Habilitation à Diriger des Recherches

Contributions à l'Estimation Fonctionnelle

Synthèse des travaux de recherche

Pascal Sarda

Gérard Collomb a guidé mes premiers pas en recherche et m'a initié à la théorie de l'estimation fonctionnelle. La qualité et l'apport de ses travaux, reconnus au plan international, l'éthique et l'enthousiasme qui animaient son travail de recherche sont des modèles. Je tiens à saluer sa mémoire et veux souligner que les travaux présentés dans ce document lui doivent beaucoup.

Remerciements

Je tiens à exprimer ma reconnaissance à Henri Caussinus pour l'attention qu'il a porté à mon travail. Ses conseils et idées ont été précieux et ont guidé ma recherche au cours de ma thèse.

Je remercie Jean-Marc Azaïs de l'intérêt qu'il a porté à mes recherches et d'avoir accepté d'en être le Directeur.

Je remercie également Alain Berlinet avec qui j'ai eu l'occasion de travailler à l'IUT de Montpellier. Les discussions scientifiques avec lui sont toujours enrichissantes. Je tiens à lui témoigner ma reconnaissance pour avoir accepté d'être rapporteur de mon travail et pour le soin avec lequel il a relu ce document et les suggestions qui ont contribué à l'améliorer.

Au cours du séjour de Dag Tjøstheim au sein du Laboratoire de Statistique et Probabilités, j'ai eu l'occasion d'avoir de nombreuses discussions scientifiques avec lui. Ses remarques et points de vue pertinents m'ont beaucoup apporté. J'ai pu également apprécier ses qualités humaines. Je tiens à le remercier sincèrement d'avoir accepté d'être rapporteur de mes travaux et de participer au jury.

Je voudrais remercier également Anestis Antoniadis pour l'honneur qu'il m'a fait en acceptant d'être rapporteur de ce travail et Jim Ramsay pour celui de s'intéresser à mon travail et de participer au jury.

Je tiens à remercier particulièrement Philippe Vieu avec lequel travailler en collaboration a toujours été enrichissant et dynamisant. Merci pour son soutien et son amitié.

Je voudrais aussi remercier Làzlò Györfi, Wolfgang Härdle et Michel Lejeune pour leur accueil lors de mes séjours dans leurs laboratoires et pour les pistes de recherche qu'ils m'ont apporté. Mes remerciements vont également à Hervé Cardot et Frédéric Ferraty avec lesquels mes derniers travaux ont été réalisés. Leur collaboration a été enrichissante sur les plans scientifique et humain. Merci aussi à Nour Eddin El Faouzi, Elie Youndjé et André Mas pour leur collaboration fructueuse.

Je remercie également Yves Romain pour ses remarques et idées pertinentes lors des nombreuses discussions que nous avons eues.

Merci à tous les membres du Laboratoire de Statistique de Montpellier et ceux du Laboratoire de Statistique et Probabilités de l'Université Paul Sabatier qui m'ont apporté aide et soutien et parmi eux Roger Lafosse, Elisabeth Lambert pour sa disponibilité et Françoise Michel pour son efficacité et son dévouement.

Table des matières

Introduction

Le cadre général des travaux de recherche présentés dans ce rapport de synthèse, est celui de l'estimation fonctionnelle. Cette spécialité a donné lieu à une littérature abondante depuis une trentaine d'années et le nombre des publications dans les revues les plus reconnues à l'échelle internationale indique qu'elle est un des centres d'intérêt de la Statistique actuelle. Les utilisateurs ont désormais fréquemment recours à des méthodes non paramétriques, rendues accessibles par ces travaux scientifiques et par leur intégration dans les logiciels de Statistique. Notre souci dans ce document est de préciser notre apport à cette problématique générale.

Nous avons découpé l'ensemble de notre travail en quatre parties principales. Nous décrivons tout d'abord trois grands axes de recherche. Il nous a paru logique de présenter les différents thèmes abordés dans un ordre respectant leur chronologie. Ce choix a été guidé par la volonté de replacer notre travail dans un contexte général, en soulignant l'évolution des préoccupations. Dans une quatrième partie nous avons regroupé des travaux plus ponctuels ne se rattachant pas directement à l'un de ces trois thèmes mais qui devraient donner lieu à des développements futurs.

La première partie de ce travail a trait pour une part aux propriétés mathématiques d'estimateurs non paramétriques. Différents cadres d'estimation fonctionnelle ont été abordés : régression, densité et fonction de hasard. Notre étude a porté sur le cas d'observations non nécessairement indépendantes, plus précisément soumises à une hypothèse de mélange. La plupart de ces travaux ont été réalisés entre 1985 et 1990 et un ouvrage général **(11)**[1] synthétise l'ensemble des résultats concernant ce sujet. Un autre aspect étudié dans cette même période et plus récemment, est celui du problème du choix du paramètre de lissage dans une estimation non paramétrique. Nous avons plus particulièrement étudié les propriétés de méthodes de sélection du paramètre de lissage reposant sur la validation croisée. Bien qu'ils ne constituent pas le point central de nos recherches récentes, nous présentons brièvement au chapitre 1 l'ensemble de ces premiers travaux initiés durant la thèse d'Université par Gérard Collomb.

[1] Les références en gras renvoient à la liste de publications à la fin de ce document.

Un aspect important des méthodes non paramétriques est lié à la "dimensionalité" du problème étudié. Dans les cas les plus usuels (densité et régression par exemple), l'étude théorique de ces méthodes mais également leur application à des cas concrets, ont mis à jour le "fléau de la dimension". Les travaux fondateurs à ce sujet sont probablement ceux de Stone (1982 et 1983) dans lesquels l'auteur donne les vitesses de convergence optimales pour un estimateur non paramétrique de la densité ou de la régression : ces vitesses dépendent de manière cruciale de la dimension du vecteur aléatoire étudié. Ainsi, les méthodes non paramétriques "pures" sont souvent inadaptées lorsque la dimension est grande ou/et lorsque la taille de l'échantillon n'est pas suffisamment importante. Dans le cadre de la régression essentiellement, des méthodes alternatives ont été proposées dont le but est de réduire la dimensionalité : il s'agit par exemple du modèle additif pour lequel la régression s'écrit comme la somme de fonctions univariées des covariables. Tout en restant dans un cadre relativement flexible, celui de l'estimation non paramétrique, l'erreur d'estimation n'est pas, d'un point de vue asymptotique, soumise à la dimensionalité du problème étudié. La deuxième partie de notre travail a trait à l'estimation non paramétrique multidimensionnelle et regroupe deux aspects. Le premier concerne l'estimation par noyaux des composantes additives d'une fonction de régression. Le second a pour objet l'estimation de Composantes Principales Additives pour lesquelles les vitesses de convergence d'estimateurs splines sont étudiées. Nous décrivons ces travaux au chapitre 2.

Plus récemment, nous nous sommes intéressés au cas de variables fonctionnelles. Dans le contexte de la régression de nombreuses études traitent ce type de données : la variable explicative est une courbe et la variable à expliquer un scalaire. Les méthodes utilisées dans ce type de problème sont la plupart du temps les méthodes paramétriques comme la Ridge Regression. Plusieurs études en ont montré les limites lorsque la courbe explicative n'est pas "lisse" par exemple. Il nous est apparu intéressant de nous tourner vers des méthodes "fonctionnelles" : dans ce cadre nous avons plus particulièrement étudié le modèle linéaire fonctionnel pour lequel la liaison entre la courbe et la variable à expliquer s'écrit à l'aide d'un opérateur linéaire continu. Plusieurs travaux ont été réalisés à ce jour : ils sont présentés au chapitre 3 de ce document.

Outre les travaux cités ci-dessus, nous avons regroupé au chapitre 4 des travaux concernant l'estimation de quantiles conditionnels ou encore l'estimation de points de ruptures. Ces travaux ne se rattachent pas directement aux problématiques des trois parties décrites ci-dessus. Pourtant ils nous semblent avoir leur place dans cette synthèse et être des points de départ de recherches ultérieures.

D'un point de vue global, nous avons, dans chaque partie du document, situé nos travaux dans le cadre général de la littérature en donnant les références qui nous paraissent les plus caractéristiques et fondatrices sur le sujet, sans chercher à atteindre l'exhaustivité, tâche qui nous paraît par ailleurs difficile compte tenu de l'abondance en la matière. Comme tout travail de recherche, le travail présenté s'est nourri de nombreuses collaborations mais également de discussions plus informelles. Il est assez difficile de rendre compte des secondes dans ce type de document. Nous soulignons au travers du manuscrit les premières et le contexte dans lequel elles ont été établies.

1 Quelques propriétés d'estimateurs non paramétriques. Choix d'estimateurs.

1.1 Généralités et repères historiques.

Nous fixons tout d'abord brièvement le cadre général de l'estimation fonctionnelle. Reprenons les notations de Bosq et Lecoutre (1987) et considérons un **modèle statistique** $(E, \mathcal{A}, \mathcal{P})$, où E est l'espace des observations, $\mathcal{A}$ une tribu sur E et $\mathcal{P}$ une famille de mesures de probabilités sur $(E, \mathcal{A})$. On écrit généralement la famille $\mathcal{P}$ sous la forme :

$$\mathcal{P} \; = \; \{P_\theta, \; \theta \in \Theta\} \,.$$

Le problème posé est celui de l'estimation du paramètre $g(P_\theta)$, où :

- g est une fonction définie sur $\mathcal{P}$ et à valeurs dans un espace Θ',
- P_θ est une loi inconnue dans $\mathcal{P}$,
- Θ' est un espace vectoriel de dimension infinie, autrement dit un espace fonctionnel.

Bien que la frontière entre modèle paramétrique et modèle non paramétrique soit assez floue (*cf.* Bosq et Lecoutre, 1987, à ce sujet), nous adoptons dans notre travail la définition suivante, définition implicitement admise dans une grande partie de la littérature.

> Le modèle d'estimation du paramètre $g(P_\theta)$ est dit **non paramétrique** lorsque les hypothèses sur $(E, \mathcal{A}, \mathcal{P})$ ne permettent pas d'écrire $g(P_\theta)$ en fonction d'un nombre fini de paramètres réels.

Compte tenu de la nature de l'espace Θ' et de la définition ci-dessus, nous parlons dans la suite indifféremment d'estimation fonctionnelle ou d'estimation non paramétrique.

De nombreux estimateurs non paramétriques de la régression ont en commun l'approche suivante. Considérons un couple aléatoire (X, Y) à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ et r la fonction de régression définie par $r(.) = \mathbb{E}(Y|X = .)$. Etant donné n couples (X_i, Y_i) de même loi que (X, Y), pour tout x de $\mathbb{R}^d$, l'estimateur $r_n(x)$ de $r(x)$ est une moyenne pondérée des observations Y_i, la pondération de Y_i tenant compte de la "distance" entre X_i et x.

Soit $X_i, i = 1, \ldots, n$ des variables à valeurs dans $I\!R^d$ de même densité f et x un élément de $I\!R^d$, le principe de construction de plusieurs estimateurs non paramétriques de la densité $f(x)$ est basé sur la proportion d'observations (pondérées) X_i "proches" de x.

Parmi les estimateurs reposant sur ces principes, les plus abondamment étudiés et utilisés ont été les estimateurs à noyau et les estimateurs splines. Dans la première partie de notre travail, nous avons plus particulièrement étudié les premiers (*cf.* section 1.2) tandis que nous nous sommes intéressés aux seconds dans des travaux plus récents[2]. Les résultats décrits ci-dessous ont été établis sous l'hypothèse minimale de continuité de la fonction à estimer puisque nous nous sommes intéressés principalement à des résultats de convergence ponctuelle ou uniforme sur un compact. Notons que des résultats de convergence universelle (voir Collomb, 1981, pour cette notion) ont été établis pour des estimateurs à noyaux ou des plus proches voisins lorsque la fonction n'est pas nécessairement continue : *cf.* par exemple Stone (1977) et Devroye et Wagner (1979 et 1980).

La principale particularité des estimateurs non paramétriques étudiés est qu'ils dépendent d'un paramètre contrôlant leur régularité : le paramètre de lissage. Ainsi, les techniques d'estimation non paramétriques ne sont pas, pour la plupart d'entre elles, "automatiques". Cette particularité a induit une littérature abondante sur le problème crucial du choix de ce paramètre. Nous revenons sur ce point à la section 1.3.

D'un point de vue historique, on trouve des traces relativement anciennes d'estimateurs non paramétriques. Pearson et Lee (1903), qui introduisent la droite de régression, utilisent pour calculer les coefficients de cette droite un régressogramme, équivalent de l'histogramme pour la régression. Macauley (1931) introduit une méthode d'ajustement local de la tendance d'une série chronologique par un polynôme, méthode reprise par Cleveland (1979) dans le contexte de la régression et dont une variante est devenue un des

[2]Notons qu'il existe deux types d'estimateurs splines dans la littérature : ceux reposant sur les "splines de lissage" et ceux reposant sur les "splines de régression". Nous avons utilisé dans notre travail les splines de régression (ainsi que les splines hybrides : *cf.* section 3). Nous renvoyons à l'ouvrage de référence de Wahba (1990) sur l'estimation par splines de lissage. Afin d'alléger le texte, nous parlons d'estimateurs splines étant entendu qu'il s'agit ici de splines de régression.

standards en matière d'estimateurs non paramétriques de la régression. Pour autant l'étude mathématique des méthodes d'estimation non paramétriques date d'une quarantaine d'années et correspond au développement des moyens automatiques de calcul. Citons en densité les travaux de Rosenblatt (1956), Parzen (1962) et en régression ceux de Tukey (1961), Nadaraya (1964) et Watson (1964).

1.2 Propriétés de convergence.

Les travaux présentés ci-dessous ont été réalisés au cours de la thèse et dans les deux années qui ont suivi, c'est-à-dire entre 1985 et 1990. Ils ont été amorcés sous la direction de Gérard Collomb et ont été le fruit de collaborations avec P. Vieu et S. Hassani au laboratoire de Statistique et Probabilités, W. Härdle, lors de deux séjours à l'Université de Bonn et L. Györfi, lors d'un séjour à l'Université de Budapest.

De nombreux auteurs se sont attachés à établir des propriétés mathématiques d'estimateurs non paramétriques : étude du biais, de la variance, convergence. Au début des années quatre-vingts, la plupart des résultats sont donnés pour des estimateurs construits à l'aide d'un échantillon de variables indépendantes et identiquement distribuées. Par ailleurs, dans un grand nombre de travaux, l'hypothèse minimale sur la fonction à estimer (densité ou régression) est que celle-ci est continûment dérivable à l'ordre k, $k \geq 0$ (c'est-à-dire continuité pour $k = 0$) en un point x lorsque des résultats ponctuels sont établis ou sur un voisinage d'un compact C lorsqu'on s'intéresse à des propriétés uniformes sur le compact C (voir cependant les travaux cités à la section précédente pour des résultats sous des hypothèses plus générales).

Parmi les résultats les plus caractéristiques, citons l'obtention des vitesses de convergence en norme L_q, $1 \leq q \leq \infty$, d'estimateurs non paramétriques, c'est-à-dire :

$$n^{-k/(2k+d)}, \qquad q < \infty,$$

$$\left(\frac{n}{log\, n}\right)^{-k/(2k+d)}, \qquad q = \infty.$$

Ces vitesses sont les vitesses de convergence optimales données par Stone (1982 et 1983). Outre les hypothèses de régularité sur la fonction à estimer, les auteurs font intervenir des hypothèses probabilistes plus ou moins restrictives suivant la nature des résultats démontrés. Le présent document n'est pas centré sur la discussion de ce problème et nous renvoyons pour cela aux revues bibliographiques et ouvrages généraux de Collomb (1981 et 1985), Silverman (1986), Devroye (1987), Bosq et Lecoutre (1987), Härdle (1990) et Izenman (1991). Nous avons pour notre part réalisé un travail de synthèse reprenant les propriétés de convergence pour l'estimateur à noyau de la régression et celui de la densité dans ce cadre général. Ce travail constitue un chapitre d'un ouvrage collectif à visée didactique **(12)**.

Dans la première partie de notre travail et suivant les travaux de Collomb (1984), nous nous sommes intéressés à l'estimation non paramétrique à partir d'observations non nécessairement indépendantes. Les estimateurs étudiés dans différents cadres fonctionnels reposent sur la méthode du noyau ou sur des méthodes apparentées à cette méthode (méthode combinant noyau et k plus proches voisins par exemple). Nous avons tout d'abord étudié les propriétés de ces estimateurs sous une hypothèse de φ-mélange. Les résultats établis sont pour l'essentiel des résultats de convergence presque complète uniforme sur un compact C. Nous donnons ci-dessous un rapide descriptif de ces travaux.

– **Estimation de la régression. Application à la prédiction de séries temporelles.**
  Etant donné n couples (X_i, Y_i) de variables aléatoires à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ (non nécessairement équidistribués), on suppose l'existence d'une fonction de régression commune r :

$$\mathbb{E}(Y_i|X_i) \ = \ r(X_i), \quad i = 1, \ldots, n.$$

  Etant donné un noyau K sur $\mathbb{R}^d$ et une suite $(h_n)_n$ de nombres positifs, l'estimateur à noyau de r (Nadaraya, 1964, Watson, 1964) est défini en un point x de $\mathbb{R}^d$ par :

$$r_n(x) \ = \ \frac{(nh_n^d)^{-1} \displaystyle\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right)}{(nh_n^d)^{-1} \displaystyle\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)}.$$

  Nous montrons tout d'abord dans **(14)** la convergence uniforme presque complète de l'estimateur r_n vers r sur un compact C. Une borne supérieure de la vitesse de convergence de ce même estimateur est donnée dans **(5)** tandis que la vitesse de convergence pour l'estimateur d'une dérivée de la régression obtenu par dérivation de r_n, est étudiée dans **(6)**. L'ensemble de ces articles généralisent des résultats précédents de Collomb (1984) et Collomb et Härdle (1986). La généralisation porte sur les hypothèses sur Y_i : dans nos travaux nous ne supposons pas que Y_i est borné mais nous imposons une hypothèse d'existence d'un moment exponentiel ou une hypothèse plus faible d'existence d'un moment d'ordre $\beta > 2$. Pour traiter ces deux cas nous utilisons dans les démonstrations une technique de troncation. Nous appliquons ensuite ces résultats au problème de la prédiction pour une série temporelle.

– **Estimation de la fonction de hasard**
Etant donné une variable aléatoire X à valeurs dans $I\!\!R^d$, de fonction de répartition F et de densité f, la fonction de hasard est définie en tout point x de $I\!\!R^d$, tel que $F(x) < 1$, par :

$$\lambda(x) \;=\; \frac{f(x)}{1 - F(x)}.$$

Dans le cas $d = 1$ et lorsque la variable X représente une durée de vie, la fonction λ correspond à une probabilité "instantanée" de décès à l'instant x. Nous avons tout d'abord réalisé dans **(15)** une revue bibliographique des différentes méthodes d'estimation non paramétrique de la fonction λ.

Nous avons par la suite étudié l'estimateur à noyau (*cf.* Murthy, 1965 et Rice et Rosenblatt, 1976) défini par :

$$\lambda_n(x) \;=\; \frac{f_n(x)}{1 - n^{-1}\sum_{i=1}^{n} 1\!\!1_{[X_i \le x]}} \;=\; \frac{(nh_n^d)^{-1}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)}{1 - n^{-1}\sum_{i=1}^{n} 1\!\!1_{[X_i \le x]}}.$$

Dans **(4)** et **(13)**, nous donnons des résultats de convergence presque complète uniforme sur un compact C de l'estimateur λ_n sous l'hypothèse de φ-mélange. Une borne supérieure pour la vitesse de convergence sous cette même hypothèse a été donnée dans **(18)**.

Un autre estimateur a été également étudié dans **(4)** et **(13)** : il s'agit de l'estimateur obtenu en combinant dans le numérateur ci-dessus l'estimateur à noyau et l'estimateur des k-plus proches voisins (Loftsgaarden et Quesenberry, 1965). On obtient alors le nouvel estimateur :

$$\tilde{\lambda}_n(x) = \frac{(nR(k_n,x)^d)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{R(k_n,x)}\right)}{1 - n^{-1} \sum_{i=1}^{n} 1\!\!1_{[X_i \leq x]}},$$

où

$$R(k_n,x) = \inf\{\lambda \in I\!R^+ \ : \ \#\{X_i, i = 1, \ldots, n, |x - X_i| \leq \lambda\} \geq k_n\},$$

et où $(k_n)_n$ est une suite de nombres entiers non nuls. Nous avons établi la convergence uniforme presque complète de cet estimateur dans **(4)** et **(13)**. L'ensemble de ces travaux a été repris dans un article synthétique **(20)**.

Notons que compte tenu de la structure des estimateurs ci-dessus dans lesquels le numérateur est l'estimateur à noyau, respectivement des k-plus proches voisins, de la densité f, des résultats de convergence en estimation de la densité sont établis dans ces mêmes articles.

– **Estimation de la fonction de répartition**
Au vu de l'écriture des estimateurs de la fonction de hasard ci-dessus, nous nous sommes naturellement intéressés à la fonction de répartition empirique dans **(4)**, **(13)** et **(18)**. Nous avons établi des résultats de convergence presque complète uniforme pour cet estimateur de la fonction de répartition sous l'hypothèse de φ-mélange. Il est à noter que l'estimation de la fonction de répartition est un problème d'estimation fonctionnelle un peu différent de ceux étudiés ci-dessus. En effet, la fonction de répartition empirique est un estimateur sans biais et on montre que sa vitesse de convergence est, quelle que soit la dimension du vecteur X, de l'ordre de $n^{-1/2}$.

Dans l'ouvrage général **(11)**, nous avons d'une part repris l'ensemble des travaux ci-dessus et d'autre part étendu ces résultats aux cas du ρ-mélange et de l'α-mélange, hypothèses moins restrictives que le φ-mélange. Notons que les démonstrations reposent principalement sur l'utilisation de résultats généraux d'Analyse (du type de celui donné par le Théorème de Bochner) pour traiter le biais de l'estimateur et celle d'une inégalité de type Bernstein pour la variance. Ces inégalités ont été établies par Collomb (1984) pour le cas du φ-mélange, par Peligrad (1988) pour le ρ-mélange et Carbon (1983) pour l'α-mélange.

Pour conclure cette section, notons que les résultats établis dans le cadre ci-dessus et dont le travail **(11)** constitue un des premiers ouvrages généraux sur ce sujet, ont depuis été améliorés dans la littérature. Nous pensons en particulier à la vitesse de convergence dont il a été établi qu'elle est la même que celle obtenue dans le cas indépendant et est donc la vitesse de convergence optimale, au moins dans le cas de la densité et de la régression. Nous renvoyons pour ces résultats aux ouvrages de Doukhan (1994) et de Bosq (1996) (voir également la section 2.1 ci-dessous).

1.3 Choix du paramètre de lissage.

Pour une part les travaux présentés dans cette section ont été réalisés en collaboration avec G. Collomb et P. Vieu dans la même période que les travaux décrits à la section précédente. Ils ont été poursuivis à la suite de la thèse et jusqu'en 1993. Ces derniers travaux comprennent une collaboration avec J. Mielniczuk lors de son séjour au sein de notre laboratoire, des résultats établis avec P. Vieu pour l'estimation de la fonction de hasard et enfin des résultats établis avec E. Youndjé et qui constituent une partie de sa thèse que nous avons co-encadrée.

Au vu des propriétés asymptotiques des estimateurs non paramétriques, et particulièrement des estimateurs à noyau, nul doute que le choix du paramètre de lissage est crucial pour la qualité de ces estimateurs. Celui-ci influe de manière prépondérante sur l'aspect, aussi bien local que global, de la courbe estimée et par conséquent sur l'interprétation que l'on peut en faire. On dispose, dans la plupart des cas, de l'expression du paramètre de lissage minimisant l'erreur quadratique moyenne intégrée asymptotique mais cette largeur de fenêtre "idéale" (relativement au critère d'erreur retenu) n'est pas directement calculable puisqu'elle dépend généralement de la dérivée seconde de la fonction à estimer. Les auteurs se sont alors attachés à introduire des procédures de sélection automatiques et donc moins subjectives que le simple choix "a l'oeil" du paramètre de lissage. L'étude de ce problème a nourri une littérature abondante, notamment vers le milieu des années quatre-vingts.

Plusieurs méthodes ont été proposées dans la littérature. Les méthodes reposant sur la validation croisée sont parmi les plus populaires. D'un point de vue pratique, un des prinicpaux intérêts de ces méthodes est leur caractère "direct". Parmi les travaux théoriques sur ces méthodes (optimalité asymptotique) dans le cadre des estimateurs à noyau, citons Stone (1984) et Marron (1987), pour la densité, et Härdle et Marron (1985), pour la régression. Une méthode de ré-injection (ou plug-in) a également été introduite. Son principe repose sur l'estimation de la (ou des) quantité(s) inconnue(s) (dépendant principalement de la dérivée seconde) dans l'expression du paramètre de lissage minimisant l'erreur quadratique moyenne intégrée. Bien que la littérature sur le plug-in soit moins abondante que pour la validation croisée, elle suscite depuis une dizaine d'années un intérêt croissant : voir par exemple à ce sujet Hall *et al.* (1991) ou Ruppert *et al.* (1993) pour les estimateurs à noyau de la densité et de la régression respectivement. Sur les méthodes plug-in, nous renvoyons à la synthèse bibliographique de Biau (1999). De manière générale, parmi les nombreuses revues bibliographiques concernant les méthodes de sélection du paramètre de lissage, citons celles de Marron (1988), Vieu (1993), Berlinet et Devroye (1994) et Cao *et al.* (1994).

Nous avons pour notre part étudié les méthodes reposant sur la validation croisée. Un premier article **(16)** porte sur l'estimation de la régression. Dans cet article nous établissons un résultat de convergence de l'estimateur à noyau construit à l'aide de la largeur de fenêtre choisi par validation-croisée. Nous décrivons ci-dessous plus en détails les travaux réalisés par la suite.

- **Estimation de la densité**
  Nous reprenons les notations de la section précédente avec $d = 1$ et nous revenons au cas de variables indépendantes. Dans **(19)**, nous adaptons la méthode introduite par Rudemo (1982) et Bowman (1984) pour le choix d'une largeur de fenêtre locale c'est-à-dire dépendant du point où la fonction est estimée. Le critère de validation croisée s'écrit :

$$CV_x(h_n) \;=\; \int_a^b f_n^2(y) w_{n,x}(y) dy \;-\; 2n^{-1} \sum_{i=1}^{n} f_{n,-i}(X_i) w_{n,x}(X_i),$$

  où f_n est l'estimateur à noyau de la densité et $f_{n,-i}$ est l'estimateur à noyau construit à l'aide de l'échantillon privé de la $i^{\text{ème}}$ observation. La suite de fonctions $(w_{n,x})_n$ concentrées autour du point x rend le choix de la largeur de fenêtre locale. Il est facile de montrer que $CV_x(h_n)$ est un estimateur sans biais des deux premiers termes du développement de l'erreur quadratique moyenne intégrée "locale" définie par :

$$EQMI_x(h_n) = \mathbb{E}\int_a^b (f_n(y) - f(y))^2 w_{n,x}(y)dy.$$

Nous montrons dans **(19)** l'optimalité asymptotique de la largeur de fenêtre $h(x)$ minimisant $CV_x(h_n)$, c'est-à-dire :

$$\sup_{x\in[a,b]} \frac{EQMI_x(h(x))}{\inf_{h_n\in H_n} EQMI_x(h_n)} \longrightarrow 1, \; p.s.$$

Nous renvoyons à **(19)** pour la définition de l'ensemble H_n.

- **Estimation de la fonction de hasard**
  Dans **(21)** nous nous intéressons au choix de la largeur de fenêtre pour l'estimateur à noyau de la fonction de hasard défini ci-dessus. Le cadre général est celui d'un échantillon de variables indépendantes et identiquement distribuées. Le critère de choix repose sur la validation croisée et s'écrit :

$$CV(h) = \int \lambda_n^2(x)w(x)dx - 2n^{-1}\sum_{i=1}^n \frac{f_{n,-i}(X_i)w(X_i)}{\left(1 - n^{-1}\sum_{i=1}^n \mathbb{1}_{[X_i\le x]}\right)^2},$$

  où w est une fonction de poids positive. Nous montrons alors que la largeur de fenêtre $\hat{h}_n$ minimisant ce critère vérifie :

$$\frac{EQI(\hat{h}_n)}{\inf_{h_n\in H_n} EQI(h_n)} \longrightarrow 1, \; p.co.,$$

  où EQI est l'erreur quadratique intégrée. Ce résultat est repris dans **(20)** où une généralisation du critère ci-dessus est donnée pour traiter le cas d'observations dépendantes.
- **Estimation d'une densité conditionnelle**
  Etant donné un couple aléatoire (X, Y) à valeurs dans $\mathbb{R}^2$, l'estimateur à noyau de la densité conditionnelle φ de Y sachant X est défini, à partir

de l'échantillon $(X_i, Y_i), i = 1, \ldots, n$, par :

$$\varphi_n(y|x) \;=\; \frac{(nh_n^2)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) K\left(\frac{y - Y_i}{h_n}\right)}{(nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)}.$$

Le critère de validation croisée s'écrit :

$$CV(h_n) \;=\; n^{-1} \sum_{i=1}^{n} \int \varphi_{n,-i}^2(y|X_i) w_1(y) dy w_2(X_i)$$

$$-2n^{-1} \sum_{i=1}^{n} \varphi_{n,-i}^2(Y_i|X_i) w_1(Y_i) w_2(X_i),$$

où $\varphi_{n,-i}$ est l'estimateur à noyau défini à l'aide de l'échantillon privé de la $i^{\text{ème}}$ observation et où w_1 et w_2 sont des fonctions de poids positives. Dans **(23)** nous montrons l'optimalité asymptotique de la largeur de fenêtre choisie par validation croisée dans le cas d'un échantillon i.i.d. :

$$\frac{EQI(\hat{h}_n)}{\inf_{h_n \in H_n} EQI(h_n)} \;\longrightarrow\; 1, \; p.co..$$

Le cas d'observations α-mélangeantes est étudié dans **(7)**. Enfin, dans **(25)** nous modifions l'estimateur à noyau ci-dessus en remplaçant au dénominateur la fonction de répartition empirique par l'estimateur à noyau de la fonction de répartition (voir également **(22)** ci-dessous). Nous montrons dans cet article que la largeur de fenêtre choisie par validation croisée est asymptotiquement optimale, résultat analogue aux résultats d'optimalité ci-dessus.

2 Estimation non paramétrique multidimensionnelle

La plupart des résultats de convergence présentés à la section 1.2 sont donnés dans un cadre multidimensionnel. Comme nous l'avons souligné plus haut, les vitesses de convergence obtenues dans le cas d'un échantillon i.i.d. sont les vitesses optimales et dépendent de la dimension d. Les travaux fondateurs de Stone (1982 et 1983) fournissent une justification théorique au phénomène du "fléau de la dimension" propre à la plupart des méthodes non paramétriques. Nous avons vu en effet, à la section 1.1 que pour un modèle non paramétrique de régression, les estimateurs de la fonction en un point x s'écrivent principalement à l'aide des observations (X_i, Y_i) avec X_i "proche" de x. La "raréfaction" des observations dans un voisinage de x pour une dimension grande conduit à prendre un voisinage de diamètre important pour conserver une proportion d'observations convenable (voir Huber, 1985, pour un vecteur uniforme dans $[0,1]^d$). Cependant en augmentant le volume du voisinage autour de x on perd sur le biais de l'estimateur et le gain sur la variance, qui dépend de la dimension d, ne compense pas cette perte (*cf.* Huber, 1985).

On le voit, ce problème joue contre les méthodes non paramétriques dans le cas de la régression par exemple où le vecteur des prédicteurs est souvent de dimension relativement importante. A partir des travaux de Huber (1985), mais également de travaux antérieurs comme ceux de Friedman et Stuetzle (1981), des auteurs se sont attachés à proposer des solutions au fléau de la dimension et le problème général de l'estimation non paramétrique multidimensionnelle a suscité un intérêt croissant dans la littérature.

Cette littérature regroupe deux aspects principaux. Tout d'abord des modèles non paramétriques permettant de réduire la dimensionalité ont été proposés : l'étude d'estimateurs pour ces modèles fait l'objet d'une littérature spécifique. Sur ce premier aspect nous avons réalisé une revue bibliographique sélective **(8)** présentant les principaux modèles de régression et les méthodes d'estimation s'y rattachant. Un autre aspect est celui du choix de modèle. Il s'agit d'introduire des méthodes de sélection de modèles : non paramétriques purs ou non paramétriques de dimensionalité inférieure par exemple. Nous discutons ci-dessous le premier problème et renvoyons aux travaux de Vieu

(1994) et Sperlich *et al.* (2000) pour le second point.

Nous avons, lors d'un séjour à l'université de Bonn (1987) réalisé un premier travail d'ordre bibliographique et informatique en collaboration avec W. Härdle. Ce travail portait sur le modèle ACE (Alternating Conditional Expectation) proposé par Breiman et Friedman (1985). Nous nous sommes par la suite intéressés principalement à deux directions que nous présentons ci-dessous. La première a trait au modèle additif de régression et fait l'objet de la thèse de C. Camlong que nous co-encadrons. La seconde réalisée en collaboration avec N.E. El Faouzi à la suite de sa thèse, porte sur l'estimation de Composantes Principales Additives.

2.1 Modèle additif de régression

Parmi les modèles non paramétriques de régression multidimensionnelle, un des plus populaires est probablement le modèle additif. Etant donné un vecteur aléatoire (X, Y) où $X = {}^t(X^1, \ldots, X^d)$ est à valeurs dans $\mathbb{R}^d$ et Y dans $\mathbb{R}$, la régression de Y sur X s'écrit :

$$r(x) \;:=\; \mathbb{E}(Y|X = x) \;=\; \sum_{j=1}^{d} r_j(x^j), \qquad x \in \mathbb{R}^d,$$

où $r_j, j = 1, \ldots, d$ sont des fonctions définies sur $\mathbb{R}$ vérifiant :

$$\mathbb{E}(r_j(X^j)) \;=\; 0,$$

condition assurant l'identifiabilité du modèle. Pour simplifier la présentation, nous n'envisageons ici que le cas où la fonction r est additive en notant que dans le cas général on étudie la meilleure approximation additive de r au sens de la norme L^2 (*cf.* Stone, 1985, pour les conditions d'existence et d'unicité de cette approximation). Dans le même ordre d'idées, nous ne présentons que la version la plus simple du modèle additif : des généralisations de ce modèle, par exemple au cas d'interactions entre covariables, sont relativement immédiates.

Il existe deux types d'estimateurs des fonctions r_j : les premiers reposent sur l'algorithme *backfitting* tandis que les seconds sont des estimateurs "directs". Dans notre travail nous avons principalement considéré les seconds (voir cependant **(8)**). L'avantage des seconds estimateurs est que leur écriture se prête plus facilement à une analyse des propriétés statistiques. En effet, peu de résultats théoriques existent pour les estimateurs reposant sur l'algorithme *backfitting* : notons cependant un travail récent de Mammen *et al.* (1999) dans lequel des propriétés de convergence sont données pour une version de cet estimateur. Dans la seconde classe, Stone (1985) propose un estimateur de type spline et obtient la vitesse de convergence en norme L^2 pour cet estimateur, vitesse optimale pour un estimateur non paramétrique en dimension 1. Ce résultat montre que l'objectif de réduction de la dimensionalité est atteint lorsqu'on considère le modèle additif. Pour le même type d'estimateurs, Stone (1994) a étudié les vitesses de convergence pour le modèle d'interactions dans plusieurs problèmes d'estimation fonctionnelle.

Vers le milieu des années quatre-vingt-dix d'autres estimateurs directs des fonctions r_j ont été proposés. Ces estimateurs reposent sur l'estimation non paramétrique "globale" de la fonction r (par la méthode du noyau) puis de l'intégration marginale de cet estimateur (*cf.* par exemple Tjøstheim et Auestad, 1994 ou Linton et Nielsen, 1995). De manière plus formelle, étant donnée une densité $q(x) = \Pi_{j=1}^d q_j(x^j)$ sur $\mathbb{R}^d$, on peut écrire $r(x)$ sous la forme :

$$r(x) \;=\; \sum_{j=1}^d \eta_j(x^j) \;+\; \int r(x)q(x)dx,$$

où

$$\begin{aligned} \eta_j(x^j) \;&=\; r_j(x^j) \;-\; \int r_j(x^j)q_j(x^j)dx_j \\ &= \int r(x)q_{-j}(\bar{x}^j)d\bar{x}^j \;-\; \int r(x)q(x)dx, \end{aligned}$$

avec $q_{-j} = \Pi_{l\neq j}q_l$ et $\bar{x}^j = (x^1, \ldots, x^{j-1}, x^{j+1}, \ldots, x^d)$. A partir de n couples (X_i, Y_i) de même loi que (X, Y), l'estimateur de $\eta_j(x^j)$ est alors défini par :

$$\eta_{n,j}(x^j) \;=\; \int r_n(x)q_{-j}(\bar{x}^j)d\bar{x}^j \;-\; \int r_n(x)q(x)dx,$$

où r_n est l'estimateur à noyau (dans $\mathbb{R}^d$) défini plus haut.

Nous nous sommes intéressés aux propriétés asymptotiques de l'estimateur ci-dessus. Ce travail a été réalisé en collaboration avec C. Camlong et P.

Vieu et il constitue une partie du travail de thèse de C. Camlong. Dans **(3)**, nous établissons la normalité asymptotique de l'estimateur η_j sous une hypothèse de β-mélange : nous appliquons ensuite le résultat à la prédiction de séries temporelles. Les techniques utilisées dans la preuve sont sensiblement différentes de celles utilisées dans les articles présentées à la section 1.2. Ces techniques reposent sur :

- une inégalité de moments due à Yoshihara (1984) ;
- le Lemme de Berbee (*cf.* Doukhan, 1994) ;
- un découpage en petits et grands "blocs".

Nous obtenons alors une vitesse de convergence qui est celle de la dimension un (et identique au cas i.i.d.). Notons que notre résultat est donné dans le cas où la densité de X_i est supposée connue mais que sa généralisation est immédiate dans le cas où cette densité est estimée par l'estimateur à noyau comme C. Camlong l'a établi dans sa thèse.

Pour conclure cette section, nous notons que nous ne nous sommes pas directement intéressés dans notre travail au problème du choix de modèle. Cependant, ce problème a été abordé au cours de la thèse de C. Camlong qui propose un test de l'hypothèse nulle H_0 : "la fonction r est additive".

2.2 Estimation de Composantes Principales Additives

Dans un modèle de régression linéaire, un problème important est celui de la recherche de *colinéarité* entre les covariables, c'est-à-dire de relations du type $\sum_{j=1}^{d} a_j X^j = 0$. Dans le cadre du modèle de régression additive ci-dessus, l'analogue de cette colinéarité se traduit par l'existence de fonctions $\phi_j, j = 1, \ldots, d$, telles que :

$$\sum_{j=1}^{d} \phi_j(X^j) \;=\; 0.$$

Dans le cas d'existence d'une telle relation, cela conduit à une instabilité dans le modèle additif puisqu'on a alors :

$$\sum_{j=1}^{d} \psi_j(X^j) \;=\; \sum_{j=1}^{d} (\psi_j + \epsilon\phi_j)(X^j).$$

De manière générale, il est utile en pratique de chercher la fonction additive du vecteur X (si elle existe) de variance minimum, c'est-à-dire pour laquelle la contrainte de nullité ci-dessus est aussi "vraie" que possible. Ce problème a été étudié en particulier par Donnell *et al.* (1993) qui cherchent des dépendances additives entre les covariables pour les données de concentration d'ozone déjà étudiées par Breiman et Friedman (1985). Nous donnons tout d'abord une description plus formelle de ce problème.

Nous supposons que $X = {}^t(X^1, \ldots, X^d)$ est un vecteur aléatoire à valeurs dans $[0,1]^d$. Soit pour $j = 1, \ldots, d$, H_j l'ensemble des fonctions ϕ_j telles que $I\!\!E\phi_j(X^j) = 0$ et $Var\phi_j(X^j) < \infty$. La plus petite Composante Principale Additive (CPA), lorsqu'elle existe, est définie comme le vecteur $\Phi = (\phi_1, \ldots, \phi_d)$ tel que $\phi_j \in H_j, j = 1, \ldots, d$ et :

$$Var[\tilde{\phi}(X)] \;=\; min! \text{ sous la contrainte } \sum_{j=1}^{d} Var\phi_j(X^j) \;=\; 1,$$

où

$$\tilde{\phi}(X) \;=\; \sum_{j=1}^{d} \phi_j(X^j).$$

On définit une $r^{\text{ème}}$ plus petite Composante Principale Additive de manière identique en ajoutant une contrainte d'orthogonalité (*cf.* Donnell *et al.*, 1993). Il est assez facile de montrer que la $r^{\text{ème}}$ plus petite composante principale additive est une fonction propre associée à la $r^{\text{ème}}$ plus petite valeur propre de l'opérateur P de $H = H_1 \times \ldots \times H_d$ dans lui-même, défini par ses composantes $[P\Phi]_j = P_j(\sum_{j'=1}^{d} \phi_{j'})$, où P_j est la projection orthogonale sur H_j. Nous donnons dans **(27)** une condition suffisante d'existence d'une $r^{\text{ème}}$ plus petite composante principale additive.

Bien que l'estimation de la plus petite composante principale additive soit plus particulièrement intéressante du point de vue des applications, nous nous sommes intéressés dans **(27)** au cas général de l'estimation d'une $r^{\text{ème}}$ plus petite CPA. Les estimateurs introduits reposent sur les fonctions splines. Considérons tout d'abord un échantillon $(X_i, Y_i), i = 1, \ldots, n$ de couples i.i.d. et étant donné $k+1$ noeuds équidistants sur $[0,1]$, l'ensemble $S_{k,q}$ des fonctions splines de degré q sur $[0,1]$, c'est-à-dire des fonctions $(q-1)$ fois dérivables sur $[0,1]$ et polynomiale de degré q entre deux noeuds. On définit l'estimateur spline de la $r^{\text{ème}}$ plus petite composante principale additive $\hat{\Phi}_r =$

$(\hat{\phi}_{r,1},\ldots,\hat{\phi}_{r,d})$ avec $\hat{\phi}_{r,j} \in S_{k,q}$ par :

$$n^{-1}\sum_{i=1}^{n}\left[\sum_{j=1}^{d}\hat{\phi}_{r,j}(X_i^j)\right]^2 \;=\; \min!,$$

sous les contraintes :

$$n^{-1}\sum_{i=1}^{n}\hat{\phi}_{r,j}(X_i^j) \;=\; 0, \qquad j=1,\ldots,d,$$

et

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{d}\hat{\phi}_{r,j}(X_i^j)\hat{\phi}_{l,j}(X_i^j) \;=\; \delta_{r,l}, \qquad l=1,\ldots,r.$$

Nous montrons que sous certaines hypothèses de régularité (*cf.* **(27)**), il existe une solution unique à ce problème de minimisation sur un ensemble dont la probabilité tend vers 1 lorsque n tend vers l'infini. Enfin nous obtenons pour cet estimateur une borne supérieure de la vitesse de convergence en norme L^2, c'est-à-dire $n^{-\beta/(2\beta+1)}$ où $\beta = m+\alpha$, m étant un entier positif et $0<\alpha\leq 1$ et les composantes principales additives étant supposées posséder une dérivée d'ordre m vérifiant la condition de Lipschitz d'ordre α. Cette vitesse de convergence est la vitesse optimale donnée par Stone (1982 et 1983) dans le cas de l'estimation de la régression ou de la densité et pour la dimension 1.

3 Modèle linéaire fonctionnel

Le travail présenté dans cette section a été réalisé en collaboration avec H. Cardot et F. Ferraty et pour la dernière partie avec A. Mas.

On trouve dans la littérature de nombreux exemples d'applications concernant des variables fonctionnelles : en chimie quantitative par exemple où on souhaite expliquer une variable chimique par une courbe spectrométrique (*cf.* Frank et Friedman, 1993). Nous renvoyons à Hastie *et al.* (1995), Hastie et Mallows (1993) et Ramsay et Silverman (1997) pour d'autres exemples en linguistique, météorologie, *etc*, ...

Frank et Friedman (1993) comparent les différentes méthodes paramétriques utilisées par les praticiens dans le contexte de la régression avec variable explicative fonctionnelle : Ridge Regression, Régression sur Composantes Principales et Régression PLS. L'idée de ces méthodes est de considérer les points de discrétisation des courbes explicatives comme les variables explicatives d'un modèle de régression. Frank et Friedman montrent que ces méthodes donnent des résultats satisfaisants dans un certain nombre de cas, en chimie quantitative notamment. Pourtant, comme le soulignent Hastie et Mallows (1993) dans leur commentaire et Frank et Friedman dans leur réponse, ces méthodes paramétriques échouent souvent lorsque les courbes explicatives ne sont pas lisses ou sont elles-mêmes mesurées avec une erreur. Par ailleurs, elles sont difficiles à mettre en oeuvre lorsque les courbes ne sont pas discrétisées aux mêmes points.

Plusieurs auteurs ont proposé des modèles de régression prenant en compte la nature fonctionnelle de la courbe explicative. Citons Hastie et Mallows (1993), Ramsay et Silverman (1997) et Marx et Eilers (1999). Une étude bibliographique préliminaire nous a montré que les modèles fonctionnels de régression sont l'objet d'un intérêt croissant dans la communauté des Statisticiens. Pour autant les résultats théoriques concernant les méthodes proposées dans ce contexte sont relativement rares. Dans le contexte de l'ACP pour variables à valeurs dans un espace de Banach ou de Hilbert, citons cependant les travaux de Dauxois *et al* (1982), Antoniadis et Beder (1989) et Silverman (1996) qui établissent des résultats de convergence pour des estimateurs de l'opérateur de covariance. Notons également l'article de Bosq (1991) qui s'intéresse à l'estimation de l'opérateur d'autocorrélation pour un processus autorégressif hilbertien. Une partie du travail présenté ci-dessous reprend les résultats contenus dans ce dernier article dans le cas de l'estimation de la régression.

Fixons le cadre général dans lequel X est une fonction aléatoire définie sur sur un compact de $\mathbb{R}$ (nous prenons $[0,1]$ dans la suite) et à valeurs dans $\mathbb{R}$ telle que $\mathbb{E}\int_0^1 X^2(t)dt < \infty$ et Y est une variable aléatoire réelle. Le lien entre X et Y est modélisé par :

$$Y \;=\; \Psi(X) \,+\, \varepsilon,$$

où Ψ est un opérateur linéaire continu défini sur l'ensemble H des fonctions définies sur $[0,1]$ de carré intégrable et à valeurs dans $\mathbb{R}$ (c'est-à-dire Ψ est un élément du dual topologique H' de l'espace de Hilbert H) et ε est une variable aléatoire indépendante de X, d'espérance nulle et de variance σ^2.

Nous ne présentons ici que le cas particulier où $H = L^2([0,1])$ en notant qu'une partie de nos résultats se généralisent aisément au cas où H est un espace de Hilbert réel séparable quelconque. Par ailleurs le Théorème de représentation de Riesz nous permet de réécrire le modèle ci-dessus de la manière suivante :

$$Y \;=\; \int_0^1 \alpha(t)X(t)dt \,+\, \varepsilon,$$

où $\alpha \in H$.

Suivant les travaux de Bosq (1991), nous avons proposé dans **(28)** un premier estimateur de l'opérateur Ψ, obtenu en réalisant une régression sur composantes principales fonctionnelles. En effet, en prenant dans H le produit scalaire et la norme usuels, l'opérateur de covariance de X est l'opérateur de H dans H défini par :

$$\Gamma x \;=\; \mathbb{E}(< X, x > X), \qquad x \in H,$$

et l'opérateur de covariance croisée entre X et Y est donné par :

$$\Delta x \;=\; \mathbb{E}(< X, x > Y), \qquad x \in H.$$

Il est alors facile de montrer que l'on a la relation :

$$\Delta \;=\; \Psi\Gamma.$$

Cependant l'estimation de l'inverse de l'opérateur Γ lorsqu'il existe est problématique puisque cet inverse n'est pas borné (rappelons que Γ est un opérateur autoadjoint, positif et nucléaire). A partir des versions empiriques Γ_n et Δ_n de Γ et de Δ basées sur un échantillon $(X_i, Y_i), i = 1, \ldots, n$, nous avons adapté l'estimateur proposé par Bosq (1991) dans le cas de la prédiction pour une série chronologique à temps continu. Son principe repose sur l'analyse spectrale de l'opérateur Γ_n puis sur l'inversion de Γ_n dans l'espace engendré par les fonctions propres associées aux k_n plus grandes valeurs propres de Γ_n. Nous obtenons dans **(28)** la convergence en probabilité et la convergence presque sûre de cet estimateur. Les preuves plus détaillées de ces résultats se trouvent dans **(9)**. Les vitesses de convergence sont précisées dans le cas où les valeurs propres de Γ sont de la forme $\lambda_j = ar^j$, $a > 0, r \in]0,1[$ ou $\lambda_j = aj^{-\gamma}$, $a > 0, \gamma > 1$. Les démonstrations s'appuient sur des résultats asymptotiques pour les opérateurs de covariance et covariance croisée empiriques (*cf.* Dauxois *et al.*, 1982 et Bosq, 1991) ainsi qu'une version générale de l'inégalité de Bernstein due à Yurinskiĭ (1976).

Nous avons également réalisé dans **(28)** une étude par simulations montrant que cet estimateur donne de bons résultats mais que cependant, pour les petites tailles d'échantillon le coefficient fonctionnel estimé n'est pas suffisamment lisse. Dans **(2)** et **(10)** nous avons d'une part modifié l'estimateur ci-dessus en introduisant un lissage spline préliminaire sur les courbes explicatives. Les résultats par simulation obtenus dans **(2)** montrent le gain du point de vue de la régularité du coefficient fonctionnel tandis que les résultats théoriques ne sont pas affectés par cette modification. Par ailleurs, nous avons proposé dans ces mêmes articles, un estimateur alternatif reposant sur un lissage spline du coefficient fonctionnel. C'est ce dernier que nous décrivons ci-dessous.

Etant donné un vecteur $(B_{k,1}, \ldots, B_{k,k+q})'$ de B-splines engendrant l'espace $S_{k,q}$ des fonctions splines de degré q avec $k+1$ noeuds définies sur $[0,1]$, l'estimateur $\hat{\alpha}$ est défini par :

$$\hat{\alpha} = \sum_{j=1}^{k+q} \hat{\theta}_j B_{k,j},$$

où $(\hat{\theta}_1, \ldots, \hat{\theta}_{k+q})'$ est un vecteur de coefficients minimisant :

$$n^{-1} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{k+q} < \theta_j B_{k,j}, X_i > \right)^2 \; + \; \rho \left\| \sum_{j=1}^{k+q} \theta_j B_{k,j}^{(m)} \right\|^2 ,$$

et où ρ est un paramètre de régularisation.

Dans l'article **(2)**, nous obtenons la vitesse de convergence en norme L^2 de l'estimateur $\hat{\alpha}$ ainsi que celle de l'estimateur obtenu par régression sur composantes principales lisses ci-dessus. Nous comparons les deux estimateurs à l'aide d'une étude par simulations. Comme nous l'avons souligné plus haut, peu de résultats asymptotiques existent dans le contexte de modèles de régression avec variables fonctionnelles et les résultats obtenus dans **(2)** et **(10)** sont, à notre connaissance, parmi les premiers. Citons également le récent article de Ferraty et Vieu (2000) qui obtiennent des résultats de convergence pour un estimateur de type noyau dans le cadre d'un modèle de régression non paramétrique.

Enfin, dans un travail récent **(1)**, nous nous sommes intéressés au problème du test de l'hypothèse de nullité de l'opérateur Ψ (ou de de manière équivalente de la fonction α). Nous proposons deux statistiques de test reposant sur la norme de l'opérateur Δ. Nous montrons la normalité asymptotique de la première tandis que la seconde est approximée par un χ^2. Nous avons réalisé une étude par simulations du niveau et de la puissance donnant des résultats équivalents et relativement satisfaisants pour les deux procédures de test.

4 Autres problèmes étudiés.

Nous avons regroupé dans ce chapitre des travaux qui ne se rattachent pas directement aux travaux précédents. Le premier a trait à l'étude d'estimateurs reposant sur un ajustement linéaire local. Le deuxième a trait à l'estimation non paramétrique de points de discontinuités.

4.1 Méthodes reposant sur un ajustement polynomial local

Nous nous sommes intéressés, à l'occasion d'une collaboration avec M. Lejeune, à des estimateurs non paramétriques reposant sur un ajustement polynomial local. Ce travail a tout d'abord porté sur l'estimation de quantiles conditionnels puis sur l'estimation d'une densité et d'une fonction de répartition.

L'estimateur de Nadaraya-Watson de la régression peut être vu comme la solution du problème de minimisation suivant (pour la dimension $d = 1$) :

$$r_n(x) \; = \; \arg\min_a \sum_{i=1}^{n} (Y_i - a)^2 \, h_n^{-1} K\left(\frac{x - X_i}{h_n}\right).$$

Suivant les travaux de Cleveland (1979), certains auteurs se sont intéressés à l'estimateur à noyau basé sur un ajustement polynomial local, c'est-à-dire défini par le problème de minimisation :

$$\min_{(a_0,\ldots,a_k)} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{k} a_k (x - X_i)^j\right)^2 h_n^{-1} K\left(\frac{x - X_i}{h_n}\right).$$

On pose alors :

$$r_n(x) \; = \; \hat{a}_0.$$

La généralisation à une dimension d quelconque est immédiate. Les estimateurs reposant sur un ajustement linéaire local sont les plus utilisés en pratique. Ces derniers possèdent plusieurs avantages par rapport à l'estimateur de Nadaraya-Watson, dont le fait qu'ils ne sont pas soumis aux effets de bord. Par ailleurs l'expression asymptotique du biais est plus simple puisqu'en particulier la dérivée seconde de la densité marginale n'entre pas dans cette expression. Nous renvoyons à l'ouvrage de Fan et Gijbels (1996) pour les détails des propriétés de ces estimateurs.

Dans l'article **(17)**, nous nous sommes intéressés à l'estimation de quantiles conditionnels à l'aide d'une version modifiée de l'estimateur de régression linéaire locale. Soit (X, Y) un couple de variables aléatoires et soit F_x la fonction de répartition conditionnelle de Y sachant $X = x$, le quantile conditionnel d'ordre $p \in [0, 1]$ au point x est défini par :

$$r_p(x) \;=\; F_x^{-1}(p).$$

A partir d'un échantillon $(X_i, Y_i), i = 1, \ldots, n$ de variables indépendantes et identiquement distribuées, l'estimateur de $r_p(x)$ est défini de la manière suivante. Soit $(\hat{\beta}_0, \ldots, \hat{\beta}_k)$ solution du problème de minimisation :

$$\min_{(\beta_0,\ldots,\beta_k)} \sum_{i=1}^{n} \delta(p) \left| Y_i - \sum_{j=0}^{k} \beta_j (x - X_i)^j \right| h_n^{-1} K\left(\frac{x - X_i}{h_n}\right),$$

où :

$$\delta(p) \;=\; \begin{cases} p, & \text{si } \; Y_i - \sum_{j=0}^{k} \beta_j (x - X_i)^j \geq 0 \\ 1 - p & \text{si } \; Y_i - \sum_{j=0}^{k} \beta_j (x - X_i)^j < 0. \end{cases}$$

L'estimateur de $r_p(x)$ est alors défini par $\hat{\beta}_0$. Nous traitons dans **(17)** le cas $k = 2$ et proposons un algorithme de résolution du problème de minimisation ci-dessus. Nous illustrons la méthode à l'aide d'une étude par simulations. Nous renvoyons à Fan *et al.* (1994) pour la démonstration de la normalité asymptotique de cet estimateur. La procédure d'estimation définie ci-dessus peut être qualifiée de "directe" : d'autres estimateurs de quantiles conditionnels de ce type ont été introduits dans la littérature. Citons par exemple He et Shi (1994) qui utilisent des fonctions splines de régression. Des estimateurs "indirects" construits par inversion d'un estimateur de la fonction de répartition conditionnelle ont également été étudiés : voir entre autres Stone (1977), Bhattacharya et Gangopadhyay (1990) et Berlinet *et al.* (1998).

Dans **(22)** nous utilisons le principe d'ajustement polynomial local pour le lissage de la fonction de répartition empirique. L'estimateur obtenu est la fonction cumulée d'un estimateur à noyau de la densité. Par ailleurs, la procédure mise en oeuvre permet d'obtenir l'expression de noyaux d'ordre supérieur, produit d'une densité (le noyau initial) et d'un noyau reproduisant polynomial (rappelons que le noyau est d'ordre k lorsque ses moments d'ordre 1 à $k-1$ s'annulent et son moment d'ordre k est non nul). Ces résultats particularisent des résultats établis par Berlinet (1991) dans le cadre général d'un espace de Hilbert à noyau reproduisant. Dans cet article, l'auteur s'intéresse à la recherche de noyaux d'ordre supérieur et propose une procédure de construction de tels noyaux ainsi qu'une méthode de choix de l'ordre et du paramètre de lissage. Notons que l'utilisation de noyaux d'ordre supérieur (pour l'estimation de la densité par exemple) permet une réduction du biais au prix d'une augmentation de la variance. En pratique se pose donc le problème du choix de l'ordre, problème traité également par Hall et Marron (1987). Dans notre travail nous ne nous intéressons pas directement à cet aspect mais à l'étude de l'estimateur à noyau de la fonction de répartition pour lequel nous donnons les expressions asymptotiques du biais et de la variance.

4.2 Estimation de points de discontinuité

Ce travail a porté initialement sur l'étude d'un problème météorologique posé par J.L. Brenguier du Centre National Météorologique de Toulouse. Le but est de déterminer l'évolution de la concentration des particules d'eau dans un nuage et d'étudier son effet sur les développements des précipitations et les propriétés radiatives du nuage. On dispose d'une série de temps d'arrivée de gouttes d'eau mesuré par un spectromètre, le long de la trajectoire rectiligne d'un avion. Dans un premier travail **(24)**, qui a donné lieu au mémoire de DEA de S. Moncaup, nous nous sommes intéressés à l'estimation non paramétrique de l'intensité d'un processus.

De manière plus formelle, on dispose de N points X_i sur un intervalle de temps $[0, T]$ que l'on suppose être la réalisation d'un processus de Poisson doublement stochastique de processus d'intensité $\{\Lambda(x), x \in \mathbb{R}\}$:

- $\{\Lambda(x), x \in \mathbb{R}\}$ est un processus stationnaire, à valeurs positives, de moyenne μ ;
- Conditionnellement à la réalisation $\lambda(x)$ de $\Lambda(x)$, les points X_i sont générés selon un processus de Poisson non homogène, de fonction d'intensité $\lambda(x)$.

Nous avons étudié l'estimateur à noyau (Diggle, 1985) de la fonction λ défini par :

$$\lambda_n(x) \;=\; \sum_{i=1}^{N} \frac{1}{h_n} K\left(\frac{x - X_i}{h_n}\right), \qquad x \in [0,T],$$

où K est un noyau et h_n le paramètre de lissage. Par ailleurs, nous modifions la procédure de validation croisée de Brooks et Marron (1990) afin de choisir un paramètre de lissage local. Le critère proposé s'écrit :

$$CV_x(h) \;=\; \int_0^T \lambda_n^2(y) 1\!\!1_{[x-\epsilon,x+\epsilon]}(y)dy$$

$$- \sum_{i=1}^{N} \sum_{j \neq i} h^{-1} K\left(\frac{X_j - X_i}{h}\right) 1\!\!1_{[x-\epsilon,x+\epsilon]}(X_i),$$

où $\epsilon > 0$. Le paramètre de lissage est choisi en minimisant le critère ci-dessus. Nous comparons ensuite la méthode globale de Brooks et Marron à cette méthode locale sur les données météorologiques.

A partir de cette première étude, nous nous sommes intéressés au problème de l'estimation de points de discontinuités. En effet, l'étude précédente a mis en évidence des ruptures dans la structure du processus de gouttes d'eau. Ce deuxième travail a donné lieu à la thèse de V. Couallier soutenue le 9 juin 2000 et que nous avons co-encadrée. Dans **(26)**, nous proposons un estimateur d'un point de discontinuité (et de la hauteur du saut) pour l'intensité d'un processus de Poisson non homogène. Celui-ci repose sur la maximisation de l'écart entre deux estimateurs à noyaux construits à l'aide de noyaux dissymétriques à gauche et à droite respectivement (*cf.* Wu et Chu, 1993, dans le contexte de la régression).

Etant donnés deux noyaux K_1 et K_2 à support $[-1,w]$ et $[-w,1]$ respectivement, l'estimateur d'une discontinuité θ dans la fonction λ est défini par :

$$\theta_n \;=\; \arg\max_{x\in[0,T]} |S_n(x)|\,,$$

où

$$S_n(x) \;=\; \frac{\lambda_{n,1}(x) - \lambda_{n,2}(x)}{k_w},$$

$\lambda_{n,1}$ et $\lambda_{n,2}$ sont les estimateurs à noyau construits à l'aide des noyaux K_1 et K_2 respectivement et où k_w est une constante dépendant de K_1 et K_2. La hauteur du saut est estimé par $S_n(x)$. Nous donnons l'expression de l'erreur quadratique moyenne pour l'estimateur du saut et étudions la qualité de l'estimateur θ_n à partir de simulations ainsi que sur les données réelles. Des résultats de convergence analogues à ceux établis dans **(26)** ont été montrés dans le contexte de la régression par Wu et Chu (1993).

5 Projet de recherche

Nous consacrons cette dernière partie aux prolongements possibles de nos travaux et aux perspectives de recherche.

De manière générale, les travaux que nous avons présentés dans ce document de synthèse relèvent pour une bonne part, comme on a pu le voir, de la Statistique Mathématique. Pour autant, les aspects appliqués de la Statistique n'en sont pas absents : qu'il s'agisse des problèmes de choix du paramètre de lissage (section 1.3), de problèmes d'estimation multidimensionnelle (section 2) ou encore de l'étude du modèle linéaire fonctionnel (section 3), les travaux que nous avons réalisés trouvent le plus souvent des motivations et des prolongements dans la pratique (voir par exemple la section 4.2). Par ailleurs, la mise en oeuvre de méthodes d'estimation non paramétrique (au travers d'études concrètes par exemple), mais également leur enseignement, nous paraissent être primordiales. Ces derniers aspects n'ont pas été développés dans ce document, qui s'attache principalement à présenter nos travaux publiés, mais notons, toutefois, qu'ils sont présents dans notre travail et que nous avons réalisé plusieurs études de cas.

C'est dans cet esprit, c'est-à-dire la mise en perspective de résultats théoriques au travers d'applications des méthodes étudiées, que nous souhaitons poursuivre notre recherche.

Axes de recherche

Il est clair que les divers travaux présentés dans ce document pourraient donner lieu à des prolongements, qu'il s'agisse par exemple du problème du choix du paramètre de lissage, qui reste encore d'actualité (voir à ce sujet les articles récents de Biau, 1999 et de Loader, 1999), ou encore de l'étude des propriétés théoriques des estimateurs des quantiles conditionnels présentés en 4.1. Cependant, nous avons choisi de nous en tenir ci-dessous aux pistes de recherche que nous souhaitons privilégier.

1. Parmi les différents thèmes décrits dans ce document, l'étude du modèle linéaire fonctionnel, et de manière plus générale de modèles pour variables fonctionnelles, devrait plus particulièrement retenir notre attention. Comme nous l'avons souligné plus haut, les problèmes liés à cette étude suscitent un grand intérêt dans la communauté des Statisticiens et ce thème devrait être, à notre avis, parmi les axes de recherche

privilégiés en matière d'estimation fonctionnelle. Il concerne de nombreuses applications comme en témoigne le récent ouvrage de Ramsay et Silverman (1997) (voir également les travaux cités à la section 3). La bibliographie sur le sujet reste relativement modeste et un grand nombre de problèmes ouverts demeurent.

A court terme, il nous semble nécessaire de pousser plus loin l'étude des propriétés des estimateurs : expressions du biais et de la variance, normalité asymptotique, *etc.* La difficulté tient au fait que l'on traite des variables à valeurs dans un espace de Hilbert de dimension infinie et des techniques de démonstrations généralement utilisées en estimation non paramétrique ne s'appliquent pas nécessairement ici. Par ailleurs le rôle et le choix du nombre de noeuds et ceux du paramètre de régularisation dans l'estimateur spline devraient être étudiés plus en détail. Les simulations que nous avons réalisées semblent indiquer que le second est primordial (pour un nombre de noeuds assez grand). La validation croisée a donné également de bons résultats pour le choix du paramètre de régularisation (*cf.* (**2**)).

D'autres problèmes devraient encore être appréhendés. Parmi eux citons l'étude du modèle linéaire fonctionnel pour des variables dépendantes : les travaux réalisés dans le contexte de la régression non paramétrique devraient trouver des prolongements dans le cas du modèle linéaire fonctionnel. Par ailleurs, nous envisageons d'étudier le problème de l'estimation de quantiles conditionnels lorsque la variable explicative est une courbe : dans ce sens, un estimateur spline minimisant un critère du type de celui de la section 4.1 devrait être introduit. L'application de tests de permutations ou encore des tests basés sur des techniques de rééchantillonnages devrait fournir une alternative aux procédures de tests proposées dans **(1)** (*cf.* par exemple Bowman et Azzalini, 1997, pour l'utilisation de tests de permutations dans le cas de modèles de régression non paramétrique). Des extensions du modèle linéaire fonctionnel et des comparaisons avec d'autres modèles devraient être également envisagées. Nous pensons en particulier à la comparaison avec les modèles paramétriques (comme la Ridge Regression) ou encore avec un modèle non paramétrique comme celui introduit par Ferraty et Vieu (2000). Enfin, un projet de comparaison du modèle linéaire fonctionnel avec des modèles paramétriques dans le contexte de "panels" de séries chronologiques devrait être mené dans un proche futur.

2. Les travaux présentés dans la 3$^{\text{ème}}$ partie (et notamment ceux concernant le modèle additif), devraient donner lieu à des recherches futures. Nous pensons notamment à l'étude d'estimateurs combinant l'estimateur d'intégration marginale et celui reposant sur l'algorithme *backfitting* : l'estimateur d'intégration marginale est utilisé lors de l'étape initiale de l'algorithme et une seule itération est ensuite réalisée. Des travaux prometteurs ont été réalisé en ce sens par Mammen *et al.* (1999) qui obtiennent des propriétés asymptotiques pour cet estimateur. Ces résultats mettent à jour le gain obtenu en terme de réduction de la dimensionalité. Le cas de variables dépendantes et l'application aux séries chronologiques, mais aussi des applications à d'autres modèles de régression, peuvent être envisagées.

3. Un autre sujet sur lequel nous avons amorcé une collaboration avec M. Goulard, chercheur à l'INRA, est celui de l'utilisation de méthodes bayésiennes en estimation non paramétrique. Plusieurs travaux ont paru dans la littérature portant en particulier sur les estimateurs splines. L'*a priori* porte sur la distribution des noeuds. L'intérêt de la méthode est de fournir une solution au problème du choix du nombre et de la position des noeuds, c'est-à-dire au problème du choix du paramètre de lissage (voir à ce sujet Denison *et al.*, 1998). L'étude que nous avons réalisée concerne l'estimation de l'intensité d'un processus de Poisson non homogène avec variable concomitante et son application à la répartition d'un peuplement d'arbres en fonction de l'altitude. Ce travail devrait se poursuivre par l'étude plus circonstanciée des méthodes utilisées.

Projets de collaborations

Comme nous l'avons souligné tout au long de ce document, les travaux que nous avons réalisés ont donné lieu à de nombreuses collaborations au sein de notre laboratoire (que ce soit à l'Université Montpellier II ou celle de Toulouse III), mais également avec des chercheurs de laboratoires extérieurs. D'autres projets de collaborations, pour lesquels des contacts ont d'ores et déjà eu lieu, devraient se concrétiser dans le futur.

Concernant le premier axe de recherche ci-dessus, nous collaborons d'une part avec des chercheurs toulousains et particulièrement avec H. Cardot (INRA), F. Ferraty (Université Toulouse II), et P. Vieu (Université Toulouse III). Nous co-encadrons avec ce dernier le travail de thèse de A. Goia qui travaille actuellement sur des modèles pour variables fonctionnelles pour lesquels il s'attache à établir des résultats dans le cas d'observations dépendantes. Une collaboration avec S. Guillas (Université Paris 6) a également été amorcée. Ce projet porte sur l'estimation de quantiles conditionnels (voir ci-dessus) et son application à l'estimation de pics de pollution. L'étude de "panels" de séries chronologiques évoquée ci-dessus doit donner lieu à un travail en collaboration avec D. Tjøstheim (Université de Bergen). Notons enfin que ce thème regroupe plusieurs aspects de la Statistique et du Calcul des Probabilités (estimation non-paramétrique, statistique multidimensionnelle, environnement fonctionnel de la statistique ou encore calcul des probabilités dans les espaces de Hilbert) et que nous pensons que des collaborations avec d'autres chercheurs devraient se nouer. C'est dans ce sens que nous animons, avec H. Cardot, F. Ferraty, Y. Romain et P. Vieu le groupe de travail "Statistique Fonctionnelle", STAPH, dont un des buts est de creuser les liens entre estimation non paramétrique et environnement fonctionnel de la statistique (*cf.* Romain, 1997, concernant ce dernier aspect).

D'autres thèmes de recherche devraient également donner lieu à des collaborations. Outre le travail amorcé avec M. Goulard évoqué ci-dessus (point 3), une visite à l'Université de Prague sur l'invitation de J. Antoch est prévue afin de poser les bases d'un travail en commun concernant l'estimation de points de discontinuités.

Références

Antoniadis, A. and Beder, J. (1989). Joint estimation of the mean and the covariance of a Banach valued gaussian vector. *Statistics*, **20**, 73-93.

Berlinet, A. (1991). Reproducing kernels and finite order kernels. In : Roussas, G (Ed.), Nonparametric Functional Estimation and Related Topics, NATO, ASI Series, 3-19.

Berlinet, A. et Dvroye, L. (1994). A comparison of kernel density estimates. *Pub. de l'Institut de Statistique de l'Université de Paris*, **38**, 3-59.

Berlinet, A., Gannoun, A. et Matzner-Lober, E. (1998). Propriétés asymptotiques d'estimateurs convergents des quantiles conditionnels. *C. R. Acad. Sci.*, Paris, Série 1, 611-614.

Battacharya, P.K. and Gangopadhyay, A.K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.*, **18**, 1400-1415.

Biau, G. (1999). Estimateurs à noyaux itérés : synthèse bibliographique. *Journal de la Société Française de Statistique*, **140**, 41-67.

Bosq, D. (1991). Modelization, non-parametric estimation and prediction for continuous time processes. In : Roussas, G (Ed.), Nonparametric Functional Estimation and Related Topics, NATO, ASI Series, 509-529.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes, Estimation and Prediction.* Lecture Notes in Statistics, **110**, Springer-Verlag.

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

Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353-360.

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis.* Oxford University Press, Oxford, UK.

Breiman, L. and Friedman, J.H. (1985). Estimating Optimal Transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.*, **80**, 580-619.

Brooks, M.M. and Marron, J.S. (1990). Asymptotic optimality of the lesat-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stoch. Proc. and their Appl.*, **38**, 157-165.

Cao, R., Cuevas, A. et Gonzalez-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Comp. Statist. and Data Analysis*, **17**, 153-176.

Carbon, M. (1983). Inégalité de Bernstein pour les processus fortement mélangeants non nécessairement stationnaires. *C. R. Acad. Sciences, Série I*, **297**, 303-306.

Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothnig Scatterplots. *J. Amer. Statist. Assoc.*, **74**, 829-836.

Collomb, G. (1976). Estimation non-paramétrique de la régression par la méthode du noyau. Thèse, Université Paul Sabatier, Toulouse.

Collomb, G. (1981). Estimation non-paramétrique de la régression : revue bibliographique. *Intern. Statist. Reviews*, **49**, 75-93.

Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahr. und verw. Geb.*, **66**, 441-460.

Collomb, G. (1985). Nonparametric regression : an up-to-date bibliography. *Statistics*, **2**, 309-324.

Collomb, G. and Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis. *Stoch. Proc. Appl.*, **23**, 77-89.

Dauxois, J., Pousse, A. et Romain, Y.(1982). Asymptotic theory for the principal component analysis of a random vector function : some applications to statistical inference. *J. Mult. Analysis*, **12**, 136-154.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998). Automatic Bayesian curve fitting. *J. R. Statist. Soc. B*, **60**, 333-350.

Devroye, L. (1987). *A course on density estimation.* Birkhauser.

Devroye, L. and Wagner, T.J. (1979). The L_1 convergence of kernel density estimates. *Ann. Statist.*, **7**, 1136-1139.

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

Diggle, P. (1985). A Kernel Method for Smoothnig Point Process Data. *Appl. Statist.*, **34**, 793-800.

Donnell, D.J., Buja, A. and Stuetzle, W. (1993). Analysis of Additive dependencies and concurvities using smallest principal components (with discussion). *Ann. Statist.*, **22**, 1635-1673.

Doukhan, P. (1994). *Mixing, Properties and Examples.* Lecture Notes in Statistics, **85**, Springer-Verlag.

Fan, J., Hu, T.C and Truong, Y.K. (1994). Robust nonparametric function estimation. *Scand. J. Statist.*, **21**, 433-446.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* Chapman & Hall, London.

Ferraty, F. et Vieu, Ph. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *Comptes Rendus de l'Académie des Sciences*, **330**, Série I, 139-142.

Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chemometric regression tools (with doiscussion). *Technometrics*, **35**, 109-148.

Friedman, J.H. and Stuetzle, W. (1993). Projection Pursuit Regression. *J. Amer. Statist. Assoc.*, **76**, 817-823.

Hall, P. and Marron, J.S. (1987) Choice of kernel order in density estimation. *Ann. Statist.*, **16**, 161-173.

Hall, P., Sheather, S., Jones, C. and Marron, J.S. (1991). Optimal data-driven bandwidth selection in kernel density estimation. *Biometrika*, **78**, 263-269.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

Härdle, W. and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression estimation. *Ann. Statist.*, **13**, 1465-1481.

Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized Discriminent Anakysis. *Ann. Statist.*, **23**, 73-102.

Hastie, T. and Mallows, C. (1993). A discussion of "A statistical view of some chemometric regression tools" by E.I Frank and J.H. Friedman. *Technometrics*, **35**, 140-143.

He, X. and Shi, P. (1996). Convergence rate of B-spline estimators of non parametric conditional quantile functions. *Nonparametric Statist.*, **3**, 56-59.

Huber, P.J. (1985). Projection Pursuit. *Ann. Statist.*, **13**, 435-475.

Izenman, A.J. (1991). Recent Developments in Nonparametric Density Estimation. *J. Amer. Statist. Assoc.*, **86**, 205-223.

Linton, O.B. and Auestad, J.B. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82**, 93-100.

Loader, C.R. (1999). Bandwidth selection : classical or plug-in ? *Ann. Statist.*, **27**, 415-438.

Lofstgaarden, D.O. and Quesenberry, C.D. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.*, **36**, 1049-1051.

Macauley, F.R. (1931). *The smoothing of time series* New York : National Bureau of Economic Research.

Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm. *Ann. Statist.*, **27**, 1443-1490.

Marron, J.S. (1987). A comparison of cross-validation techniques in density estimation. *Ann. Statist.*, **15**, 152-162.

Marron, J.S. (1988). Automatic smoothing parameter selection : a survey. *Empirical Economics*, **13**, 187-208.

Marx, B.D. and Eilers, P.H. (1999). Generalized Linear Regression on Sampled Signals and Curves : A P-Spline Approach. *Technometrics*, **41**, 1-13.

Murphy, V. (1965). Estimation of jumps, reliability and hazard rate. *Ann. Statist.*, **36**, 1032-1040.

Nadaraya, E.A. (1964). On estimating regression. *Theory Probab. Appl.*, **9**, 141-142.

Parzen, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.*, **35**, 1065-1076.

Pearson, K. and Lee, A. (1903). On the laws of inheritance in a man. *Biometrika*, **2**, 357-462.

Peligrad, M. (1988). Properties of uniform consistency of the kernel estimator of density and of regression. Manuscrit non publié.

Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis* Springer-Verlag, Berlin.

Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function. *Sankhya, Ser. A*, **38**, 60-78.

Romain, Y. (1997). Une introduction à l'approche fonctionnelle stochastique de la statistique. *Pub. lab. Stat. Prob.*, **02**-97, 1-28, Université Paul Sabatier, Toulouse.

Ruppert, D., Sheather, S.J. and Wand, M.P. (1993). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, **90**, 1257-1270.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Satist.*, **27**, 832-837.

Rudemo, M. (1982). Empirical choices of histograms and kernel density estimates. *Scand. J. Statist.*, **9**, 65-78.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Silverman, B.W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, **24**, 1-24.

Sperlich, S., Tjøstheim, D. and Yang L. (2000). Nonparametric estimation and testing of interaction in additive models. Preprint.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **4**, 595-645.

Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040-1053.

Stone, C.J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives. *Recent Advances in Statistics : Papers presented in Honor of Herman Chernoff's Sixtieth Birthday*, M.H. Rizvi, J.S. Rustagi, and D. Siegmund, (eds.) Academic Press, New York.

Stone, C.J. (1984). An asymptotically window selection rule for kernel density estimaes. *Ann. Statist.*, **12**, 1285-1297.

Stone, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, **13**, 689-705.

Tjøstheim, D. and Auestad, B. (1994). Nonparametric identification of nonlinear time series : projections. *J. Amer. Statist. Assoc.*, **89**, 1398-1409.

Tukey, J.W. (1961). Curves as parameters and touch-estimation. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 681-694.

Vieu, P. (1993). Bandwidth selection for kernel regression : a survey. In *Computer Intensive Methods in Statistics*, Statistics and Computing, Physica Verlag, Berlin, Vol. I, W. Härdle et L. Simar Eds, 134-149.

Vieu, P. (1994). Choice of regressors in nonparametric estimation. *Comp. Statist. Data Analysis*, **17**, 575-594.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia, PA.

Watson, G.S. (1964). Smooth regression analysis. *Sankhya, Ser. A*, **26**, 359-472.

Wu, J.S. and Chu, C.K. (1993). Kernel-type estimators of jump points and values of a regression function. *Ann. Statist.*, **21**, 1545-1566.

Yoshihara, K. (1984). Density estimation for samples satisfying a certain absolute regularity condition. *J. Statist. Plan. Inf.*, **9**, 19-32.

Yurinskiĭ, V.V. (1976). Exponential inequalities for sums of random vectors. *J. Mult. Analysis*, **6**, 473-499.

Liste de Publications

Publications dans des revues à comité de lecture.

(28) Cardot, H., Ferraty, F. et Sarda, P. 1999. Functional Linear Model. *Stat. Prob. Letters*, **45**, 11-22.

(27) El Faouzi, N. E. et Sarda, P. 1999. Rates of Convergence for Additive Principal Components, *J. Mult. Analysis*, **68**, 120-137.

(26) Couallier, V., Sarda, P. et Vieu, P. 1997. Estimation non paramétrique de discontinuités d'une fonction d'intensité, *Revue de Statistique Appliquée*, **XLV**, 89-106.

(25) Youndjé, E., Sarda, P. et Vieu, P. 1996. Optimal Smooth Hazard Estimates, *TEST*, **5**, 379-394.

(24) Moncaup, S., Sarda, P. et Vieu, P. 1995. Une mise en oeuvre d'estimateurs non paramétriques d'une fonction d'intensité sur des données météorologiques, *Revue de Statistique Appliquée*, **XLIII**, 77-88.

(23) Youndjé, E., Sarda, P. et Vieu, P. 1994. Validation croisée pour l'estimation non-paramétrique de la densité conditionnelle, *Pub. Inst. Stat. Univ. Paris*, **38**, 57-80.

(22) Lejeune, M. et Sarda, P. 1992. Smooth estimators of distribution and density functions, *Computational Statistics and Data Analysis*, **14**, 457-471.

(21) Sarda, P. et Vieu, P. 1991. Smoothing parameter selection in hazard estimation, *Statistics and Probability Letters*, **11**, 429-434.

(20) Sarda, P. et Vieu, P. 1989c. Estimation non-paramétrique de la fonction de hasard, *cahiers du C.E.R.O.*, **31**, 241-265.

(19) Mielniczuk, J., Sarda, P. et Vieu, P. 1989b. Local data-driven bandwidth choice for density estimation, *Journal of Statistical Planning and Inference*, **23**, 53-69.

(18) Sarda, P. et Vieu P. 1989a. Empirical distribution function for mixing random variables. Application in nonparametric hazard estimation, *Statistics*, **20**, 4, 559-571.

(17) Lejeune, M. et Sarda, P. 1988. Quantile regression : a nonparametric approach, *Computational statistics and data analysis*, **6**, 229-239.

(16) Collomb, G., Sarda, P. et Vieu, P. 1986c. Weak pointwise consistency of the cross-validatory window estimate in nonparametric estimation, *C.M.U.C. Université Carlova*, Prague, **26**, 789-798.

(15) Hassani, S., Sarda, P. et Vieu, P. 1986b. Approche non paramétrique en théorie de la fiabilité, *Revue de Statistique Appliquée*, **35**, 27-41.

(14) Sarda, P. et Vieu P. 1986a. Régression non-paramétrique et prédiction d'un processus markovien, *Cahiers du C.E.R.O.*, Bruxelles, **28**, 203-209.

(13) Collomb, G., Hassani, S., Sarda, P et Vieu P. 1985. Estimation non-paramétrique de la fonction de hasard pour des observations dépendantes, *Statistique et Analyse des données*, **10**, 42-49.

Ouvrages ou participation à des ouvrages collectifs

(12) Sarda, P et Vieu P. 2000. Kernel Regression, M. Schimek (ed), in *Smoothing and Regression : Approaches, Computation and Application*, Wiley, New-York.

(11) Györfi, L., Härdle, W., Sarda, P. et Vieu, P. 1990. *Nonparametric curve estimation from time series*, Lecture Notes in Statistics, Springer Verlag, VI.

Notes, Publications internes et Comptes rendus de congrès

(10) Cardot, H., Ferraty, F. et Sarda, P. 2000. Etude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel. *Comptes Rendus de l'Académie des Sciences*, **330**, Série I, 501-504.

(9) Cardot, H., Ferraty, F. et Sarda, P. 1998. Modèle Linéaire Fonctionnel, *Pub. Lab. Stat. Prob.*, **04**-98, 1-18, Université Paul Sabatier, Toulouse.

(8) Pelegrina, L., Sarda, P. et Vieu, P. 1996. On multidimensional nonparametric regression, **COMPSTAT**, *Proceedings in Computational Statistics*, Ed. A. Prat.

(7) Youndjé, E., Sarda, P. et Vieu, P. 1993. Estimateur à noyau d'une densité conditionnelle : choix de la fenêtre pour des observations dépendantes,*Comptes Rendus de l'Académie des Sciences*, **316**, Série I, 935-938.

(6) Sarda, P. et Vieu, P. 1988. Vitesse de convergence d'estimateurs non-paramétriques de la régression et de ses dérivées, *Comptes Rendus de l'Académie des Sciences*, **306**, Série I, 83-86.

(5) Sarda, P. et Vieu, P. 1986. Vitesse de convergence uniforme de l'estimateur à noyau de la régression pour des observations dépendantes, *Comptes Rendus de l'Académie des Sciences*, **302**, Série I, 423-426.

(4) Collomb, G., Hassani, S., Sarda, P. et Vieu P. 1985. Convergence uniforme d'estimateurs de la fonction de hasard pour des observations dépendantes : méthodes du noyau et des k-points les plus proches, *Comptes Rendus de l'Académie des Sciences*, **301**, Série I, 653-656.

Articles soumis.

(3) Camlong, C., Sarda, P. et Vieu, P. 1999. Additive Time Series : the Kernel Integration Method. *soumis*.

(2) Cardot, H., Ferraty, F. et Sarda, P. 1999. Spline Estimators for the Functional Linear Model. *soumis*.

(1) Cardot, H., Ferraty, F., Mas, A. et Sarda, P. 2000. Testing Hypotheses in the Functional Linear Model. *soumis*.