\mathbb{R}$. Ce problème a été abondamment étudié dans la littérature, et notre objectif ici n'est pas de présenter une discussion exhaustive des différentes méthodes d'estimation existantes. Nous nous limiterons à l'étude des estimateurs de type noyau, cela est à cause de la simplicité de leur construction et de leur facilite, et en particulier de leur bonnes propriétés asymptotiques.

2.1.2 L'expression explicite de l'estimateur

Les estimateurs de type noyau, introduits indépendamment par Nadaraya (1964) et Watson (1964), sont une des techniques les plus populaires d'estimation sous des modèles de régression de type 2.1. Pour comprendre les idées qui ont amené a l'introduction de ces estimateurs, il faut se remonter au régressogramme de Tukey (1961) défini de la manière suivante :

$$\hat{r}_{reg}(x) = \frac{\sum_{i=1}^{n} Y_i \mathbb{1}(X_i \in B_j)}{\sum_{i=1}^{n} \mathbb{1}(X_i \in B_j)}, \forall x \in B_j, \tag{2.4}$$

où, $B_j, j = 1, ..., J$ est une partition du support de X fixée a priori. Comme dans le cas de l'histogramme en estimation de la densité, cet estimateur primitif présente comme inconvénient d'avoir à choisir à la fois la finesse de la discrétisation (i.e. le nombre J de subdivisions) et la position exacte des bornes des intervalles B_j.

Afin de résoudre ce second problème, un nouvel estimateur peut être construit en remplaçant la discrétisation a priori en intervalles B_j par un seul intervalle mais qui varie de manière continue. Concrètement, cela donne l'estimateur de la fenêtre mobile défini par :

$$\hat{r}_{FM}(x) = \frac{\sum_{i=1}^{n} Y_i \mathbb{1}(X_i \in [x-h; x+h])}{\sum_{i=1}^{n} \mathbb{1}(X_i \in [x-h; x+h])}, \forall x, \tag{2.5}$$

où, h est un paramètre réel strictement positif.

L'estimateur précédent présente encore le désavantage d'être discontinu par nature. Ainsi sa généralisation naturelle est l'estimateur à noyau, appelé aussi estimateur de Nadaraya-Watson, défini de la manière suivante :

$$\hat{r}_{NW}(x) = \frac{\sum_{i=1}^{n} Y_i K(\frac{x-X_i}{h})}{\sum_{i=1}^{n} K(\frac{x-X_i}{h})}, \forall x, \tag{2.6}$$

Dans cette définition K est une fonction de $\mathbb{R}$ dans $\mathbb{R}$, et h est un paramètre réel strictement positif (dont nous verrons qu'il sera intéressant de le faire dépendre de n). Cet estimateur sera abondamment étudié dans le reste de ce chapitre. Il faut noter que tous les résultats que nous obtiendrons pour cet estimateur $\hat{r}_{NW}$ resteront valables pour l'estimateur de la fenêtre mobile $\hat{r}_{FM}$ puisque ce dernier est un estimateur à noyau particulier correspondant au cas où K est le noyau uniforme $K(t) = \mathbb{1}(|t| \leq 1)$. D'autres exemples de noyaux classiques sont :

$$\begin{aligned}
K(t) &= (1-|t|)\mathbb{1}(|t| \leq 1) \ (\textit{noyau triangulaire}),\\
K(t) &= \frac{3}{4}(1-t^2)\mathbb{1}(|t| \leq 1) \ (\textit{noyau d'Epanechnikov}),\\
K(t) &= \frac{15}{16}(1-t^2)^2\mathbb{1}(|t| \leq 1) \ (\textit{noyau biweight}),\\
K(t) &= \frac{1}{\sqrt{2\pi}}\exp(-\frac{u^2}{2}) \ (\textit{noyau gaussien}),\\
K(t) &= \frac{1}{2}\exp(-|t|/\sqrt{2})\sin(|t|/\sqrt{2}+\pi/4) \ (\textit{noyau de Silverman}).
\end{aligned}$$

2.1.3 La convergence presque complète de l'estimateur

Nous allons commencer par donner un résultat de convergence presque complète sous le modèle 2.3. Cette notion de convergence presque complète entraîne à la fois la convergence presque sûre et la convergence en probabilité. Pour avoir un résultat de convergence prèsque complète, on suppose les hypothèses suivantes sur le modèle 2.1 :

– Sur la densité f de la variable X :
$$f(x) > 0, \tag{2.7}$$
– Sur le paramètre de lissage $h = h(n)$:
$$lim_{n\to\infty} h = 0 \ et \ lim_{n\to\infty}\frac{nh}{\log n} = \infty, \tag{2.8}$$
– Sur le noyau K :

$$K \; est \; bornée, \; intégrable \; et \; à \; support \; compact, \tag{2.9}$$

- Sur l'ordre du noyau K :

$$\int t^j K(t)dt = 0, \forall j = 1, ..., k-1 \; et \; 0 < |\int t^k K(t)dt| < \infty, \tag{2.10}$$

- Sur la variable aléatoire à expliquer Y :

$$|Y| < M < \infty, \tag{2.11}$$

Le théorème suivant est un résultat de convergence presque complète ponctuelle sous condition de dérivabilité.

Théorème 2.1. *Considérons le modèle 2.3 avec $k > 0$ et supposons que les conditions 2.7, 2.8, 2.9, 2.10 et 2.11 soient réalisées. On a*

$$\hat{r}_{NW}(x) - r(x) = O(h^k) + O(\sqrt{\frac{logn}{nh}}), \quad p.co.$$

Les idées de base pour la démonstration de ce théorème remontent au travaux de Collomb (cf. [16] et [17]). Pour une preuve plus détaillée on peut consulter [55].

Une version uniforme du résultat précédent peut être énoncée, quitte à modifier quelques hypothèse de la façon suivante : En se place sur un compact S de $\mathbb{R}$ et on suppose :

- Sur la version uniforme de la condition 2.3 :

$$r \; et \; f \; sont \; k \; fois \; continûment \; dérivables \; autour \; de \; S, \tag{2.12}$$

- Sur la densité f :

$$\exists \theta > 0, \; tel \; que \; \inf_{x \in S} f(x) > \theta, \tag{2.13}$$

- sur le noyau K :

$$\exists \beta > 0,\ \exists C > 0,\ telque\ |K(x) - K(y)| \leq C|x - y|^{\beta}\ \forall x \in S,\ \forall y \in S, \quad (2.14)$$

On obtient le théorème suivant, sur la vitesse de convergence presque complète uniforme :

Théorème 2.2. *(cf. [16], [17] et [55])*
Considérons le modèle 2.12 avec $k > 0$ et supposons que les conditions 2.8, 2.9, 2.10, 2.11, 2.13 et 2.13 soient réalisées. On a

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O(h^k) + O(\sqrt{\frac{logn}{nh}}), \quad p.co.$$

2.1.4 La convergence en moyenne quadratique

Nous allons maintenant nous intéresser à des résultats asymptotiques en termes de convergence quadratique. A l'exceptions de quelques détails, les hypothèses dont nous aurons besoin sont sensiblement les mêmes que pour les résultats de convergence presque complète, de la manière suivantes.

- Sur le paramètre de lissage :

$$\lim_{n \to \infty} h = 0\ et\ \lim_{n \to \infty} nh = \infty. \quad (2.15)$$

- Sur la loi (X, Y) :

$$\phi(u) = E(Y^2 | X = u)\ est\ continue\ au\ point\ x\ fix\acute{e}. \quad (2.16)$$

- Sur le noyau K :

$$K\ est\ born\acute{e},\ int\acute{e}grable,\ sym\acute{e}trique\ et\ \grave{a}\ suuport\ compact. \quad (2.17)$$

Le théorème suivant établit la convergence en moyenne quadratique sous hypothèses de dérivabilité en un point fixé.

Théorème 2.3. *[16] (Convergence en moyenne quadratique ponctuelle) Considérons le modèle 2.3 avec $k \geq 2$ et supposons que les conditions 2.7, 2.11, 2.15, 2.16 et 2.17 soient réalisées. On a*

$$E[\hat{r}_{NW}(x) - r(x)]^2 = B^2(x)h^4 + V(x)\frac{1}{nh} + o(h^4 + \frac{1}{nh}),$$

où

$$B(x) = \frac{(g^2(x) - r(x)f^2(x))}{f(x)} \cdot \frac{\int t^2 K(t)dt}{2} \text{ avec la notation } g = rf,$$

et

$$V(x) = \frac{(\phi(x) - r^2(x))}{f(x)} \int K^2(t)dt.$$

2.2 Régression non-paramétrique vectorielle

2.2.1 Présentation du modèle

On se place dans cette section dans le cas de l'estimation de la fonction de régression

$$r(x) = E(Y|X = x),$$

d'une variable réelle Y sur une variable X à valeurs dans $\mathbb{R}^p$, $p > 1$. L'échantillon $(X_i, Y_i), i = 1, ..., n$ est constitué de couples indépendants et ayant chacun même loi que (X, Y). On note f la densité marginale de la variable X relativement à la mesure de Lebesgue sur $\mathbb{R}^p$.

2.2.2 Présentation de l'estimateur

La façon naturelle pour généraliser l'estimateur proposé dans la formule 2.6 pour le modèle non-paramétrique univarié, au modèle vectoriel est de poser :

$$\hat{r}_{NW}(x) = \frac{\sum_{i=1}^{n} Y_i K(\frac{x-X_i}{h})}{\sum_{i=1}^{n} K(\frac{x-X_i}{h})}, \forall x \in \mathbb{R}^p. \tag{2.18}$$

Dans cette formule le noyau K est défini de $\mathbb{R}^p$ dans $\mathbb{R}$, $h = h_n$ est un paramètre réel strictement positif. Avant de présenter quelques résultats de convergence pour l'estimateur $\hat{r}_{NW}$, nous définissons l'ordre du noyau de la façon suivante :

Pour tout $p - uplet$ d'entiers positifs $(i_1, ..., i_p)$, on pose

$$T_K(i_1, ..., i_p) = \int_{\mathbb{R}^p} u_1^{i_1} ... u_p^{i_p} K(i_1, ..., i_p) du_1 ... du_p,$$

et pour un entier fixé $k \in \mathbb{N}^*$, on pose pour tout $j \in \{1, ..., p\}$

$$T_K(j) = \int_{\mathbb{R}^p} u_j^k K(i_1, ..., i_p) du_1 ... du_p,$$

L'ordre d'un noyau multidimensionnel est défini par :

Définition 2.2. *On dit qu'une fonction K de $\mathbb{R}^p$ dans $\mathbb{R}$ est un noyau d'ordre k, $k \in \mathbb{N}^*$, lorsque :*

$$\begin{aligned} \forall (i_1, ..., i_p) \in \mathbb{N}^{*p}, \forall j, i_j < k \quad &\Longrightarrow \quad T_K(i_1, ..., i_p) = 0, \\ \forall j \quad &, \quad T_K(j) \neq 0. \end{aligned}$$

2.2.3 La convergence presque complète

On suppose avoir le modèle de régression sous une condition de régularité de type

$$r \ et \ f \ sont \ k \ fois \ continûment \ dérivables \ autour \ de \ x, \tag{2.19}$$

x étant un point fixé de $\mathbb{R}^p$.
Pour obtenir un résultat de convergence complète ponctuèlle pour ce modèle, nous introduisons les hypothèses suivantes :

- Sur la densité marginale f :

$$f(x) > 0, \tag{2.20}$$

- Sur le paramètre de lissage $h = h(n)$:

$$lim_{n\to\infty} h = 0 \ et \ lim_{n\to\infty} \frac{nh^p}{\log n} = \infty, \tag{2.21}$$

- Sur le noyau K :

$$K \ est \ borné, \ intégrable \ et \ à \ support \ compact, \tag{2.22}$$

- Sur l'ordre du noyau K :

$$K \ est \ un \ noyau \ d'ordre \ k, \tag{2.23}$$

- Sur la variable aléatoire à expliquer Y :

$$|Y| < M < \infty, \tag{2.24}$$

On aura donc le résultat suivant :

Théorème 2.4. *(cf. [60], [61] et [55])*
Considérons le modèle 2.19 avec k>0 et supposons que les conditions 2.20, 2.21, 2.22, 2.23 et 2.24 soient réalisées. Alors

$$\hat{r}_{NW}(x) - r(x) = O(h^k) + O(\sqrt{\frac{logn}{nh^p}}), \quad p.co.$$

Un noyau vérifiant les conditions du théorème précédent est donné par :

Exemple 1.

$$K(t_1, ..., t_p) = (\frac{3}{4})^p(1 - t_1^2) \times ... \times (1 - t_p^2),$$

est un noyau p-dimensionnel d'ordre 2.

Un théorème de convergence presque complète uniforme (similaire au cas de la régression réelle) peut être obtenu avec des hypothèses vonvenables sur un compact S de $\mathbb{R}^p$.
On considère le modèle non-paramétrique de type

$$r \; et \; f \; sont \; k \; fois \; continûment \; différentiables \; autour \; de \; S. \tag{2.25}$$

On suppose avoir aussi les deux hypothèses :

$$\exists \theta > 0, \; \inf_{x \in S} f(x) > \theta, \tag{2.26}$$

$$\exists \beta > 0, \; \exists C < \infty, \; |K(x) - K(y)| \leq C\|x - y\|^\beta, \; \forall (x, y) \in S \times S. \tag{2.27}$$

Alors on a le :

Théorème 2.5. *(cf. [60], [61] et [55])*
Considérons le modèle 2.25 avec k>0 et supposons que les conditions 2.21, 2.22, 2.23, 2.24, 2.26 et 2.27 soient réalisées. Alors

$$\sup_{x \in S} |\hat{r}_{NW}(x) - r(x)| = O(h^k) + O(\sqrt{\frac{logn}{nh^p}}), \quad p.co.$$

2.2.4 La convergence en moyenne quadratique

De même que pour le cas de régression réelle en section précédente, on trouve dans la littérature des résultats sur les propriétés asymptotiques de l'estimateur à noyau en termes de convergence quadratique. Et de la même façon, des modifications sur les hypothèses de le convergence presque complète nous permettent d'établir un résultat similaire au cas univarié :

- Sur le paramètre de lissage :

$$\lim_{n\to\infty} h = 0 \textit{ et } \lim_{n\to\infty} nh^p = \infty. \tag{2.28}$$

- Sur la loi (X, Y) :

$$\phi(u) = E(Y^2|X = u) \textit{ est continue au point } x \textit{ fixé.} \tag{2.29}$$

- Sur le noyau K :

$$K \textit{ est borné, intégrable, symétrique et à suuport compact.} \tag{2.30}$$

On trouve donc le théorème suivant qui établit la convergence en moyenne quadratique ponctuelle.

Théorème 2.6. *[25]*
Considérons le modèle 2.19 avec $k \geq 2$ et supposons que les conditions 2.20, 2.24, 2.28, 2.29 et 2.30 soient réalisées. On a

$$E[\hat{r}_{NW}(x) - r(x)]^2 = B^2(x)h^4 + V(x)\frac{1}{nh^p} + o(h^4 + \frac{1}{nh^p}),$$

où

$$B(x) = \frac{1}{2f(x)} \sum_{j=1}^{p} T_K(j) \left(\left[\frac{\partial^k g}{\partial x_j^k} \right](x) - \left[r(x) \frac{\partial^k f}{\partial x_j^k} \right](x) \right) \textit{ avec la notation } g = rf,$$

et

$$V(x) = \frac{(\phi(x) - r^2(x))}{f(x)} \int_{\mathbb{R}^p} K^2(t)dt.$$

Corollaire 2.1. *[25](Vitesse optimale de convergence en moyenne quadratique). Supposons que les conditions du théorème précédent soient vérifiées. On suppose de plus que la fenêtre h soit de la forme :*

$$h = Cn^{-\frac{1}{4+p}}, \ 0 < C < \infty.$$

Alors on a

$$E[\hat{r}_{NW}(x) - r(x)]^2 = O(n^{-\frac{4}{4+p}}).$$

Remarque 2.1. *La vitesse de convergence obtenue au corollaire précédent est optimale mais elle est instable relativement à la dimension du modèle p. Ce problème, connu dans la littérature comme le fléau de la dimension. Un problème lié à la rareté des données dans un espace à grande dimension, cette rareté impose dans la pratique un choix de largeur de fenêtre h d'autant plus grand que p est grand afin de stabiliser la variance de l'estimateur, mais cela conduit à une augmentation du biais, à moins d'avoir à disposition des tailles d'échantillon gigantesques (n assez grand).*

2.3 Régression non-paramétrique fonctionnelle

Cette partie du chapitre est consacrée à l'étude d'un modèle de régression non-paramétrique lorsque la variable explicative est à valeurs dans un espace vectoriel semi-normé, tandis que la variable réponse est réelle. On parle alors

d'un modèle non-paramétrique fonctionnel.
On dit qu'une variable aléatoire X est fonctionnelle (v.a.f) si X est une application mesurable définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ à valeurs dans un espace probabilisable $(E, \mathcal{B}_E)$ où E est un espace de dimension infinie, $\mathcal{B}_E$ étant la tribu borélienne de E. La terminologie *fonctionnelle* viens du fait que la dimension de E est infinie.
L'idée dans l'étude de ces modèle consiste à généraliser les résultats sur les modèles classiques pour variables aléatoires vectorielles en se plaçons dans un cadre fonctionnel, et le cas où l'espace E est un espace vectoriel de dimension finie se ramène naturellement au cadre vectoriel évoqué à la section précédente (cas où $E = \mathbb{R}^p$).

2.3.1 Modèle de régression fonctionnelle

On s'intéresse à un modèle de régression non-paramétrique de Y en X, où Y est une variable aléatoire réelle et X est une variable aléatoire à valeurs dans un espace vectoriel semi-normé $(E,\ ||.||)$, X et Y étant définies sur un même espace probabilisé $(\Omega,\ A,\ P)$. Dans un tel cadre, nous cherchons à estimer 1'opérateur fonctionnel Φ défini par

$$Y = \Phi(X) + \epsilon, \tag{2.31}$$

où ϵ est une variable réelle centrée et indépendante de X. De tels modèles offrent de nouvelles perspectives dans de nombreux domaines d 'application lorsque la variable explicative X est de nature fonctionnelle} on parle alors génériquement d'analyse de données fonctionnelles (cf. [50], [51] et [52]).
Pour les techniques développées ci-dessous, on se réfère à [22]. La démarche utilise essentiellement la notion de la dimension fractale qui repose elle même

sur une hypothèse fondamentale qui s'exprime de la façon suivante :
Supposons que pour un point fixé $x \in E$, on ait :

$$\lim_{\alpha \to 0^+} \frac{P(X \in B(x, \alpha))}{\alpha^{\delta(x)}} = c(x), \tag{2.32}$$

où, $\delta(x)$ et $c(x)$ sont deux réels strictement positifs et où $B(x,\ \alpha)$ désigne la boule de centre x et de rayon α pour la topologie associée à la semi-norme de E.

2.3.2 L'estimateur de la régression fonctionnelle

A partir de n observations indépendentes $(X_i,\ Y_i)_{i=1,\ldots,n}$ du couple $(X,\ Y)$, on s'intéresse à l'estimation de l'opérateur de régression

$$\Phi(x) = E(Y|X = x)$$

. L'estimateur $\hat{\Phi}_n(x)$ de $\Phi(x)$ est défini par :

$$\hat{\Phi}_n(x) = \sum_{i=1}^{i=n} w_i(x) Y_i \tag{2.33}$$

avec

$$w_i(x) = \frac{K(\frac{||X_i - x||}{h})}{\sum_{i=1}^{i=n} K(\frac{||X_i - x||}{h})}, \tag{2.34}$$

Le paramètre h étant une suite de nombres positifs. $\hat{\Phi}_n(x)$ est une version de 1'estimateur de Nadaraya-Watson adaptée au cas où la variable explicative est à valeurs dans un espace semi-normé.

2.3.3 Quelques propriétés asymptotiques

Pour présenter quelques résultats de convergence pour cet estimateur, nous allons donner quelques hypothèses utiles. Notons que les techniques de démonstration employées dans le cadre de v.a.f. s'inspirent largement de celles utilisées dans le cas réel et vectoriel. Pour la convergence presque complète, on note les hypothèses suivantes :

- La largeur de la fenêtre h est telle que :

$$\lim_{n\to\infty} h = 0 \ et \ \lim_{n\to\infty} \frac{nh^{\delta(x)}}{logn} = \infty, \tag{2.35}$$

- le noyau K est tel que :

$$K \ Lipschitzien \ d'ordre \ 1 \ et \ \grave{a} \ support = [0, \ \xi] \ avec \ \xi \in \mathbb{R}_*^+, \tag{2.36}$$

- la variable aléeatoire réelle Y vérifie :

$$|Y| \leq M < \infty \ p.s. \tag{2.37}$$

Le prochain théorème établi la convergence presque complète ponctuelle de l'estimateur $\hat{\Phi}_n(x)$.

Théorème 2.7. *(cf. [22], [25])*
Sous les hypothèses 2.32, 2.35, 2.36 et 2.37 et si

$$\hat{\Phi} \ est \ continu \ en \ x,$$

alors on a :

$$\lim_{n\to\infty} \hat{\Phi}_n(x) = \Phi(x) \ p.co.$$

Pour la convergence presque complète uniforme sur un compact S de l'espace semi-normé $(E, ||.||)$, Il est nécessaire de procéder au renforcement de certaines hypothèses tel qu'il était le cas avec le modèle vectoriel. En particulier l'hypothèse 2.32 est modifiée de la façon suivante :

$$\lim_{\alpha\to 0^+} \sup_{x\in S} \left\{ \frac{P(X \in B(x,\alpha))}{\alpha^\delta} - c(x) \right\} = 0 \ et \ \inf_{x\in S} c(x) > 0, \tag{2.38}$$

où δ est une constante réelle strictement positive ne dépendant pas de x, on peut alors énoncer le résultat suivant.

Théorème 2.8. *(cf. [22], [25])*
Sous les hypothèses 2.35, 2.36, 2.37 et 2.38, et si

$$\Phi \; est \; uniformément \; continu \; sur \; S,$$

alors on a :

$$\lim_{n\to\infty} \sup_{x\in S} |\hat{\Phi}_n(x) - \Phi(x)| = 0 \; p.co.$$

Remarque 2.2. *On a évoqué dans ce chapitre quelques résultats est propriétés asymptotiques pour l'estimateur à noyau de la régression pour des modèles réelles, vectoriels et fonctionnels. Il est important de signaler que cet estimateur possède beaucoup plus de bonnes propriétés asymptotiques que ce qu'on a présenté ici, des résultats similaires peuvent être démontrés pour des hypothèses de régularité différentes (cas continu, de type Lipschitz,...). On trouve dans [1], [2], [3] et [49] des résultats de la convergence en moyenne quadratique pour différents cas de régularités.*
Pour des références clés dans ce sujet on cite aussi [60],[61] , [51], [52], [55] et [25]. Les deux articles [18] et [31] présentent un suvrol assez riche pour les problèmes liées à la statistique des données fonctionnelles ainsi que les résultats les plus récents.

Chapitre 3

Etat de l'art sur le phénomène de Stein

3.1 Introduction

Pour la plupart des lois de probabilité usuelles, l'estimateur du maximum de vraisemblance est défini de façon unique, et se calcule explicitement. Sur le plan théorique, il présente de nombreux avantages. Sous des hypothèses vérifiées par de nombreux modèles courants, on démontre qu'il est asymptotiquement sans biais et convergent. On démontre de plus que sa variance est minimale. La méthode du maximum de vraisemblance est donc théoriquement la meilleure des méthodes d'estimation.

Mais cette méthode d'estimation par maximum de vraisemblance a aussi ses défauts, on cite parmis ses défauts :

1. La difficulté de maximisation de la fonction de vraisemblance.

2. Une technique de maximisation donne forcément des estimateurs peu lisses.
3. L'approche du maximum de vraisemblance n'admet pas toujours de justification probabiliste et décisionelle.

3.2 Admissibilité et Minimaxité

Dans cette première partie, en se référant à Lehmann et Casella [45], et Robert [53], on se trouve devant le cadre de la théorie de décision statistique, quelques outils de base dans le cas d'un problème d'estimation. Soient χ et Θ deux espaces euclidiens quelconques et soit X une variable aléatoire définie sur l'espace d'échantillonnage χ et suivant une distribution P_θ appartenant à une famille $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. Cette loi P_θ de la famille $\mathcal{P}$ est inconnue au sens où le paramètre θ qui indexe P_θ est supposé inconnu ; nous nous référons à X comme étant un échantillon de la loi P_θ. L'objectif du problème d'estimation est la détermination d'une fonction δ définie sur l'espace d'échantillonnage, afin d'estimer une fonction $g(\theta)$ du paramètre θ. Bien entendu, on ne peut prétendre que $\delta(X)$ égale l'inconnu $g(\theta)$, puisqu'il s'agit d'une variable aléatoire. Pour cela, il est nécessaire de spécifier une mesure de l'écart entre $\delta(X)$ et $g(\theta)$. À cette fin, on introduit une fonction de perte (ou de coût) $L(\theta, \delta)$, qui est une fonction réelle positive de θ et de δ et, qu'usuellement, on suppose telle que $L(\theta, g(\theta)) = 0$ pour tout θ. Un exemple de telles fonctions est le coût quadratique $L(\theta, \delta) = (\delta, g(\theta))^2$ lorsque δ et $g(\theta)$ sont réels ou

$$L(\theta, \delta) = \|\delta - g(\theta)\|^2$$

lorsque δ et $g(\theta)$ sont à valeurs dans $\mathbb{R}^p$ muni de la norme usuelle $\|.\|$.

Comme, pour un estimateur $\delta(X)$ de $g(\theta)$ et pour $X = x$ observé, le coût encouru $L(\theta, \delta(x))$ dépend de l'observation x, la fonction de coût ne suffit pas par elle-même à évaluer les performances globales de $\delta(X)$. Pour résoudre ce problème, une approche classique consiste à considérer le coût "moyen" qui est la fonction risque définie par

$$\mathcal{R}(\theta, \delta) = \mathbb{E}_\theta L(\theta, \delta)$$

où E_θ désigne l'espérance par rapport à la loi P_θ. Pour une classe d'estimateurs obtenus, nous pouvons donc les comparer en utilisant les fonctions de risque définies par cette formule.

Définition 3.1. *Pour deux estimateurs δ_1 et δ_2, on dit que δ_1 est au moins aussi bon que δ_2 si :*

$$\mathcal{R}(\theta, \delta_1) \leq \mathcal{R}(\theta, \delta_2)$$

pour tout $\theta \in \Theta$, cette inégalité étant stricte pour au moins une valeur de θ.

Dans le cadre de la définition du risque, un estimateur δ est dit inadmissible s'il est dominé (autrement dit, l'existence d'un estimateur au moins aussi bon que δ) par au moins un estimateur δ'. Inversement il est dit admissible s'il n'existe aucun estimateur qui le domine. Comme l'admissibilité est une propriété importante, il est intéressant de déterminer des estimateurs admissibles.

Définition 3.2. *Une classe d'estimateurs $\mathcal{C}$ est dite complète si, pour tout δ n'appartenant pas à $\mathcal{C}$, il existe δ' appartenant à $\mathcal{C}$ tel que δ' domine δ.*

Cette définition implique que tout estimateur en dehors de la classe complète est inadmissible. Donc il est raisonnable de chercher l'estimateur optimal parmi les estimateurs de la classe complète. Il faut noter qu'il existe plusieurs estimateurs admissibles (tout estimateur dont la fonction de risque se croise avec toutes les autres est admissible) d'une part et d'autre part l'admissibilité est un critère d'optimalité faible. En effet, pour un coût quadratique, l'estimateur constant $\delta(X) = \theta_0$ est typiquement admissible, puisque sa fonction de risque

$$\mathcal{R}(\theta, \delta) = \|\theta_0 - \theta\|^2 = 0$$

pour $\theta = \theta_0$. Sans avoir utilisé l'information fournie par X.
Ce qui signifie que nous ne pouvons pas résoudre le problème du choix de l'estimateur optimal en se basant seulement sur l'admissibilité. Un deuxième critère de comparaison des estimateurs souvent utilisé est le critère de minimaxité, qui consiste à minimiser le risque dans le cas le plus défavorable.

Définition 3.3. *Un estimateur δ^M qui minimise le maximum de la fonction risque, c'est-à-dire tel que :*

$$\inf_{\delta} \sup_{\theta} \mathcal{R}(\theta, \delta) = \sup_{\theta} \mathcal{R}(\theta, \delta^M)$$

est appelé estimateur du minimax.

Dans la suite, nous donnons des relations entre admissibilité et minimaxité.

Théorème 3.1. *Si un estimateur du minimax est unique, alors il est admissible*

Preuve.

Si l'estimateur est inadmissible, alors il est dominé par autre estimateur au

sens du risque. D'où l'existence d'un autre estimateur du minimax.
Pour les estimateurs dont le risque est constant, nous avons le résultat suivant.

Théorème 3.2. *Si un estimateur a un risque constant et s'il est admissible, alors il est aussi un estimateur du minimax.*

Preuve.
Si δ' est un estimateur de risque constant égal à k alors :

$$\sup_{\theta \in \Theta} \mathcal{R}(\theta, \delta') = k$$

Si, de plus, δ' est admissible on a. pour tout autre estimateur δ, l'implication suivante :

$$\forall \theta \in \Theta \; \mathcal{R}(\theta, \delta) \leq k \Longrightarrow \mathcal{R}(\theta, \delta) = k.$$

Il en résulte que :

$$\sup_{\theta \in \Theta} \mathcal{R}(\theta, \delta') \leq k \Longrightarrow \sup_{\theta \in \Theta} \mathcal{R}(\theta, \delta') = k,$$

et ainsi que k est le maximum du risque du modèle et donc que δ' est un estimateur du minimax.
S'il existe un unique estimateur du minimax δ, selon le théorème 3.1, δ est admissible et il sera le meilleur estimateur pour le critère considéré. Maintenant, s'il n'est pas unique, il existera d'autres estimateurs du minimax améliorant δ. En particulier, si l'estimateur du minimax de risque constant est inadmissible, alors tout autre estimateur du minimax a un risque uniformément plus petit.
Notons que si δ est un estimateur du minimax de risque constant égal à

k, alors tout autre estimateur du minimax δ' est tel que $sup_\theta \mathcal{R}(\theta, \delta') = k$, c'est le contexte du phénomène de Stein. Dans ce cas, et en se basant sur la minimaxité et l'admissibilité, nous ne pouvons pas déterminer l'unique et le meilleur estimateur, puisque il peut exister plusieurs estimateurs du minimax et admissibles. D'où la pertinence du problème de caractérisation des classes d'estimateurs vérifiant la minimaxité et l'admissibilité. Pour prouver l'admissibilité ou la minimaxité d'un estimateur δ, l'analyse statistique bayésienne fournit un cadre pertinent.

Définition 3.4. *Supposons que Θ soit muni d'une loi a priori propre π (c'est-à-dire telle que $\int_\Theta d\pi(\theta) = 1$). Un estimateur δ minimisant le risque de Bayes défini par :*

$$r(\pi, \delta) = \int_\Theta \mathcal{R}(\theta, \delta) d\pi(\theta)$$

est appelé estimateur de Bayes par rapport à π.

Cette dernière définition induit une méthode de déduction de l'admissibilité qui est donnée dans le résultat suivant.

Théorème 3.3. *Si l'estimateur de Bayes relativement à une loi a priori propre π est unique $((P_\theta)_\theta \in \Theta)-$presque surement, alors celui-ci est admissible.*

Preuve.

Soit δ un estimateur au moins aussi bon que l'estimateur de Bayes δ_π, c'est-à-dire tel que :

$$\forall\ \theta \in \Theta \ \ \mathcal{R}(\theta, \delta) \leq \mathcal{R}(\theta, \delta_\pi).$$

Alors son risque de Bayes vérifie :

$$r(\theta, \delta) = \int_\Theta \mathcal{R}(\theta, \delta) d\pi(\theta) \leq \int_\Theta \mathcal{R}(\theta, \delta_\pi) d\pi(\theta) = r(\theta, \delta_\pi).$$

Ceci signifie que l'estimateur δ est un estimateur de Bayes. Par hypothèse d'unicité, il en résulte que, $((P_\theta)_\theta \in \Theta)-$presque sûrement, $\delta = \delta_\pi$ et donc que δ_π est un estimateur admissible.

Théorème 3.4. *Lorsque la fonction de risque est continue en θ pour tout estimateur δ, si π est une distribution absolument continue de densité ν positive sur Θ, un estimateur de Bayes δ^ν associé à π avec risque de Bayes fini est admissible.*

Preuve.

Supposons que δ^ν soit inadmissible, alors il existe un estimateur δ tel que $\mathcal{R}(\theta, \delta) < \mathcal{R}(\theta, \delta^\nu)$ pour tout θ avec une inégalité stricte pour au moins une valeur de θ. La continuité de la fonction risque implique qu'il existe un sous-ensemble ouvert Θ_0 de l'espace des paramètres dans lequel $\mathcal{R}(\theta, \delta) < \mathcal{R}(\theta, \delta^\nu)$ pour tout θ dans Θ_0. Par conséquent, nous obtenons :

$$\int \mathcal{R}(\theta; \delta)\nu(\theta)d\theta < \int \mathcal{R}(\theta; \delta^\nu)\nu(\theta)d\theta < \infty$$

puisque ν est une fonction strictement positive, ce qui est en contradiction avec la définition de δ^ν.

Dans l'analyse bayésienne, la loi π et la densité ν utilisées précédemment sont appelées respectivement, la distribution a priori et la densité a priori. Autrement dit, θ est une réalisation d'un vecteur aléatoire θ de distribution a priori π. Un échantillon X est alors décrit à partir de sa loi $P_\theta = P(X|\theta)$, considérée comme une distribution conditionnelle de X sachant $\theta = \theta$. Cependant, l'échantillon observé $X = x$ est utilisé pour obtenir une actualisation de la distribution a priori, appelée distribution a posteriori. Dans ce cas, notons que la distribution conjointe de X et de θ est la mesure de probabilité définie sur $\chi \times \Theta$ et vérifiant :

$$P(A \times B) = \int_B P_{X|\theta}(A) d\pi(\theta), \quad A \in \mathbf{B}_\chi, \quad B \in \mathbf{B}_\Theta$$

où, $\mathbf{B}_\chi$ et $\mathbf{B}_\Theta$ sont respectivement les tribus boréliennes de χ et de Θ. Cependant, nous remarquons que la distribution a posteriori de θ, sachant $X = x$, est exprimée par la distribution conditionnelle $P_{\theta|x}$. En plus, si $P_{X|\theta}$ et π ont des densités de probabilité, par rapport à la mesure de Lebesgue, notées respectivement $f(.|\theta)$ et $\nu(\theta)$, alors la loi marginale M admet une densité marginale donnée par :

$$m(x) = \int_\Theta f(x|\theta)\nu(\theta) d\theta,$$

et la loi a posteriori $P_{\theta|x}$ admet une densité a posteriori (si $m(x) \neq 0$) donnée par :

$$\nu(\theta|x) = \frac{f(x|\theta)\nu(\theta)}{m(x)}.$$

En utilisant le concept de loi a posteriori, la détermination des estimateurs de Bayes est basée sur le résultat suivant.

Théorème 3.5. *En supposant qu'il existe au moins un estimateur à risque fini, et que pour tout x, il existe une valeur $\delta^\pi(x)$ minimisant l'expression*

$$\mathbb{E}\{L(\theta, \delta(x))|X = x\}, \tag{3.1}$$

où l'espérance est calculée par rapport à la distribution a posteriori $P_{\theta|x}$. Alors δ^π est un estimateur de Bayes.

Preuve.

Soit δ un estimateur à risque fini. Alors l'expression

$$\mathbb{E}\{L(\theta, \delta(x))|X = x\}$$

est finie, puisque L est positive. Cependant, nous avons :

$$\mathbb{E}\{L(\theta,\delta(x))|X=x\} \geq \mathbb{E}\{L(\theta,\delta^{\pi}(x))|X=x\},$$

et en calculant l'espérance, par rapport à la loi marginale M, des deux membres de l'inégalité nous obtenons le résultat voulu.
L'expression 3.1 est appelée coût a posteriori. Ainsi l'estimateur de Bayes est interprété comme un estimateur qui minimise le coût a posteriori.

Corollaire 3.1. *En supposant que les conditions du théorème 3.5 soient valides et en supposant également que $L(\theta,d)=\|d-g(\theta)\|^2$, alors l'estimateur de Bayes de $g(\theta)$ est donné par*

$$\delta^{\pi}(x)=\mathbb{E}[g(\theta)|X=x],$$

où l'espérance $\mathbb{E}$ est calculée par rapport à la distribution a posteriori de $\theta|x$.

Preuve.
L'estimateur de Bayes, d'après le théorème 3.5, est obtenu en minimisant l'espérence :

$$\mathbb{E}\{(g(\theta)-\delta(x))^2|X=x\}.$$

Le développement de cette expression en introduisant $\mathbb{E}[g(\theta)|X=x]$ donne :

$$(\delta(x)-\mathbb{E}[g(\theta)|X=x])^2-\mathbb{E}[g(\theta)|X=x]^2+\mathbb{E}[g(\theta)^2|X=x],$$

qui atteint son minimum pour $\delta(x)=\mathbb{E}[g(\theta)|X=x]$.
Maintenant, nous allons étendre la définition de l'estimateur de Bayes au cas où π est une mesure telle que $\int d\pi(\theta)=\infty$, appelée distribution a priori impropre. Dans le cas où l'expression 3.1 est finie pour tout x, l'estimateur de Bayes associé à cette distribution peut être bien défini.

Définition 3.5. *Un estimateur δ^π est dit estimateur de Bayes généralisé par rapport à la mesure $\pi(\theta)$ si l'espérance a posteriori,*

$$\mathbb{E}\{L\theta, \delta(x))|X = x\},$$

atteint son minimum en $\delta(x) = \delta^\pi(x)$ pour tout x.

Une condition suffisante de minimaxité pour quelques estimateurs est donnée par le résultat suivant, en utilisant les propriétés de limite des estimateurs de Bayes.

Lemme 3.1. *Soit π_n une suite de distributions a priori propres et soient δ_n les estimateurs de Bayes associés respectivement à π_n. Si*

$$\lim_{n\to\infty} r(\pi_n, \delta_n) < \infty$$

et s'il existe un estimateur δ tel que

$$\mathcal{R}(\theta, \delta) = \lim_{n\to\infty} r(\pi_n, \delta_n)$$

alors δ est un estimateur du minimax.

Preuve.

Soit δ' un estimateur quelconque. Alors nous avons :

$$\sup_\theta \mathcal{R}(\theta, \delta') \geq \int \mathcal{R}(\theta, \delta') d\pi_n(\theta) \geq r(\pi_n, \delta_n),$$

ceci pour tout n. En passant à la limite, nous en déduisons que :

$$\sup_\theta \mathcal{R}(\theta, \delta') \geq \sup_\theta \mathcal{R}(\theta, \delta),$$

c'est à dire que δ est un estimateur du minimax.

Nous allons aussi définir l'estimateur sans biais du risque, ce qui nous permettra de montrer la minimaxité de certains estimateurs.

Définition 3.6. *Pour un estimateur quelconque $\delta(X)$ dont le risque est $\mathcal{R}(\theta,\delta)$, s'il existe une statistique $\widehat{\mathcal{R}}(\delta(X))$, indépendante de θ, telle que :*

$$\mathcal{R}(\theta,\delta) = \mathbb{E}_\theta \widehat{\mathcal{R}}(\delta(X))$$

Alors cette dernière statistique est appelée un estimateur sans biais du risque.

Lemme 3.2. *Soient $\delta'(X)$ un estimateur du minimax et $\delta(X)$ un estimateur quelconque. En désignant par $\widehat{\mathcal{R}}(\delta'(X))$ et $\widehat{\mathcal{R}}(\delta(X))$, des estimateurs sans biais du risque associés respectivement aux estimateurs $\delta'(X)$ et $\delta(X)$, alors on a :*

1. *$\delta(x)$ est un estimateur du minimax si, pour tout $x \in \chi$,*

$$\widehat{\mathcal{R}}(\delta(X)) \leq \widehat{\mathcal{R}}(\delta'(X)).$$

2. *$\delta'(x)$ est inadmissible si, de plus,*

$$\widehat{\mathcal{R}}(\delta(X)) \leq \widehat{\mathcal{R}}(\delta'(X))$$

sur un ensemble de valeurs de x de probabilité non nulle par rapport à la famille $(P_\theta)_{\theta\in\Theta}$.

Preuve.

Pour la première partie, en calculant l'espérance des deux membres de l'inégalité $\widehat{\mathcal{R}}(\delta(X)) \leq \widehat{\mathcal{R}}(\delta'(X))$, nous obtenons

$$\mathcal{R}(\theta,\delta) \leq \mathcal{R}(\theta,\delta')$$

et δ est un estimateur du minimax, puisque δ' est supposé du minimax. La deuxième partie est une déduction directe de la définition de l'inadmissibilité.

3.3 Phénomène de Stein

Pour définir et clarifier le phénomène de Stein, on propose de se servir de deux modèles statistiques, l'un est paramétrique et l'autre est non-paramétrique. Toutefois il est important à signaler que les démonstrations qu'on trouve dans la littérature se diffèrent l'une à l'autre par des détails techniques. Sauf si le cas contraire est indiqués, les preuves présentées ici sont due à Stein dans [59] (pour le lemme de Stein) et à Lehmann et Casella [45] pour le reste du chapitre.

3.3.1 Modèle et position du problème

On considère les deux modèles suivants :

– Modèle 1. C'est le modèle d'une suite gaussiènne :

$$y_j = \theta_j + \varepsilon \xi_j, \ j = 1, ..., d,$$

avec $0 < \varepsilon < 1$, les ξ_j étant des $v.a.i.i.d.$ de loi $\mathcal{N}(0, 1)$. On note y, θ, ξ les vecteurs

$y = (y_1, ..., y_d)$, $\theta = (\theta_1, ..., \theta_d)$ et $\xi = (\xi_1, ..., \xi_d)$ $\mathcal{N}_d(0, I)$,

où, $\mathcal{N}_d(0, I)$ désigne la loi normale standard en dimension d, le modèle peut donc s'écrire

$$y = \theta + \varepsilon \xi, \ avec \ \xi \in \mathcal{N}_d(0, I)$$

Le problème statistique consiste à estimer le paramètre inconnu $\theta \in \mathbb{R}^d$.

– Modèle 2. On observe les vecteurs aléatoires $X_1, ..., X_n$ vérifiant

$$X_i = \theta + \eta_i, \quad i = 1, ..., n,$$

avec $\theta \in \mathbb{R}^d$, les η_i étant des vecteurs gaussiens i.i.d. de loi $\mathcal{N}_d(0, I)$. Et le Le problème statistique consiste à estimer le paramètre $\theta \in \mathbb{R}^d$. Le vecteur $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est une statistique exhaustive dans ce modèle. On peut écrire

$$\overline{X} = \theta + \varepsilon\xi$$

avec

$$\varepsilon = \frac{1}{\sqrt{n}} \text{ et } \xi = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \eta_i \sim \mathcal{N}_d(0, I).$$

Dans la suite $\mathbb{E}_\theta$ désigne l'espérence relative à la loi de y dans le Modèle 1 ou à celle de $\overline{X}$ dans le modèle 2 et $\|.\|$ désignera la norme euclidienne de $\mathbb{R}^d$.

Remarque 3.1. *Alors que le modèle 2 est classique en statistique paramétrique, le modèle 2 sera utile dans le contexte d'applications à l'estimation non-paramétrique. Compte tenu de l'équivalence des deux modèles, il suffit d'étudier le modèle 1, le calcul pour le modèle 2 étant analogue.*

3.3.2 Admissibilité

Rappelons la définition présentée de l'admissibilité (et la dominance) dans la section précédente :

Définition 3.7. *Un estimateur θ^* du paramètre θ est dit **inadmissible** sur $\Theta \subseteq \mathbb{R}^d$ par rapport au risque quadratique s'il existe un autre estimateur $\hat{\theta}$ tel que*

$$\mathbb{E}_\theta \left\|\hat{\theta} - \theta\right\|^2 \leq \mathbb{E}_\theta \left\|\theta^* - \theta\right\|^2 \qquad \textit{pour tout } \theta \in \Theta$$

et il existe $\theta_0 \in \Theta$ tel que

$$\mathbb{E}_{\theta_0} \left\| \hat{\theta} - \theta_0 \right\|^2 \leq \mathbb{E}_{\theta_0} \left\| \theta^* - \theta_0 \right\|^2$$

*Dans le cas contraire, l'estimateur θ^* est dit **admissible**, et on dit alors que l'estimateur θ^* domine l'estimateur $\hat{\theta}$.*

Exemple 2. *Le risque quadratique de l'estimateur $\overline{X}$ dans le modèle 2 vaut*

$$\mathbb{E}_\theta \left\| \overline{X} - \theta \right\|^2 = \frac{d}{n} = d\varepsilon^2 \ , \qquad \forall \theta \in \mathbb{R}^d.$$

*En considérant le modèle 2, **Stein** a montré que si $d \geq 3$ l'estimateur $\overline{X}$ est **inadmissible**. Cette propriété se nomme **phénomène de Stein**. En outre, Stein a proposé un estimateur qui est meilleur que $\overline{X}$ partout sur $\mathbb{R}^d$ si $d \geq 3$. La construction est basée sur une **contraction** qui rapproche les valeurs de $\overline{X}$ vers 0 en fonction de $\left\| \overline{X} \right\|$.*

3.3.3 La contraction de Stein et l'estimateur de James-Stein

Pour expliquer l'idée de la contraction de stein pour le modèle 1 (le raisonnement pour le modèle 2 étant analogue), on a besoin des lemmes suivants :

Lemme 3.3. *(Stein, 1981) Supposons qu'une fonction*

$$f \ : \mathbb{R}^d \longrightarrow \mathbb{R} \ \ vérifie$$

1.

$$f(u_1, ..., u_d) \text{ est absolument continue en chaque coordonne } u_i, \ \ p.s,$$

2.

$$\mathbb{E}_\theta[\frac{\partial f}{\partial y_i}(y)] < \infty \ , \qquad i = 1, ..., d.$$

Alors

$$\mathbb{E}_\theta[(\theta_i - y_i)f(y)] = -\varepsilon^2 \mathbb{E}_\theta[\frac{\partial f}{\partial y_i}(y)] \ , \qquad i = 1, ..., d.$$

Preuve.[59]

La preuve utilise essentiellement l'intégration par parties avec une légère modification due au fait que la fonction f n'est pas différentiable au sens usuel.

Lemme 3.4. *Pour* $d \geq 3$, *pour tout*$\theta \in \mathbb{R}^d$,

$$0 < \mathbb{E}_\theta(\frac{1}{\|y\|^2}) < \infty \ .$$

Preuve.

La preuve repose sur la propriété de symétrie sphyrique de la loi gaussienne, dans le sens :

$$\begin{aligned} & Pour \ \ \xi \quad \sim \quad \mathcal{N}_d(\mu, \sigma^2) \ \ on \ a, \\ \forall v, \ v' \ \in \mathbb{R}^d \ : \ \|v\| = \|v'\| \ \ & \Longrightarrow \ \ \|\xi + v\| \overset{\text{loi}}{=} \|\xi + v'\| \end{aligned}$$

et pour $d \geq 3$, il existe une constante $C > 0$ telle que

$$\int_{\mathbb{R}^d} \exp(-\frac{\|u\|^2}{6}) \, \|u\|^{-2} \, du = C \int_0^\infty e^{-r^2/6} r^{d-3} dr < \infty.$$

3.3.4 La contraction de Stein

Stein a introduit la classe des estimateurs de la forme $\hat{\theta} = g(y)y$, où $g\ : \mathbb{R}^d \longrightarrow \mathbb{R}$ est une fonction à choisir. Les coordonnées du vecteur $\hat{\theta}$ sont de la forme

$$\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_d),\ \text{avec}\ \hat{\theta}_j = g(y)y_j\ .$$

Le vecteur aléatoire y est lui même un estimateur de θ (analogue à $\overline{X}$ dans le modèle 2). Son risque vaut

$$\mathbb{E}_\theta \left\| y - \theta \right\|^2 = d\varepsilon^2$$

Cherchons une fonction g telle que le risque de $\hat{\theta} = g(y)y$ soit inférieur à celui de y.

$$\begin{aligned}
\mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 &= \sum_{i=1}^{d} \mathbb{E}_\theta[(g(y)y_i - \theta_i)^2] \\
&= \sum_{i=1}^{d} \mathbb{E}_\theta[(y_i - \theta_i)^2] + 2\mathbb{E}_\theta[(\theta_i - y_i)(1 - g(y))y_i] + \mathbb{E}_\theta[y_i^2(1 - g(y))^2].
\end{aligned}$$

Supposons que la fonction g est telle que les conditions du lemme 3.3 soient vérifiées pour $f_i(y) = (1 - g(y))y_i,\ i = 1, ..., d$. Alors,

$$\mathbb{E}_\theta[(\theta_i - y_i)(1 - g(y))y_i] = -\varepsilon^2 \mathbb{E}_\theta[1 - g(y) - y_i \frac{\partial g}{\partial y_i}(y)]\ ,$$

et

$$\mathbb{E}_\theta[(\hat{\theta}_i - \theta_i)^2] = \varepsilon^2 - 2\varepsilon^2 \mathbb{E}_\theta[1 - g(y) - y_i \frac{\partial g}{\partial y_i}(y)] + \mathbb{E}_\theta[y_i^2(1 - g(y))^2]$$

En prenant la somme sur i, on obtient

$$\mathbb{E}_\theta \left\| \hat{\theta} - \theta \right\|^2 = d\varepsilon^2 + \mathbb{E}_\theta[W(y)]$$

avec

$$W(y) \;=\; -2d\varepsilon^2(1-g(y)) + 2\varepsilon^2 \sum_{i=1}^{d} y_i \frac{\partial g}{\partial y_i}(y) + \|y\|^2(1-g(y))^2.$$

Pour que le risque de $\hat{\theta}$ soit plus petit que celui de y, il suffit de choisir g telle que

$$\mathbb{E}_\theta[W(y)] < 0$$

Afin de satisfaire à cette inégalité, Stein (1956) a proposé de chercher g parmi les fonctions de la forme

$$g(y) = 1 - \frac{c}{\|y\|^2}$$

avec une constante $c > 0$ convenablement choisie. Dans ce cas les fonctions $f_i(y) = (1-g(y))y_i$ vérifient les conditions du lemme 3.3, on obtient donc

$$W(y) = \frac{1}{\|y\|^2}(-2d\varepsilon^2 c + 4\varepsilon^2 c + c^2)$$

La valeur c fournissant le minimum de $W(y)$ vaut

$$c_{opt} = \varepsilon^2(d-2)$$

Dans ce cas $W(y) = -\frac{\varepsilon^4(d-2)^2}{\|y\|^2}$, et $-\infty < \mathbb{E}_\theta[W(y)] < 0$ dès que $d \geq 3$, d'après le lemme 3.4. Et La fonction g et l'estimateur $\hat{\theta} = g(y)y$ associés à ce choix de c valent

$$g(y) = 1 - \frac{\varepsilon^2(d-2)}{\|y\|^2} \quad \text{et } \hat{\theta}_{JS} = (1 - \frac{\varepsilon^2(d-2)}{\|y\|^2})y. \tag{3.2}$$

On appelle $\hat{\theta}_{JS}$ estimateur de *James-Stein* (James et Stein 1961). Si la norme $\|y\|$ est assez grande, la fonction g effectue une contraction de y vers 0, que l'on appelle *contraction de Stein* (Stein's shrinkage).

De façon similaire, pour le modèle 2 l'estimateur de James-Stein s'obtient si l'on remplace y par $\overline{X}$ et ε par $1l\sqrt{n}$ dans (1) :

$$\hat{\theta}_{JS} = (1 - \frac{(d-2)}{n\left\|\overline{X}\right\|^2})\overline{X}.$$

On a donc le résultat suivant.

Théorème 3.6. *(Phénomène de Stein).Soit $d \geq 3$, L'estimateur $\hat{\theta} = y$ est inadmissible sur $\mathbb{R}^d$ dans le modèle 1 et l'estimateur $\hat{\theta} = \overline{X}$ est inadmissible sur $\mathbb{R}^d$ dans le modèle 2.*

3.3.5 Autres estimateurs à contraction

Le calcul précédent montre qu'il existe toute une famille d'estimateurs qui sont meilleurs que y dans le modèle 1 : il suffit de prendre $c > 0$ dans la définition de g, de sorte que $-2d\varepsilon^2 c + 4\varepsilon^2 c + c^2 < 0$. Par exemple, si $c = d\varepsilon^2$, on obtient *l'estimateur de Stein :*

$$\hat{\theta}_S = (1 - \frac{d\varepsilon^2}{\|y\|^2})y$$

D'autres estimateurs encore plus performants correspondent aux fonctions g positives :

$$g(y) = \left(1 - \frac{c}{\|y\|^2}\right)_+$$

On obtient pour $c = (d-2)\varepsilon^2$ et $c = d\varepsilon^2$ respectivement les estimateurs de *James-Stein et de Stein à poids positifs :*

$$\hat{\theta}_{JS^+} = \left(1 - \frac{(d-2)\varepsilon^2}{\|y\|^2}\right)_+ y$$

et

$$\hat{\theta}_{S^+} = \left(1 - \frac{d\varepsilon^2}{\|y\|^2}\right)_+ y$$

Un des outils de comparaison entre ces estimateurs est la notion de **Super-efficacité :**

3.3.6 Superefficacité

Définition 3.8. *On dit qu'un estimateur θ^*_ε est superefficace (dans le modèle 1) si*

$$\lim_{\varepsilon \to 0} \sup \frac{\mathbb{E}_\theta \|\theta^*_\varepsilon - \theta\|^2}{d\varepsilon^2} \leq 1 \ , \qquad \forall \theta \in \mathbb{R}^d \tag{3.3}$$

et s'il existe $\theta = \overline{\theta} \in \mathbb{R}^d$ tel que l'inégalité est stricte. Les points $\overline{\theta}$ vérifiant l'inégalité stricte sont appelés points de superefficacité de θ^*_ε .

Un exemple de superefficacité

Proposition 3.1. *Dans le modèle 1, les estimateurs $\hat{\theta}_{JS}$ et $\hat{\theta}_{JS^+}$ sont superefficaces si $d \geq 3$, les estimateurs $\hat{\theta}_S$ et $\hat{\theta}_{S^+}$ sont superefficaces si $d \geq 5$.*

Preuve.[45]

Notons que L. Lecam [46] a montré que, pour tout d fini (i.e. dans le cas paramétrique), l'ensemble des points de superefficacité d'un estimateur est un ensemble de mesure nulle. La situation est totalement différente dans le cas non-paramétrique (par exemple un modèle de suite gaussienne où $d = \infty$), il existe des estimateurs qui sont superefficaces sur un ensemble de mesure non-nulle.

3.4 Commentaires et compléments bibliographiques

3.4.1 Comparaison des estimateurs, efficacité et performance

Dans les quartes dernières décennies, de nombreux chercheurs s'intéressent à la comparaison entre les estimateurs de Stein , James-Stein et les autres estimateurs de rétrécissement. On cite à titre d'exemple les travaux de Lehmann, Lehmann et Casilla, Casella et Hwang (cf. [12], [13]).
Dans [13] G. Casella et J. T. Huang présentent une expression pour la limite (quand la dimension d tend vers l'infini) du risque de l'estimateur de James-Stein divisé par le risque de l'estimateur du maximum de vraisemblance. Une expression semblable est également obtenue quand l'estimateur de James Stein est remplacé par sa version tronquée. Dans les deux cas, la technique utilisée consiste à prendre le quotient en sandwich (ratio) entre un majorant et un minorant qui convergent vers une même limite lorsque d tend vers l'infini. Les bornes ont l'avantage d'être faciles à calculer et de rester précises mime pour des valeurs de d relativement petites.

3.4.2 A propos de la fonction du coût

A la fin de son article historique (Inadmissibility of the usual estimator for the mean of a multivariate normal distribution) Charles Stein a signalée des remarques importantes qui ont donner naissance à une nouvelle vision de la théorie de l'estimation, notamment une remarque sur la fonction de perte, je cite : "For certain loss functions, for example $L(\xi, d) = \sup_i |\xi_i - d_i|$, little or no improvement over the usual estimator may be possible".

Théoriquement, la situation est donc la suivante :

Relativement à une fonction de perte donnée, existe-il une familles d'estimateurs de type Stein qui sont meilleurs qu'un estimateur usuel (dans le ces paramétrique ou non-paramétrique) ?

On cite ici les travaux de T. Kubokawa pour les fonctions de coût concaves, les travaux de Strawderman et Brandwein dans le cas pramétrique avec des fonctions de coût équilibrées. Ces mêmes auteurs traitent aussi le problème d'admissibilité et dominance pour un modèle à symétrie sphérique (cf. [6], [7], [8] et [44]).

Dans [11], L. Brown définit une classe spéciale de fonctions de perte (ρ-loss) et utilise une nouvelle idée d'admissibilité faible pour prouver que pour une dimension $d \geq 3$, l'estimateur usuel est "inadmissible" si les fonctions de perte ρ distinctes peuvent être regroupées en une seule fonction de perte pour le problème ensemble d'estimation de la moyenne.

L' auteur donne une formulation du problème dans laquelle les pertes ne peuvent pas être regroupées de cette manière et trouve une condition étonnamment faible, nécessaire et suffisante, dans laquelle l'estimateur usuel est inadmissible dans un sens approprié à cette formulation. L'auteurs examine également le cas de symétrie sphérique et trouve une autre condition nécessaire et suffisante.

Pour le problème canonique de l'estimation d'une moyenne normale multivariée sous perte d'erreur au carré, E. George [29] aborde le problème de

la sélection d'un estimateur de rétrécissement minimax lorsque des informations antérieures vagues ou contradictoires suggèrent que plus d'un estimateur d'une large classe pourrait être efficace.
Pour cette situation, une nouvelle classe d'estimateurs alternatifs, appelés estimateurs à rétrécissement multiple, est proposée. Ces estimateurs utilisent les données pour imiter les propriétés de comportement et de risque de l'estimateur le plus efficace considéré. Des estimations non biaisées du risque et des conditions suffisantes pour la minimaxité sont fournies.
Les motivations bayésiennes lient cette construction aux moyens a posteriori de mélange des antérieurs. Pour illustrer cette théorie, des estimateurs de Stein à rétrécissement multiple et minimax multiples sont construits, estimateurs qui peuvent rétrécir de manière adaptative les données vers un nombre quelconque de points ou de sous-espaces.

En 1996, A. Parsian et al. [48] étudient le problème de l'dmissibilité et de la dominance pour les cas de non-normalité avec une perte non quadratique. L'estimation du paramètre d'échelle sous l'effet de la perte d'entropie est considérée avec des restrictions aux principes d'invariance et de non biais de risque. Une forme explicite d'estimateur équivariant à l'échelle du risque minimum avec la perte d'entropie est obtenue.
En particulier, les auteurs dans [48] traitent l'admissibilité et l'inadmissibilité d'une classe d'estimateurs linéaires de la forme $cT + d$, où $T \sim \Gamma$ (la distribution gamma), cet étude comprend également l'admissibilité de l'estimateur de moindres carrées pour les paramètres d'intérèt.

Notez que la perte d'entropie a été introduite pour la première fois par Stein dans [58] pour l'estimation de la matrice multi-normale de variance-covariance. Plus tard, la même perte a été envisagée par Brown [10], Haft (cf. [32], [33], [34] et [35]) et dans Dey et al. [19] pour estimer la matrice multi-normale de variance-covariance ou son inverse.

Par suite, Dey et al. [20] ont aussi considérés la perte d'entropie pour l'estimation simultanée de p paramètres d'échelle $\Gamma-$indépendants ou de leurs inverses.

Ighodaro et Santner [38] et Ighodaro et al. [39] ont considérés la perte d'entropie pour l'estimation simultanée de proportions binomiales et multinomiales indépendantes.On cite aussi à se sujet les traveaux de Ghosh et Yang [30] ou ils ont considéré la perte d'entropie pour l'estimation simultanée de $p-$moyennes de Poisson indépendantes.

Dans le même contexte, Rukhin et Ananda [54] ont examiné le problème de l'estimation de la variance inconnue d'un vecteur normal multivarié sous perte quadratique et perte d'entropie. Bien plus tard, Yang [64] a considéré la perte d'entropie pour l'estimation pseudo-orthgonale de moyennes de Poisson pour le cas indépendant.

La perte de Stein est aussi utilisée dans les traveaux de Wieczorkowski et Zielinski [63] ou ils considèrent cette perte pour l'estimation minimax de la probabilité binomiale.

Dans [21], Dey et al. Utilisent la fonction de perte équilibrée de Zellener et montrent en général que les résultats fréquentistes et bayésiens pour une perte équilibrée découlent et impliquent également des résultats connexes pour les fonctions de perte quadratiques ne reflétant que la précision de l'estimation.

Plusieurs exemples sont donnés pour les modèles avec condition de normalité, et plus généralement pour la distribution d'erreur à symétrie sphérique.

Parmis les résultats obtenus dans ce travail, on cite :

Théorème 3.7. *[21]*
Soit le modèle de régression paramétrique $p-$*dimentionnel,*

$$y = \theta + u,$$

avec : $E(u) = 0$ *et* $Cov(u) = \sigma^2 I$, *et si on suppose* $\hat{\theta}$ *l'estimateur usuel de moindres carrés et on définit la fonction de perte équilibrée de Zellner, par*

$$L_\omega(\theta, \tilde{\theta}) = \omega(y - \tilde{\theta})'(y - \tilde{\theta}) + (1 - \omega)(\theta - \tilde{\theta})'(\theta - \tilde{\theta}),$$

Alors
L'estimateur $\hat{\theta} + (1 - \omega)g(y)$ *domine* $\hat{\theta}$ *sous la fonction de perte équilibrée* L_ω *pour tout* $0 \leq \omega < 1$ *si et seulement si* $\hat{\theta} + g(y)$ *domine* $\hat{\theta}$ *sous la perte quadratique* L_0.

Par conséquent, la classe d'estimateurs qui domine l'estimateur des moindres carrés sous la perte usuelle conduit à une classe d'estimateurs qui domine $\hat{\theta}$ sous la fonction de perte équilibrée L_ω (et inversement).
Des résultats similaires à ce théorème sont établis par Jozani et al. dans [42] avec la perte équilibrée avec poids. Dans ce dernier travail, les auteurs traitent l'admissibilité, la dominance et la minimaxité pour des modèles bayésiens et non-bayésiens, avec plus d'exemples en cas de non-normalité (loi de Poisson, Binomial, Négative Binomial, Gamma...).
A partir de ce dernier travail, on s'est inspiré pour poser les questions suivantes :

1. **Peut-il y avoir une relation similaire dans le contexte non-paramétrique entre l'admissibilité et la dominance sous la perte quadratique et celle de Zellner (généralisée en suite par Jozani et al. [42]), en choisissant l'estimateur à noyau comme estimateur cible (target estimator) ?**
2. **La possibilité d'améliorer l'estimateur à noyau par rétrécissement ?**

Il est clair que cette question touche au phénomène de Stein et à sa rigidité relativement au problème de la dimension.

Le chapitre suivant traite les questions précédentes sur un modèle de régression fonctionnel. Il faut signaler que pour des difficultés techniques du calcul, on se restreint actuellement au cas des hypothèses d'indépendance et de normalité du modèle choisit.

Chapitre 4

On Improved Shrinkage Estimators of the Regression Operator for Functional Data under Balanced Loss Function

Ce chapitre a fait l'objet d'une publication dans le Journal "C. R. Acad. Sci. Paris, 357, Issues 1112, 912-916"

4.1 Introduction

In this chapter, we are interested in the regression model

$$Y = r(X) + \varepsilon \tag{4.1}$$

where the response Y and the error ε are a real valued variables, and the explanatory data X is of fixed-design and functional type.
The nonparametric statistical methods have been developed intensively whereas a lot of scientists today collect samples of curves and other functional observations. For an introduction and applications of this field, the books by Ramsay and silverman (cf. [51], [52]), Bosq [5], Ferraty and Vieu [25] provide some methods of analysis along with diverse case studies in several areas including criminology, economics, archeology and neurophysiology.

Note that, in the parametric case, Stein in ([57]) proved the inadmissibility of the best invariant estimator $Y = (Y_1, ...; Y_p)$ of a normal mean $\theta = (\theta_1, ...; \theta_p)$ under quadratic loss, and showed that the estimators of the form :

$$\left(1 - \frac{b}{a + \|Y\|^2}\right) Y$$

dominated Y for a (respectively b) sufficiently small (respectively large) when $p \geq 3$. A few years later, James and Stein in [40] give a new proof of the result of [57] where other distributions and other loss functions are considered. Since then, much research has been devoted to improving upon the best invariant estimator of a location vector θ by realxing the normality assumption, using different loss functions, or considering more general estimators. To quote only a few of them, Bradwein et al. ([6], [7]) considered a generalizations of James-Stein estimators under spherical symmetry distributions and in another work, they present the shrinkage estimators of the location parameter for certain spherically symmetric distributions. A same approach is given by Cellier et al. ([14], [15]) under elliptically symmetric distributions. Fourdrinier et al. ([26]) focused on the estimation of a loss function for spherically symmetric distribution in the general linear model and present in ([28]) an expository development of loss estimation with substantial emphasis on the setting where the distributional context is normal and its extension to the case where the underlying distribution is spherically symmetric.

Consequently, we are interested in the Stein improvement of the Kernel estimator $\hat{r}(x)$ in the model (4.1) using balanced loss function

$$L(r(x), \hat{r}_g(x)) = \omega\phi(\hat{r}(x), \hat{r}_g(x)) + (1-\omega)\phi(r(x), \hat{r}_g(x)) \qquad (4.2)$$

where $0 < \omega < 1$, $\phi(s,t)$ is a differentiable function from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$, $\hat{r}_g$ is a estimator defined by $\hat{r}_g(x) = \hat{r}(x) - g(x, Y)$, where g is some positive operator from $E \times \mathbb{R}^n$ to $\mathbb{R}$ where (E, d) is a semi-metric space.

Contrary to what has been shown by Dey et al. in ([21]), in some cases the question about admissibility and dominance, may depend mainly on ω on one hand. In other hand, we note that results for balanced loss function (BLF) may be inferred directly from corresponding results for quadratic and weighted quadratic loss as shown by Jozani et al. in ([42]) .

This work is organized as follows : some basic assumptions are provided in section 2, in particular the fundamental theorem on the kernel estimator of model (4.1). The loss functions generalizes the Zellner's balanced loss function (proposed in [65]) is defined in section 3. An important proposition is proved in this section, this last is a tool for comparing the difference in risk relative to the quadratic loss function with the difference in risk under the balanced loss function. The main results with proofs are presented in section 4, including admissibility and dominance under proposed balanced loss functions.

4.2 Model

We consider the regression model 4.1 when the response Y is a real valued random variable, with $\mathbb{E}\{|Y|\} < \infty$, and the explanatory data X belongs to a semi-metric space (E, d) equipped to be normally distributed with zero mean and variance σ_ε^2. From the sample $(X_1, Y_1), ..., (X_n, Y_n)$ we estimate the regression operator r by :

$$\hat{r}(x) = \sum_{i=1}^{n} Y_i W_i(x) = (Y_1, ..., Y_n)^t (W_1(x), ..., W_n(x)) := Y^t W \tag{4.3}$$

$\Delta_i(x) = K\left(\dfrac{d(x, X_i)}{h}\right)$ and $W_i(x) = \dfrac{\Delta_i(x)}{\sum_{i=1}^n \Delta_i(x)}$ with K is a real valued function defined on $\mathbb{R}^+$ and $h = h(n)$ is the bandwidth, such that : $h \in \mathbb{R}^+$ and $\lim\limits_{n\to\infty} h = 0$. For all $x \in E$, we denote by :

$$n_x(h) = \#\{X_i : X_i \in \mathbb{B}(x, h)\} = \textstyle\sum_{i=1}^n \mathbf{1}_{\mathbb{B}(x,h)}(X_i)$$

the number of curves belonging to the closed ball $\mathbb{B}(x, h)$ of center x and radius h, where $\#(A)$ denotes the cardinality of the set A.
The empirical probability $\varphi_x(h)$ is defined as :

$$\varphi_x(h) = \frac{n_x(h)}{n}$$

To establish the convergence of the estimator $\hat{r}(x)$, the following assumptions will be needed in the sequel of this work.

– About the kernel : we assume that the kernel K is strictly decreasing on $[0,1]$ and there exist some positive numbers a and b such that :

$$a\mathbf{1}_{[0,1]}(t) \leq K(t) \leq b\mathbf{1}_{[0,1]}(t) \quad t \in \mathbb{R} \tag{4.4}$$

– About the moments : the response variable Y satisfies

$$\exists \nu \geq 2, \exists C_\nu > 0 \quad \mathbb{E}\{|Y|^\nu\} \leq C_\nu < \infty \tag{4.5}$$

– About the regression operator : there exists $0 < k < \infty$ and $\beta > 0$:

$$\forall x, y \in E, |r(x) - r(y)| \leq kd(x,y)^\beta \tag{4.6}$$

$$\forall x, y \in E, \exists C_\nu \in \mathbb{R} : r(x) = r(y) + C_\nu d(x,y)^\beta + \circ(d(x,y)) \tag{4.7}$$

Noting by,

$$M_\nu(h) = K^\nu(1) - \int_0^1 \frac{\varphi_x(hs)}{\varphi_x(h)(K^\nu)'}(s)ds$$

for $\nu = 1, 2$, the behaviour of the estimator $\hat{r}(x)$ is then given by the following theorem.

Theorem 4.1. *([25]) Under the assumptions (4.4), (4.5) and (4.6) or (4.7), if $n\varphi_x(h) \to 0$ as $n \to \infty$, and $M_\nu(h) \to M_\nu$ as $h \to 0$*
Then the risk of $\hat{r}$ is given by :

$$\mathcal{R}(r(x), \hat{r}(x)) = B_n^2(x) + \frac{M_2}{M_1^2}\frac{\sigma_\epsilon}{n\varphi_x(h)}$$

Where

$$B_n(x) = \begin{cases} \mathcal{O}(h^\beta) \ \textit{under assumption } (4.6) \\ \\ C_x D_{x,h} h + \mathrm{o}(h) \ \textit{under assumption } (4.7) \end{cases}$$

is the bias of $\hat{r}(x)$, *with*

$$D_{x,h} = \sum_{\substack{i=1 \\ X_i \in B(x,h)}}^{n} W_i(x) \frac{d(x, X_i)}{h} < 1$$

From Theorem 4.1 and through the variance term of the estimator, it appears that the problem of design choice is much crucial here than of the ordinary finite dimensional regression.

4.3 Balanced loss functions (BLF)

let $x \in E$. We wish to estimate r under the balanced loss

$$L(r(x), \hat{r}_g(x)) = \omega\phi(\hat{r}(x), \hat{r}_g(x)) + (1-\omega)\phi(r(x), \hat{r}_g(x))$$

Where,

$$\hat{r}_g(x) = \hat{r}(x) - g(x, Y)$$

$\hat{r}(x)$ is the kernel estimator, g is some positive operator from $E \times \mathbb{R}^n$ to $\mathbb{R}$, and $0 < \omega < 1$. The risk of any estimator $\hat{r}_g$ will be denoted by

$$\mathcal{R}(r(x), \hat{r}_g(x)) = \mathbb{E}\{L(r(x), \hat{r}_g(x))\}$$

Where $\mathbb{E}$ holds for the expectation with respect to the sampling distribution of ε. We have to determine g in order to reduce the balanced risk $\mathcal{R}(r(x), \hat{r}_g(x))$.

Our first purpose is to establish a sufficient condition on g which allows us to improve the estimator $\hat{r}(x)$. In what follows, we study the alternative estimator $\hat{r}_g(x)$. Along this article we consider distance-based balanced loss functions, this means that we have :

$$\phi(s,t) = \psi(s-t),$$

Where the representing function ψ is a positive function such that $\psi(0) = 0$. We are now able to give an expression for the difference of the risk :

$$\begin{aligned}
\triangle\mathcal{R}(x) &= \mathcal{R}(r(x), \hat{r}_g(x)) - \mathcal{R}(r(x), \hat{r}(x)) \\
&= \mathbb{E}\{\ L(r(x), \hat{r}_g(x)) - L(r(x), \hat{r}(x))\} \\
&= \omega\mathbb{E}\{\phi(\hat{r}(x), \hat{r}_g(x)) - \phi(\hat{r}(x), \hat{r}(x))\} \\
&+ (1-\omega)\mathbb{E}\{\phi(r(x), \hat{r}_g(x)) - \phi(r(x), \hat{r}(x))\} \\
&= \omega\mathbb{E}\{\psi(g(x,Y))\} + (1-\omega)\mathbb{E}\{\psi(r(x) - \hat{r}(x) + g(x,Y)) - \psi(r(x) - \hat{r}(x))\}
\end{aligned}$$

First note that, in order that the estimation problem makes sense, we need the risk $\mathcal{R}(r(x), \hat{r}_g(x))$ of $\hat{r}_g(x)$ to be finite, which means :

$$Bias^2(\hat{r}_g(x)) + var(\hat{r}_g(x)) < \infty$$

Under that last condition, it is also easy to check that the risk $\mathcal{R}(r(x), \hat{r}_g(x))$ is finite if

$$\mathbb{E}\{\psi(g(x,Y))\} < \infty.$$

The method used later is to compare the difference in risk $\triangle\mathcal{R}(x)$ with that obtained by considering the quadratic loss function $L_0(r(x), \hat{r}_g(x)) = (r(x) - \hat{r}_g(x))^2$. If we note $\mathcal{R}_0(x)$ the risk obtained relatively to the quadratic loss function and $\triangle\mathcal{R}_0(x)$ the difference in risk, we have :

$$\begin{aligned}\triangle\mathcal{R}_0(x) &= \mathcal{R}_0(r(x),\hat{r}_g(x)) - \mathcal{R}_0(r(x),\hat{r}(x)) \\ &= \mathbb{E}\{(\hat{r}_g(x)-r(x))^2 - (\hat{r}(x)-r(x))^2\} \\ &= \mathbb{E}\{(\hat{r}(x)-g(x,Y)-r(x))^2 - (\hat{r}(x)-r(x))^2\} \\ &= \mathbb{E}\{(g^2(x,Y) - 2(\hat{r}(x)-r(x))g(x.Y)\}.\end{aligned}$$

the following proposition gives us the formula of the difference in risk $\triangle\mathcal{R}_0(x)$. Notice first that $\partial g(x,Y)/\partial Y_i$ is the partial derivative of $g(x,Y)$ with respect to the ith variable, when Y is considered as the vector $(Y_1,...,Y_n)$.

Proposition 4.1. *For every diffentiable function $g : E\times\mathbb{R}^n \longrightarrow \mathbb{R}$, we have*

$$\triangle\mathcal{R}_0(x) = \mathbb{E}\{g^2(x,Y)\} - 2\sigma_\epsilon\sum_{i=1}^{n} W_i(x)\mathbb{E}\{\frac{\partial(g(x,Y))}{\partial Y_i}\}$$

Proof. Let $Y^{i,t} = (Y_1,...,Y_{i-1},t,Y_{i+1},...,Y_n)\in\mathbb{R}^n$. By applying the integration by parts formula, we obtain that

$$\begin{aligned}\mathbb{E}\{(\hat{r}(x)-r(x))g(x,Y)\} &= \sum_{i=1}^{n} W_i(x)\mathbb{E}\{(Y_i - r(x))g(x,Y)\} \\ &= \frac{1}{\sigma_\epsilon\sqrt{2\pi}}\sum_{i=1}^{n} W_i(x)\int_{\mathbb{R}}(t-r(x))g(x,Y^{i,t})(\exp -\frac{1}{2}(\frac{t-r(x)}{\sigma_\epsilon})^2)dt \\ &= \frac{\sigma_\epsilon}{\sqrt{2\pi}}\sum_{i=1}^{n} W_i(x)\int_{\mathbb{R}} g(x,Y^{i,t})\frac{\partial}{\partial t}(-\exp -\frac{1}{2}(\frac{t-r(x)}{\sigma_\epsilon})^2)dt \\ &= \frac{\sigma_\epsilon}{\sqrt{2\pi}}\sum_{i=1}^{n} W_i(x)\int_{\mathbb{R}}\frac{\partial}{\partial t}(g(x,Y^{i,t}))\exp -\frac{1}{2}(\frac{t-r(x)}{\sigma_\epsilon})^2 dt \\ &= \sigma_\epsilon\sum_{i=1}^{n} W_i(x)\mathbb{E}[\frac{\partial}{\partial Y_i}g(x,Y)] \\ &= \sigma_\epsilon\sum_{i=1}^{n} W_i(x)\mathbb{E}[\nabla_i g(x,Y)]\end{aligned}$$

So,

$$\begin{aligned}\triangle\mathcal{R}_0(x) &= \mathbb{E}\{g^2(x,Y)\} - 2\sigma_\epsilon \sum_{i=1}^{n} W_i(x)\mathbb{E}[\nabla_i g(x,Y)] \\ &= \mathbb{E}\{g^2(x,Y) - 2\sigma_\epsilon \prec W(x), \nabla g(x,Y) \succ\}\end{aligned}$$

Where

$$\nabla g(x,Y) = (\nabla_1 g(x,Y), ..., \nabla_n g(x,Y))$$

with

$$\nabla_i g(x,Y) = \frac{\partial g(x,Y)}{\partial Y_i}$$

and $\prec .,. \succ$ denote the euclidian inner product on $\mathbb{R}^n$.

4.4 Admissibility and dominance, Main results

The objective in this section is to show the influence of the choice of the balanced loss function on the possibility of improvement of the usual estimator (the kernel estimator). We begin with examples where the choice of BLF leads to finding a solution to this problem of improvement by shrinkage, then we cite examples of the opposite case where the usual estimator can not be dominated by another estimator obtained by contraction .

4.4.1 Quadratic Balanced loss function

We define ϕ by :

$$\phi(r(x), \hat{r}(x)) = \psi(r(x) - \hat{r}(x)) = (r(x) - \hat{r}(x))^2,$$

So the quadratic balanced loss function is defined by :

$$\begin{aligned} L(r(x), \hat{r}_g(x)) &= \omega(\hat{r}(x) - \hat{r}_g(x))^2 + (1-\omega)(r(x) - \hat{r}_g(x))^2 \qquad (4.8) \\ &= \omega(g(x,Y))^2 + (1-\omega)(r(x) - \hat{r}(x) + g(x,Y))^2, \end{aligned}$$

The associated difference in risk to this loss function is :

$$\begin{aligned} \triangle\mathcal{R}(x) &= \omega\mathbb{E}\{g^2(x,Y)\} + (1-\omega)\mathbb{E}\{(r(x) - \hat{r}(x) + g(x,Y))^2 - (r(x) - \hat{r}(x))^2\} \\ &= \omega\mathbb{E}\{g^2(x,Y)\} + (1-\omega)\triangle\mathcal{R}_0(x) \\ &= \mathbb{E}\{g^2(x,Y)\} - 2(1-\omega)\mathbb{E}\{(\hat{r}(x) - r(x))g(x,Y)\}, \end{aligned}$$

Applying the last equation and the proposition 4.1, we are able to give sufficient conditions on function g such that the estimator $\hat{r}_g$ dominate $\hat{r}$ through the following theorem.

Theorem 4.2. *Let $g : E \times \mathbb{R}^n \to \mathbb{R}$ be a positive differentiable function. A sufficient condition for the estimator $\hat{r}(x) - g(x,Y)$ to dominate $\hat{r}(x)$ under the quadratic balanced loss function (4.8) is :*

$$g^2(x,Y) - 2(1-\omega)\sigma_\epsilon \prec W(x), \nabla g(x,Y) \succ \leq 0 \qquad (4.9)$$

4.4.2 Weighted quadratic balanced loss function

Let us define the following function :

$$\phi(r(x), \hat{r}(x)) = \psi(r(x) - \hat{r}(x)) = \tau(r(x) - \hat{r}(x))^2, \qquad \tau > 0,$$

So the weighted quadratic balanced loss function is defined by :

$$\begin{aligned} L(r(x), \hat{r}_g(x)) &= \omega\tau(\hat{r}(x) - \hat{r}_g(x))^2 + (1-\omega)\tau(r(x) - \hat{r}_g(x))^2 \qquad (4.10) \\ &= \omega\tau(g(x,Y))^2 + (1-\omega)\tau(r(x) - \hat{r}(x) + g(x,Y))^2, \end{aligned}$$

The associated difference in risk to this loss function is :

$$\begin{aligned}\triangle\mathcal{R}(x) &= \omega\mathbb{E}\{\tau g^2(x,Y)\} + (1-\omega)\mathbb{E}\{\tau[(r(x)-\hat{r}(x)+g(x,Y))^2 - (r(x)-\hat{r}(x))^2]\} \\ &= \omega\mathbb{E}\{\tau g^2(x,Y)\} + (1-\omega)\mathbb{E}\{\tau[g^2(x,Y) + 2g(x,Y)(r(x)-\hat{r}(x,Y))]\} \\ &= \mathbb{E}\{\tau g^2(x,Y)\} - 2(1-\omega)\mathbb{E}\{\tau(\hat{r}(x)-r(x))g(x,Y)\},\end{aligned}$$

From the last equation and the proposition 4.1, the next theorem give sufficient conditions on function g such that the estimator $\hat{r}_g$ dominate $\hat{r}$ under the BLF (4.10).

Theorem 4.3. *Let $g : E \times \mathbb{R}^n \to \mathbb{R}$ be a positive differentiable function. A sufficient condition for the estimator $\hat{r}(x) - g(x,Y)$ to dominate $\hat{r}(x)$ under the weighted quadratic balanced loss function (4.10) is*

$$\tau g^2(x,Y) - 2(1-\omega)\tau\sigma_\epsilon \prec W(x), \nabla g(x,Y) \succ \leq 0 \qquad (4.11)$$

The difference in risk $\triangle\mathcal{R}(x)$ can also be written :

$$\begin{aligned}\triangle\mathcal{R}(x) &= \mathbb{E}\{\tau g^2(x,Y)\} - 2(1-\omega)\mathbb{E}\{\tau(\hat{r}(x)-r(x))g(x,Y)\} \\ &= \mathbb{E}\{\tau g^2(x,Y)\} - 2(1-\omega)[\mathbb{E}\{\tau^2 g^2(x,Y)\} - \triangle\mathcal{R}_0(\tau g)] \\ &= 2(1-\omega)\triangle\mathcal{R}_0(\tau g) + \mathbb{E}\{(\tau - 2(1-\omega)\tau^2)g^2(x,Y)\},\end{aligned}$$

where $\triangle\mathcal{R}_0(\tau g) = \mathcal{R}_0(r(x), \hat{r}_{\tau g}(x)) - \mathcal{R}_0(r(x), \hat{r}(x))$ is the difference in risk between the estimator $\hat{r}_{\tau g}(x) = \hat{r}(x) - \tau g(x,Y)$ and the the target estimator $\hat{r}$ with respect to the quadratic loss function. This formula leads to the following result :

Theorem 4.4. *Let $g : E \times \mathbb{R}^n \to \mathbb{R}$ be a positive differentiable function. Assume that the estimator $\hat{r}(x) - \tau g(x,Y)$ dominate the kernel estimator $\hat{r}(x)$*

under the quadratic loss function L_0. A sufficient condition for the estimator $\hat{r}(x) - g(x, Y)$ to dominate $\hat{r}(x)$ under the weighted quadratic balanced loss function (4.10) is

$$\tau - 2(1-\omega)\tau^2 \leq 0 \tag{4.12}$$

Remark 4.1. Inequality (4.12) control the good choice of the positive weight to improve the kernel estimator under weighted BLF and assumptions of dominance respectively to the quadratic loss. This means that τ must be chosen such that $\frac{1}{2(1-\omega)} < \tau$.

4.4.3 Logistic balanced loss function

We define ϕ by :

$$\phi(r(x), \hat{r}(x)) = \psi(r(x) - \hat{r}(x)) = -\ln \frac{1}{2}(1 + e^{-(r(x)-\hat{r}(x))^2}),$$

so the logistic balanced loss function defined by this choice is :

$$\begin{aligned} L(r(x), \hat{r}_g(x)) &= \omega(-\ln \frac{1}{2}(1 + e^{-g^2(x,Y)})) \\ &+ (1-\omega)(-\ln \frac{1}{2}(1 + e^{-(r(x)-\hat{r}(x)+g(x,Y))^2})), \end{aligned} \tag{4.13}$$

The difference in risk associated to the balanced loss function is :

$$\begin{aligned} \triangle\mathcal{R}(x) &= \omega\mathbb{E}\{-\ln \frac{1}{2}(1 + e^{-g^2(x,Y)})\} \\ &+ (1-\omega)\mathbb{E}\{\ln[1 + e^{-(r(x)-\hat{r}(x))^2}] - \ln[1 + e^{-(r(x)-\hat{r}(x)+g(x,Y))^2}]\}. \end{aligned} \tag{4.14}$$

For any non-negative function g, we have :

$$-\ln \frac{1}{2}(1 + e - g^2(x, Y)) \geq 0,$$

Since the logarithm and exponential functions are increasing, it is easy to see that :

$$\triangle\mathcal{R}(x) \geq 0.$$

We have this result :

Theorem 4.5. *For the regression model (4.1) and Under the logistic balanced loss function (4.13) the kernel estimator $\hat{r}(x)$ is admissible.*

4.4.4 Reflected normal balanced loss function

We define ϕ by :

$$\phi(r(x), \hat{r}(x)) = \psi(r(x) - \hat{r}(x)) = 1 - e^{-a(r(x)-\hat{r}(x))^2}, \qquad a > 0.$$

So, the reflected normal balanced loss function defined by this choice is :

$$\begin{aligned} L(r(x), \hat{r}_g(x)) &= \omega(1 - e^{-ag^2(x,Y)}) \\ &+ (1-\omega)(1 - e^{-a(r(x)-\hat{r}_g(x))^2}), \end{aligned} \tag{4.15}$$

The difference in risk associated to the balanced loss function is :

$$\begin{aligned} \triangle\mathcal{R}(x) &= \omega\mathbb{E}\{1 - e^{-ag^2(x,Y)}\} \\ &+ (1-\omega)\mathbb{E}\{e^{-a(r(x)-\hat{r}(x))^2} - e^{-a(r(x)-\hat{r}(x)+g(x,Y))^2}\}. \end{aligned} \tag{4.16}$$

Since the function g is non-negative and the exponential function is increasing, we have :

$$\triangle\mathcal{R}(x) \geq 0.$$

consequently, we give the following theorem :

Theorem 4.6. *For the regression model (4.1) and under the reflected normal balanced loss function (4.15) the kernel estimator $\hat{r}(x)$ is admissible.*

4.4.5 Linex balanced loss function

We propose a balanced loss function based on the following function :

$$\phi(r(x), \hat{r}(x)) = \psi(r(x) - \hat{r}(x)) = e^{a(r(x)-\hat{r}(x))} - a(r(x) - \hat{r}(x)) - 1, \qquad a > 0$$

So, the Linex balanced function is defined as :

$$\begin{aligned} L(r(x), \hat{r}_g(x)) &= \omega(e^{ag(x,Y)} - ag(x, Y) - 1) \\ &+ (1-\omega)(e^{a(r(x)-\hat{r}(x)+g(x,Y))} - a(r(x) - \hat{r}(x) + g(x, Y)) - 1) \end{aligned} \tag{4.17}$$

The difference in risk for this balanced loss function is :

$$\begin{aligned} \triangle\mathcal{R}(x) &= \omega\mathbb{E}\{e^{ag(x,Y)} - ag(x, Y) - 1\} \\ &+ (1-\omega)\mathbb{E}\{e^{a(r(x)-\hat{r}(x)+g(x,Y))} - e^{a(r(x)-\hat{r}(x))} - ag(x, Y)\}. \end{aligned}$$

So, we have :

$$\begin{aligned} \triangle\mathcal{R}(x) &= \omega\, \underbrace{\mathbb{E}\{e^{ag(x,Y)} - \frac{a}{\omega}g(x, Y) - 1\}}_{\alpha} \\ &+ (1-\omega)\, \underbrace{\mathbb{E}\{e^{a(r(x)-\hat{r}(x)+g(x,Y))} - e^{a(r(x)-\hat{r}(x))}\}}_{\beta}. \end{aligned} \tag{4.18}$$

Since, the last term of the equality (4.18) is positive, it result that the non-positivity of $\triangle\mathcal{R}(x)$ depends on the sign of the term :

$$e^{ag(x,Y)} - \frac{a}{\omega}g(x, Y) - 1,$$

and we can state the following theorem :

Theorem 4.7. *Under the Linex balanced loss function (4.17), a necessary condition for the estimator $\hat{r}_g(x)$ to dominate the kernel estimator $\hat{r}(x)$ is :*

$$ag(x, Y) < \ln(\frac{1}{\omega}).$$

Remark 4.2. The above theorem shows the theoretical possibility of the domination of the kernel estimator, but in this situation, finding a sufficient condition for the domination of the kernel estimator is not evident in reason of the difficulty to comparing the terms α and β of the equality (4.18).

4.4.6 Absolute value balanced loss function

The function ϕ is given by :

$$\phi(r(x), \hat{r}(x)) = \psi(r(x) - \hat{r}(x)) = |r(x) - \hat{r}(x)|,$$

Then, the absolute value balanced loss function is :

$$L(r(x), \hat{r}_g(x)) = \omega g(x, Y) + (1 - \omega)|r(x) - \hat{r}(x) + g(x, Y)| \tag{4.19}$$

The difference in risk under this balanced loss function is :

$$\triangle\mathcal{R}(x) = \omega\mathbb{E}\{g(x, Y)\} + (1 - \omega)\mathbb{E}\{|r(x) - \hat{r}(x) + g(x, Y)| - |r(x) - \hat{r}(x)|\}. \tag{4.20}$$

We remark that for all $x \in E$:

$$|r(x) - \hat{r}(x) + g(x, Y)| - |r(x) - \hat{r}(x)| \leq g(x, Y),$$

So, for $1 - \omega \leq \omega$ we have necessary :

$$\triangle\mathcal{R}(x) \geq 0,$$

Consequently, we have the following result :

Theorem 4.8. *For the regression model (4.1) and under the absolute value balanced loss function (4.19) with $\frac{1}{2} \leq \omega < 1$, the kernel estimator $\hat{r}(x)$ is admissible.*

Remark 4.3. If $0 < \omega < \frac{1}{2}$, the sign of the second term of equality (4.20) is unknown. So, we can not prove the admissibility or the inadmissibility of the kernel estimator $\hat{r}(x)$. This result confirm the remark (ix) of Charles Stein in [57] in the parametric case.

4.4.7 Concluding remarks

1. At the beginning, our motivation is to generalize the works of Jozani et al. [42] and Dey et al. [21] on the admissibility of the Stein type estimator for a nonparametric model. In this context we confirm an absence of work on this idea. Moreover, the class of balanced loss functions used up to the present day is restricted to the quadratic or quantum loss function with weight.

2. The results obtained can be applied to parametric regression models. The main idea is that the explanatory variable X in our model is of functional type (with values in a semi-metric space), so we can have analogous results for parametric models as studied by Dey et al. [21] or for a general multivariate linear model.

3. The nonparametric methods used for improving the usual estimators (smoothing methods and cross-validation idea) leads to choose between small bias and large variance (or the contrary) with respect to interesting arguments in the posed problem. But as all can see, Stein methods (shrinkage) leads to minimize the risk (and so the variance) without taking account the bias.

Chapitre 5

Conclusions et perspectives

5.1 Conclusions

La question principale évoquée dans cette thèse est l'amélioration par rétrécissement de l'estimateur usuel pour un modèle de régression non-paramétrique fonctionnelle (fixed-design), sous une fonction de perte non-quadratique.
Notre contribution à ce sujet est la proposition de certaines classes de fonctions de perte généralisant celle proposée par Zellner dans [65] (Balanced loss function). Par conséquent, on montre que l'estimateur usuel (l'estimateur à noyau dans notre cas) est :

1. Inadmissible relativement à certaines classes de ces fonctions de perte équilibrées, d'où la possibilité de l'améliorer par rétrécissement. Cela se réalise dans le cas des fonctions de perte : Quadratic balanced loss , Weighted quadratic Balanced loss, Linex balanced loss.

2. Admissible dans certains cas, notamment dans le cas des fonctions de perte : Logistic balanced loss, Reflected normal balanced loss, Absolute balanced loss avec $1/2 \leq \omega < 1$. Cela est un phénomène qui se réalise indépendamment à la dimension, contrairement à ce qui a été reconnu sur le phénomène de Stein.

5.2 Perspectives

Le travail de recherche effectué dans cette thèse ainsi que les questions évoquées, ont permet de clarifier les points de vue et de prendre connaissance des travaux de nombreux chercheurs dans ce domaine tels que Willam Strawderman, Tatsuya Kubokawa, Eric Marchand, Jafari Jozani et Malay Ghosh. Dans les perspectives, on envisage à répondre aux questions suivantes :

1. Le problème d'inadmissibilité (et de minimaxité) dans le cas où on remplace l'hypothèse de normalité par des hypothèses plus générales de symétrie sphérique ou de mélange, relativement à une fonction de perte quelconque.
2. Expliquer le lien entre le groupe de symétrie de loi du modèle et celui de la fonction de perte choisie. Cet argument géométrique est un argument qui peut expliquer définitivement le lien entre la question d'admissibilité et de minimaxité d'une part, et le choix optimal de la fonction de perte d'une autre part.
3. Répondre aux questions ouvertes dans [47] sur la fonction de perte

équilibbrée (BLF) non quadratique convenable pour un modèle d'estimation quelconque. Celon les auteurs de [47], un choix optimal pour la fonction de perte équilibrée doit passer par un examen de la fonction de perte (regret loss) associée.

4. Généraliser l'approche de la fonction de perte intégrée, proposée par Jozani et al. Dans [43] aux fonctions de perte équilibrées non-quadratiques. Même si le travail des auteurs au [43] est restreint à l'estimation de la fonction de répartition, mais j'estime que la méthode dont ils ont employée peut-être très raisonnable pour résoudre un problème d'estimation pour différents modèles.

Bibliographie

[1] K. Benhenni, F. Ferraty, M. Rachdi, P. Vieu, *Local smoothing regression with functional data*, Comput. Statist., **22**, (22), (2007), 353-369.

[2] K. Benhenni, S. Hedli-Griche, M. Rachdi and P. Vieu, *Consistency of the regression estimator with functional data under long memory conditions*, Statist. probab. Lett., Doi : 10.1016/j.spl.2007.11.011, (2008).

[3] G. Boente, R. Fraiman, *Kernel-based functional principal components*, Stat. Probab. Lett., **48**, (04), (2000), 335-345.

[4] D. Bosq, J.P. Lecoutre, *Théorie de l'estimation fonctionnelle*, Economica, Paris, (1987).

[5] D. Bosq, *Linear processes in function space : theory and application*, Lecture Notes in Statistics, **149**, Springer Series in Statistics, Springer-Verlag, New York, (2002).

[6] A.C. Brandwein, W.E. Strawderman, *Generalizations of James-Stein estimators under spherical symmetry*, The Annals of Statistics, **19**, (3), (1991), 1639-1650.

[7] A.C. Brandwein, S. Ralescu, W.E. Strawderman, *Shrinkage estimators of the location parameter for certain spherically symmetric distributions*,

Ann. Int. Statist. Math, **45**, (3), (1991), 551-563.

[8] A.C. Brandwein, W.E. Strawderman, *Stein Estimation for Spherically Symmetric Distributions : Recent Developments*, Statistical Science, **27**, (1), (2012), 11-23.

[9] L. D. Brown, *On the admissibility of invariant estimators of one or more location parameters*, Annals of Mathematical Statistics, **37**, (1966), 1087-1136.

[10] L. D. Brown, *Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters*, Annals of Mathematical Statistics, **39**, (1968), 29-48.

[11] L. D. Brown, *Estimation with Incompletely Specified Loss Functions (the Case of Several Location Parameters)*, Journal of the American Statistical Association, **70**, (350)(1975), 417-427.

[12] G. Casella, *Minimax ridge regression estimation*, Ann. Statist., **8**, (1980), 1036-1056.

[13] G. Casella, J.T. Huang, *Limit expressions for the risk of James-Stein estimators*, The Canadian Journal of Statistics, **10**, (4) (1982), 305-309.

[14] D. Cellier, D. Fourdrinier, C. Robert, *Robust shrinkage estimators of the location parameter for elliptically symmetric distributions*, Journal of Multivariate Analysis, **29**, (1989), 39-52.

[15] D. Cellier, D. Fourdrinier, *Shrinkage estimators under spherical symmetry for the general linear model*, Journal of Multivariate Analysis, **52**(2), (1995), 338-351.

[16] G. Collomb, *Estimation non-paramétrique de la régression par la méthode du noyau*, Thèse 3eme cycle, Toulouse **3**, (1976).

[17] G. Collomb, *Propriétés de convergence presque complète du prédicteur à noyau*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, **66**, (1984), 441-460.

[18] A. Cuevas, *A partial overview of the theory of statistics with functional data*, J. Statist. Plann. Inference, **147**, (2014), 1-23.

[19] D.K. Dey, C. Srinivasan, *Estimation of a covariance matrix under Stein's loss*, Ann. Statist, **13**, (1985), 1581-1591.

[20] D.K. Dey, M. Ghosh, C. Srinivasan, *Simultaneous estimation of parameters under entropy loss*, J. Statist. Plann. Inference, **15**, (1987), 347-363.

[21] D.K. Dey, M. Ghosh, W.E. Strawderman, *On estimation with balanced loss functions*, Statistics & Probability Letters, **45**, (1999), 97-101.

[22] F. Ferraty, P. Vieu, *Dimension fraclale et estimation de la régression dans des espaces vectoriels semi-normés*, Comptes Rendus Acad. Sci. Paris, **330**, (2000), 139-142.

[23] F. Ferraty, P. Vieu, *The functional non-parametric model and application to spectrometric data*, Comput. Statist. and Data Anal., **17**, (2002), 545-564.

[24] F. Ferraty, P. Vieu, *Non-parametric models for functional data, with application in regression,times series prediction and curves discrimination*, J. Nonparametric Statist., **16**, (2004), 111-127.

[25] F. Ferraty, P. Vieu, *Nonparametric functional data analysis. Theory and Practice*, Springer Series in Statistics. New York, (2006).

[26] D. Fourdrinier, M.T. Wells, *Estimation of a loss function for spherically symmetric distributions in the general linear model*, Annals of Statistics, **23**, (2), (1995), 571-592.

[27] D. Fourdrinier, M.T. Wells, *Loss estimation for spherically symmetric distributions*, Journal of Multivariate Analysis, **53**, (2), (1995), 311-331.

[28] D. Fourdrinier, M.T. Wells, *On Improved Loss Estimation for Shrinkage Estimators*, Statistical Science, **27**, (1), (2012), 61-81.

[29] E.I. George, *Minimax Multiple Shrinkage Estimation*, The Annals of Statistics, **14**, (1), (1986), 188-205.

[30] M. Ghosh, M.C. Yang, *Simultaneous estimation of Poisson means under entropy loss*, Ann. Statist., **16**, (1988), 278-291.

[31] A. Goia, P. Vieu, *An introduction to recent advances in high/infinite dimensional Statistics*, Journal of Multivariate Analysis, Doi : org/10.1016/j.jmva.2015.12.001, (2015).

[32] L.R. Haft, *Minimax estimators for a multinormal precision matrix*, J. Multivariate Anal, **7**, (1977), 374-385.

[33] L.R. Haft, *Estimation of the inverse covariance matrix : Random mixtures of the inverse Wishart matrix and the identity*, Ann. Statist., **7**, (1979), 1264-1276.

[34] L.R. Haft, *Empirical Bayes estimation of the multivariate normal c onvariance matrix*, The Annals of Statistics, **8**, (1980), 586-597.

[35] L.R. Haft, *Identities for the Wishart distribution with applications to regression*, Sankhya Ser. B, **44**, (1982), 245-258.

[36] W. Härdle, *Applied Nonparametric Regression*, Cambridge. Univ. Press, Cambridge, UK, (1990).

[37] W. Härdle, *Smoothing Technics with implementation*, Springer-Verlag, New York,, (1991).

[38] Ighodaro, A.O., T. Santer, *Ridge type estimators of multinormal cell probabilities*, In : S.S. Gupta and J.O. Berger, Eds., Statistical Decision Theory and Related Topics III, **2**, Academic Press, New York,(1982), 31-53 .

[39] A.O. Ighodaro, T. Santer, L.D. Brown, *Admissibility and complete class results for the multinomial estimation problem with entropy and squared error loss*, J. Multivariate Anal., **12**, (1982), 469-479.

[40] W. James, C. Stein, *Estimation with Quadratic Loss*. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability : Contributions to the Theory of Statistics, University of California Press, Berkeley, Calif, **1**, (1961), 361-379.

[41] I. Johnstone, *On admissibility of some unbiased estimates of loss*. Statistical decision theory and related topics IV, (S. S. Gupta, J. O. Berger Ed.), **1**, (1988), 361-379.

[42] M.J. Jozani, E. Marchand, A. Parsian, *On estimation with weighted balanced-type loss function*, Statistics & Probability Letters, **76**, (2006), 773-780.

[43] M.J. Jozani, A. Leblac, E. Marchand, *On continuous distribution functions, minimax and best invariant estimators, and integrated balanced loss functions*, preprint.

[44] T. Kubokawa, E. Marchand, W. Strawderman, *On improved shrinkage estimators for concave loss*, Statistics and Probability Letters, **96**, (2015), 241-246.

[45] E.L. Lehmann, G. Casilla, *Theory of Point Estimation*, Springer, New York, (1998).

[46] L. Lecam, *On the assumptions used to prove asymptotic normality of maximum Likelihood estimates*, Ann. Math. Statist., **41**, (1970), 802-828.

[47] T.K. Nayak, B. Sinha, *An Appreciation of Balanced Loss Functions Via Regret Loss*, Communications in Statistics - Theory and Methods, **44**, (2015), 607-616.

[48] A. Parsian, N. Nematollahi, *Estimation of scale parameter under entropy loss function*, Journal of Statistical Planning and Inference, **52**, (9), (1996), 77-91.

[49] M. Rachdi, P. Vieu, *Nonparametric regression for functional data : automatic smoothing parameter selection*, Journal of Statistical Planning and Inference, **137**, (9), (2007), 2784-2801.

[50] J. Ramsay, C. Dalzell, *Some tools for functional data analysis. (with discussion)*, J. R. Stat. Soc., Ser. B **53**, (3), (1991), 539-572.

[51] J. Ramsay, B. Silverman, *Functional data analysis*, Springer Series in Statistics, Springer-Verlag, New York, (1997).

[52] J. Ramsay, B. Silverman, *Applied Functional Data Analysis, : Methods and case studies*, Springer Series in Statistics, Springer-Verlag, New York, (2002). .

[53] C.P. Robert, *Le choix bayésien, Principes et pratique*, Statistique et Probabilités Appliquées, Springer-Verlag, Paris, (2006).

[54] A.L. Rukhin, M.M.A Ananda, *Risk behavior of variance estimators in multivariate normal distribution*, Statist. Probab. Lett., **13**, (1992), 159-166.

[55] P. Sarda, P. Vieu, *Kernel regression. Smoothing and Regression : Approaches, computation and application,*, Wiley Series in Probability and Statistics, (2000), 43-70.

[56] M. Schimek, *Smoothing and Regression : Approaches, computation and application,*, Wiley Series in Probability and Statistics, (2000).

[57] C. Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Procedings of the third Berkeley Symposium of Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press, **1**, (1956), 197-206.

[58] C. Stein, *Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean*, Ann. Inst. Statist. Math., **16**, (1964), 155-160.

[59] C. Stein, *Estimation of the mean of multivariate normal distribution*, Annals of Statistics, **9**, (1981), 1135-1151.

[60] P. Vieu, *Quadratic errors for nonparametric estimates under dependence*, J. Multiv. Anal., **39**, (1991), 324-347.

[61] P. Vieu, *Introduction aux méthodes non-paramétriques : propriétés asymptotiques des estimateurs à noyau de la régression*, Publi. Labo. Stat. Proba. Toulouse **3**, (1994), 11-94.

[62] M. Wand, C. Jones, *Kernel Smoothing*, Chapman and Hall, London, UK, (1995).

[63] R. Wieczorkowski, R. Zielinski, *Minimax estimation of binomial probability with entropy loss function*, Statist. Deci., **10**, (1992), 39-44.

[64] M.C. Yang, *Ridge estimation of independent Poisson means under entropy loss*, Statist. Dec., **10**, (1992), 1-23.

[65] A. Zellner, *Bayesian and Non-Bayesian Estimation Using Balanced Loss Functions*, Statistical Decision Theory and Related Topics V, Springer, New York, (1994), 337-390.

Résumé

La question principale évoquée dans cette thèse est l'amélioration par rétrécissement de l'estimateur usuel pour un modèle de régression non-paramétrique fonctionnelle (fixed-design), sous une fonction de perte non-quadratique.Notre contribution à ce sujet est la proposition de certaines classes de fonctions de perte généralisant celle proposée par Zellner dans [64] (Balanced loss function). Par conséquent, on montre que l'estimateur usuel (l'estimateur ànoyau dans notre cas) est : 1- Inadmissible relativement à certaines classes de ces fonctions de perteéquilibrées, d'où la possibilité de l'améliorer par rétrécissement. Celase réalise dans le cas des fonctions de perte : Quadratic balanced loss ,Weighted quadratic Balanced loss, Linex balanced loss. 2-Admissible dans certains cas, notamment dans le cas des fonctions de perte : Logistic balanced loss, Reflected normal balanced loss, Absolute balanced loss avec $1/2 < \omega < 1$. Cela est un phénomène qui se réalise indépendamment à la dimension, contrairement à ce qui a été reconnusur le phénomène de Stein.

Summary

The main question raised in this thesis is the improvement of the usual estimator for a nonparametric functional regression model (fixed design), under a non-quadratic loss function. Our contribution on this subjectis the proposal of some classes of loss functions generalizing that proposed by Zellner in [64] (balanced loss function). Therefore, we show that the usual estimator (the kernel estimator in this case) is: 1- Inadmissible for some classes of these balanced loss functions, hence the possibility to improve it by shrinkage. This is realized in the case of loss functions : quadratic balanced loss, weighted quadratic balanced loss, Linex balanced loss. 2-Admissible in some cases, particularly for loss functions : balanced logistic loss, reflected normal balancsd loss, absolute balanced losswith$1/2 < \omega < 1$. This is a phenomenon that occurs independently of the dimension, contrary to what was recognized on the Stein phenomenon.

الملـخص

السؤال الرئيسي الذي أثير في هذه الأطروحة هو تضييق التحسن للمقدر المعتاد لنموذج الانحدار الوظيفي غيرالبارامتري تحت وظيفة الفقدان غير التربيعي. مساهمتنا في هذا الموضوع هي اقتراح بعض فئات وظائف الخسارة التي عممت من طرف Zellner لذلك، نوضح أن المقدر المعتاد (مقدر النواة في حالتنا) هو : 1-- غير مقبول فيما يتعلق بفئات معينة من وظائف الخسارة المتوازنة، وبالتالي إمكانية تحسينها عن طريق التقلص. يتم تحقيق ذلك في حالة وظائف الخسارة: الخسارة المتوازنة التربيعية، الخسارة المتوازنة التربيعية مع معامل وزن، الخسارة المتوازنة الخطية الأسيــة. 2- - مقبول في بعض الحالات، لاسيما في حالة وظائف الخسارة: الخسارة اللوجستية المتوازنة،الخسارة الطبيعية المتوازنة المنعكسة،الخسارة المتوازنة المطلقة. وهذه ظاهرة تتحقق بشكل مستقل عن البعد، على عكس ماتم الاعتراف به في ظاهرة Stein.