

Sommaire

Introduction	3
Estimation fonctionnelle	
P. Sarda et P. Vieu	9
Modélisation pour variables fonctionnelles dans un contexte explicatif	
H. Cardot et F. Ferraty	27
Sur et pour une approche fonctionnelle en statistique	
Y. Romain	45
Produit de convolution de mesures spectrales	
A. Boudou	49
The geometrical theory of estimating functions	
C. Small	51
Inférence statistique pour des estimateurs de discontinuités dans un cadre non paramétrique	
V. Couallier	55
Nonparametric estimation in null recurrent time series	
D. Tjøstheim	57
ACP de fonctions de densité. Application aux données climatiques	
T. Antoniadou <i>et al.</i>	61
Modèle non linéaire fonctionnel: une approche par régression inverse	
A.F. Yao et L. Ferré	63
Estimation bayésienne de l'intensité d'un processus de Cox non homogène par une méthode MCMC à saut réversible	
M. Goulard	65
Permutation tests in change point analysis	
J. Antoch et M. Hušková	67
Inférence statistique pour la localisation d'une discontinuité par régression linéaire locale	
G. Grégoire	69
Non causalité et discrétisation fonctionnelle, théorèmes limites pour un processus ARHX(1)	
S. Guillas	73
Data exploration using piecewise polynomial regression trees	
P. Chaudhuri	75

La naissance du groupe de travail STAPH: Une nécessité historique, scientifique

H. CARDOT, F. FERRATY, Y. ROMAIN, P. SARDA ET P. VIEU

1 Le contexte historique et scientifique

Les aspects fonctionnels de la Statistique ont toujours été très présents au Laboratoire de Statistique et Probabilités de Toulouse. Historiquement, ces aspects trouvent leurs racines dans deux familles de recherches dont les travaux sont reconnus par la communauté scientifique française et internationale :

- L'estimation non-paramétrique “à la Collomb”;
- L'environnement fonctionnel de la Statistique multidimensionnelle “à la Dauxois-Pousse”.

Ces deux domaines ont, pour l'essentiel, été développés de manière indépendante dans l'ensemble de la littérature. Cependant, l'évolution récente de la Statistique Fonctionnelle induit des liens de plus en plus étroits entre eux. Il s'agit de l'utilisation de mêmes outils mathématiques (l'analyse fonctionnelle par exemple) ou encore de problèmes statistiques pour lesquels des modèles alliant les deux approches sont introduits (modèles fonctionnels, estimation fonctionnelle multidimensionnelle,...). Enfin, dans de nombreuses études de cas les deux approches sont souvent complémentaires.

Cette évolution rend nécessaires des études approfondies sur tous les aspects fonctionnels de la Statistique: le contexte historique évoqué ci-dessus fait de Toulouse un lieu privilégié pour cette démarche. Le groupe de travail **STAPH** s'inscrit ainsi dans ce double contexte scientifique et historique.

2 Les objectifs scientifiques de STAPH

Le groupe de travail STAPH se fixe deux objectifs. Il s'agit, tout d'abord, de poursuivre l'étude des deux axes initiaux. Au delà de cette étude, la recherche et l'approfondissement de liens entre ces deux axes est une démarche essentielle du

groupe de travail. C'est là notre second objectif qui, comme nous l'avons souligné plus haut (voir également les trois premiers exposés), répond à une évolution récente de la Statistique. L'utilisation conjointe des deux approches devraient permettre d'appréhender des problématiques telles que la modélisation de variables fonctionnelles comme en témoignent d'ailleurs des travaux récents, dont certains réalisés par des membres de notre laboratoire. En ce sens, ce second objectif nous apparaît comme le plus novateur et c'est celui que nous souhaitons privilégier.

Les trois thèmes principaux qui ont ainsi été retenus sont les suivants:

- Environnement opératoire de la Statistique et de ses branches connexes;
- Estimation fonctionnelle uni et multi dimensionnelle;
- Modèles à variables fonctionnelles.

3 Les fondements de groupe de travail

Nous sommes attachés à une certaine façon de vivre et de penser la Recherche et la Recherche en Statistique en particulier. Pour ce qui concerne le groupe de travail, celle-ci s'articule autour de trois points : une exigence d'ouverture scientifique, un mode de fonctionnement respectueux des besoins et des envies de tous les participants et un ancrage profond dans la communauté scientifique toulousaine.

Ouverture scientifique

De par la position de la Statistique en tant que discipline à part entière des Mathématiques Appliquées, le groupe STAPH n'entend privilégier aucun des aspects qui en font son originalité. Les approches fondamentales doivent y trouver leur place au même titre que les études appliquées. En ce sens, nous souhaitons que ce groupe soit un lieu d'activation des interactions entre "Statistique Fondamentale" et "Statistique Appliquée", en privilégiant également les collaborations avec d'autres domaines scientifiques qui soient mutuellement profitables.

La position spécifique de la Statistique en fait une discipline ouverte sur de nombreux autres domaines, que ce soit vers d'autres secteurs des Mathématiques (Calcul des Probabilités, Mathématiques Fondamentales, Analyse Numérique, ...), mais aussi vers la Physique, l'Informatique, ... , et enfin vers des disciplines scientifiques pour lesquelles la Statistique est un outil (Biologie, Médecine, Océanographie, ...). Nous souhaitons que ce groupe de travail soit aussi un lieu d'échanges entre toutes ces disciplines.

Pour résumer ce premier aspect, les interactions intra et extra Statistique devraient être au centre de l'activité du groupe de travail qui entend œuvrer ainsi au décloisonnement entre "Statistique Fondamentale" et "Statistique Appliquée", tant il est vrai que cette séparation n'a pas lieu d'être.

Fonctionnement

Pour ce qui concerne le déroulement des séances de travail, il nous paraît impératif de se donner les moyens d'un échange véritable entre les participants, ce qui implique une ambiance détendue où chacun (orateur et participant) peut non seulement faire part de ses connaissances mais aussi (et surtout) de ses lacunes.

Par ailleurs, il ne nous semble pas souhaitable de privilégier une forme a priori mais plutôt de choisir dans chaque situation la forme la mieux adaptée pour répondre aux exigences ci-dessus. Nous souhaitons ainsi favoriser une large diversité dans la nature des séances (exposé de résultats théoriques, étude de cas, table ronde sur un thème précis, exposé de synthèse,...).

L'animation du groupe de travail est assuré par un collectif. Nous tenons beaucoup à cette idée de travail collectif, trop souvent négligé dans notre métier de chercheur. Ce collectif d'animation est, pour l'instant, composé des cinq membres fondateurs du groupe de travail.

Ancrage dans la communauté statistique et scientifique toulousaine

Le groupe de travail se situe au sein du **Laboratoire de Statistique et Probabilités de Toulouse** et se veut naturellement ouvert à tous ses membres. Compte tenu de la nature des thèmes abordés, il est clair que tous (et pas seulement les statisticiens) peuvent à un moment ou un autre apporter leur contribution à cette démarche et/ou profiter de la réflexion collective offerte par le groupe de travail.

Comme en témoigne les exposés présentés au cours de cette année, cette ouverture dépasse le cadre du L.S.P. et concerne non seulement l'ensemble des statisticiens ou utilisateurs de la statistique toulousains mais également des chercheurs en visite dans notre laboratoire ou dans un laboratoire toulousain.

Nous conclurons cette introduction en remerciant l'ensemble des orateurs et participants au groupe de travail. L'intérêt qui a été porté au groupe de travail au long de cette première année d'existence, si l'on en juge par le nombre de participants et par la diversité des exposés, témoigne d'un bilan positif. Il indique que les thèmes abordés sont porteurs de projets scientifiques et nous pousse à poursuivre et développer ce travail.

Résumés

Estimation fonctionnelle

Pascal SARDA et Philippe VIEU
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
118, Route de Narbonne
31062, Toulouse Cédex
sarda@cict.fr
vieu@cict.fr
13/12/99

1 Introduction

Cet exposé s'inscrit en préambule à l'un des thèmes du groupe de travail. Son objectif est de faire un point sur la théorie de l'estimation fonctionnelle et se situe donc implicitement dans le cadre de la statistique non paramétrique. Il ne s'agit pas de faire une revue bibliographique sur l'estimation fonctionnelle, domaine qui a donné lieu à une importante littérature depuis une quarantaine d'années, mais d'en exposer les principes fondamentaux en les situant dans leur contexte historique et de dégager les perspectives et problèmes ouverts. Par ailleurs, nous nous sommes essentiellement concentrés sur les aspects de l'estimation non paramétrique qui ont été le plus étudiés au Laboratoire de Statistique et Probabilités en ne cherchant pas à atteindre l'exhaustivité.

Ce travail doit beaucoup à Gérard Collomb qui fut le précurseur de la statistique non paramétrique au Laboratoire de Statistique et Probabilités et qui a guidé nos premiers pas en recherche. L'écho de ses travaux dépasse en outre le cadre de notre laboratoire en se situant au tout premier plan des publications internationales dans le domaine entre 1975 et 1985, date du décès de Gérard. Citons, par exemple, les résultats asymptotiques concernant les estimateurs à noyau de la régression ou encore ses derniers travaux sur l'estimation non paramétrique dans les processus et la généralisation de l'inégalité de Bernstein au cas de variables dépendantes, qui sont encore cités en référence dans la littérature (voir par exemple Collomb, 1976, 1977, 1984 et Collomb et Härdle, 1986).

2 Modèles non paramétriques et estimation fonctionnelle

2.1 Modèles non paramétriques

Considérons un **modèle statistique**, c'est-à-dire un triplet $(E, \mathcal{A}, \mathcal{P})$, où E est l'espace des observations, \mathcal{A} une tribu sur E et $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ une famille de mesures de probabilités sur (E, \mathcal{A}) . Etant donné g une fonction définie sur \mathcal{P} et à valeurs dans un espace Θ' et une loi inconnue P_θ dans \mathcal{P} , le problème posé est celui de l'**estimation** de $g(P_\theta)$ à partir d'une réalisation d'une variable aléatoire \mathcal{X} à valeurs dans E et de loi P_θ . Cela revient donc à construire une application de E dans Θ' appelée **estimateur du paramètre** $g(P_\theta)$.

La nature du modèle (paramétrique, non paramétrique) est liée à celle de l'ensemble Θ . Notons tout de suite que la frontière entre modèle paramétrique et modèle non paramétrique est assez floue. Bosq et Lecoutre (1987) définissent un modèle non paramétrique comme un modèle pour lequel Θ contient un ensemble convexe de dimension infinie tandis que le modèle est paramétrique lorsque Θ est un ouvert de \mathbb{R}^s . Cependant, la définition d'un modèle non paramétrique la plus largement utilisée dans la littérature (voir par exemple les travaux de Collomb) est la suivante :

Définition. *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.*

C'est cette définition que nous adoptons dans la suite. Il est clair qu'elle exclut par exemple le modèle linéaire non gaussien pour lequel l'estimation de la fonction de régression se résume à l'estimation de q paramètres réels. Cependant, ce modèle n'est pas non plus paramétrique puisque \mathcal{P} n'est pas décrit par un nombre fini de paramètres réels. On parle parfois dans ce cas de modèle semi paramétrique.

Nous nous limitons au modèle statistique :

$$(E, \mathcal{A}, \mathcal{P}) = (\mathbb{R}^{kn}, \mathcal{B}_{\mathbb{R}^{kn}}, (P^n)_{P \in \mathcal{P}_0}).$$

Dans ce cas la variable \mathcal{X} est un vecteur (Z_1, \dots, Z_n) de variables aléatoires Z_i à valeurs dans \mathbb{R}^k indépendantes et identiquement distribuées de loi $P \in \mathcal{P}_0$. Nous notons Z une variable générique de loi P .

2.2 Estimation fonctionnelle

Nous nous intéressons au cas où $g(P_\theta)$ est un **paramètre fonctionnel**, c'est-à-dire au cas où Θ' est un espace vectoriel de dimension infinie, autrement dit un espace fonctionnel. Nous parlons dans ce cas d'**estimation fonctionnelle**. Pour

fixer les idées donnons deux exemples les plus courants de paramètres fonctionnels: la densité et la régression.

Densité. Nous supposons que \mathcal{P}_0 est dominée par la mesure de Lebesgue sur \mathbb{R}^k . La mesure $\frac{dP}{d\lambda}$ est le paramètre fonctionnel densité à valeurs dans l'espace L^1 . Des hypothèses supplémentaires (par exemple l'existence d'une version bornée et/ou continue) restreignent le modèle non paramétrique sans changer la nature fonctionnelle du paramètre à estimer.

Régression. On pose $k = d + 1$ et $Z = (X, Y)$ où X est à valeurs dans \mathbb{R}^d et Y est à valeurs dans \mathbb{R} . Soit P_{XY} la loi inconnue de (X, Y) . Pour $x \in \mathbb{R}^d$, on note $P_Y^{X=x}$ la loi conditionnelle de Y sachant $X = x$ et on suppose que $E|Y| < +\infty$. Alors l'espérance conditionnelle $\mathbb{E}(Y|X)$ est définie par :

$$\forall x \in \mathbb{R}^d, \mathbb{E}(Y|X) = \int_{\mathbb{R}} y P_Y^{X=x}(dy) = r(x).$$

Dans la suite nous ferons un abus de langage usuel en appelant **estimateur non paramétrique** un estimateur d'un paramètre fonctionnel dans un modèle non paramétrique.

3 Remarques générales

3.1 A propos des hypothèses

Dans le cadre de l'estimation non paramétrique de la densité ou de la régression, on ne fait pas d'hypothèse autre que celle de l'existence de cette fonction (f ou r suivant le cas). Notons cependant que lorsque l'on s'intéresse à la convergence d'un estimateur f_n vers f (ou r_n vers r) ponctuelle en un point x , respectivement uniforme sur un ensemble G , les hypothèses minimales dans la littérature sont les suivantes :

Densité.

(D1) f est continue en x , respectivement sur un voisinage de G .

Régression.

(R1) r est continue en x , respectivement sur un voisinage de G ;

(R2) la loi marginale P_X admet une densité f par rapport à la mesure de Lebesgue sur \mathbb{R}^d et f est minorée par une constante strictement positive sur un voisinage de x , respectivement de G ;

(R3) $\mathbb{E}Y^2 < +\infty$ et $v(x) = \text{Var}(Y|X = x)$ est bornée sur un voisinage de x , respectivement de G .

Sous ces hypothèses, on obtient la convergence (p.s., p., m.q.) ponctuelle ou uniforme sur G des estimateurs non paramétriques les plus courants de f et de r . Notons que pour obtenir des résultats plus précis sur les estimateurs (biais et variance par exemple) nous avons besoin d'hypothèses plus restrictives sur les fonctions f ou r telles que l'existence de deux dérivées continues par exemple. Ce type de résultats sera discuté plus en détails ultérieurement.

3.2 Biais des estimateurs

Dans la plupart des modèles paramétriques, on introduit souvent des estimateurs (des paramètres tels que l'espérance, la variance, ...) sans biais et de variance minimum dans cette classe. Dans un grand nombre de modèles non paramétriques et pour la plupart des paramètres fonctionnels considérés il n'est pas possible de procéder de cette manière. Le résultat suivant qui est une conséquence directe d'un théorème de Bickel et Lehman (voir Collomb, 1976 et Bosq et Lecoutre, 1987) illustre ceci dans le cas de la densité et de la régression.

Théorème. *Sous l'hypothèse (D1), (respectivement sous les hypothèses (R1), (R2), (R3)), il n'existe pas d'estimateur sans biais de f , (respectivement de r), c'est-à-dire si \mathcal{P}_0 est l'ensemble des lois vérifiant (D1) (respectivement (R1), (R2), (R3)), alors il n'existe pas de f_n (respectivement de r_n) tel que :*

$$\forall P \in \mathcal{P}_0, \forall x \in \mathbb{R}^k, \mathbb{E}f_n(x) = f(x),$$

$$(\text{respectivement } \forall P \in \mathcal{P}_0, \forall x \in \mathbb{R}^d, \mathbb{E}r_n(x) = r(x)).$$

Il existe cependant des cas de paramètres fonctionnels pour lesquels on peut trouver des estimateurs sans biais. Par exemple la fonction de répartition est estimée sans biais par la fonction de répartition empirique et ceci quelle que soit la forme de la f.d.r. c'est-à-dire dans le cadre d'un modèle non paramétrique très large. Notons encore la fonction caractéristique empirique qui estime sans biais la fonction caractéristique là encore sans restriction sur cette fonction.

Compte tenu du résultat précédent, la première propriété généralement établie pour un estimateur non paramétrique est celle du non biais asymptotique. Ensuite pour mesurer sa qualité on introduit une norme fonctionnelle. Pour des raisons de commodité, on choisit la plupart du temps l'Erreur Quadratique Intégrée (c'est-à-dire la norme L^2) ou encore l'Erreur Quadratique Moyenne Intégrée.

3.3 Vitesses de convergence optimale

Parmi les nombreux travaux concernant les propriétés asymptotiques d'estimateurs non paramétriques, plusieurs auteurs se sont intéressés à leurs vitesses de convergence selon divers modes de convergence. Ils obtiennent généralement des bornes supérieures pour la vitesse de convergence pour la convergence en probabilité, presque sûre ou en norme L^2 . Les vitesses de convergence d'estimateurs usuels dans le cadre d'un modèle non paramétrique sont généralement moins bonnes que la vitesse $n^{-1/2}$ obtenue pour des estimateurs paramétriques. Se pose alors la question de l'optimalité de ces vitesses de convergence.

Stone répond à cette question au début des années quatre-vingts à travers plusieurs articles portant sur l'estimation de la régression et celle de la densité (cf. Stone, 1982 et 1983). Ainsi, pour $P_\theta \in \mathcal{P}$, où \mathcal{P} est une famille de mesures de probabilité et $g(P_\theta)$ un paramètre fonctionnel (régression ou densité) à estimer, l'auteur obtient la vitesse de convergence optimale ponctuelle ou uniforme pour un estimateur de $g(P_\theta)$ uniformément sur \mathcal{P} .

Dans le cas de la régression par exemple, supposons que r soit à valeurs dans un compact G de \mathbb{R}^d et considérons la norme L^q , $\|\cdot\|_q$ $1 \leq q \leq \infty$, dans l'ensemble des fonctions définies sur G et à valeurs dans \mathbb{R} . Notons alors par r_n un estimateur de la régression r_θ et par (b_n) une suite de nombre réels positifs.

Définition. 1. On dit que (b_n) est une **borne inférieure pour la vitesse de convergence** si :

$$\exists C > 0, \liminf_{n \rightarrow \infty} \sup_{r_n} \sup_{P_\theta \in \mathcal{P}} P [\|r_n - r_\theta\|_q > Cb_n] = 1.$$

2. On dit que (b_n) est une **vitesse de convergence atteinte** si :

$$\exists r_n, \exists C > 0, \lim_{n \rightarrow \infty} \sup_{P_\theta \in \mathcal{P}} P [\|r_n - r_\theta\|_q > Cb_n] = 0.$$

3. On dit que (b_n) est une **vitesse de convergence optimale**, si (b_n) est une borne inférieure pour la vitesse de convergence atteinte.

Stone (1982) montre que sous certaines hypothèses sur la loi conditionnelle de Y sachant X et lorsque la régression r est k fois continuellement différentiable et que sa dérivée d'ordre k est lipschitzienne d'ordre β , la vitesse optimale de convergence est :

$$n^{-(k+\beta)/2(k+\beta)+d}, \quad 0 < q < \infty,$$

et

$$(\log n/n)^{-(k+\beta)/2(k+\beta)+d}, \quad q = \infty.$$

On voit que cette vitesse de convergence dépend d'une part de la régularité de la fonction à estimer et d'autre part de la dimension d . La vitesse de convergence se détériore ainsi avec la dimension et ce problème a donné lieu à une importante

littérature à partir du milieu des années quatre-vingts : nous y reviendrons au chapitre 6. Notons enfin que les vitesses de convergence optimales obtenues dans le cas de la densité sont identiques.

4 Quelques estimateurs non paramétriques de la régression et de la densité

4.1 Principes généraux des estimateurs non paramétriques

Nous nous intéressons principalement aux problèmes d'estimation de la régression et de la densité qui sont les plus étudiés dans la littérature. Notons également que les principes d'estimation énoncés ci-dessous sont transposables à d'autres problèmes d'estimation fonctionnelle.

De nombreux estimateurs non paramétriques de la régression ont en commun l'approche suivante. Étant donné un échantillon $(X_i, Y_i), i = 1, \dots, n$, pour tout x de \mathbb{R}^d , $r_n(x)$ est une moyenne pondérée des observations Y_i , la pondération de Y_i tenant compte de la "distance" entre X_i et x .

Le principe de construction de plusieurs estimateurs non paramétriques de la densité est basé sur une idée analogue. En effet, pour un échantillon $X_i, i = 1, \dots, n$ et pour x élément de \mathbb{R}^k , $f(x)$ est estimée à partir d'une proportion d'observations (pondérées) X_i "proches" de x .

Parmi les estimateurs reposant sur ce principe, les estimateurs à noyau sont les plus couramment utilisés à cause de leur facilité de calcul et de leurs bonnes propriétés statistiques. Ces estimateurs feront donc l'objet d'une attention particulière dans la suite.

La deuxième particularité des estimateurs non paramétriques est qu'ils dépendent principalement d'un paramètre contrôlant leur régularité. L'existence d'un tel paramètre tient à la nature fonctionnelle de l'objet estimé. On appelle ce paramètre **paramètre de lissage**. Ainsi les techniques d'estimation non paramétriques ne sont pas a priori automatiques bien que pour la plupart d'entre elles, des méthodes de sélection du paramètre de lissage ont été introduites. Nous consacrerons le paragraphe 5 à ce problème.

Parmi les ouvrages généraux sur l'estimation fonctionnelle, outre celui de Bosq et Lecoutre (1987) déjà cité, citons les livres de Härdle (1990), Wand et Jones (1995), Fan et Gijbels (1996), Loader (1999) qui portent principalement sur l'estimation de la régression et les méthodes du noyau, l'ouvrage de Silverman (1986) qui traite de la densité et les livres de Härdle (1991) et de Azzalini et Bowman (1997) qui étudient les aspects appliqués de l'estimation fonctionnelle (implémentation dans le logiciel S-Plus).

4.2 Régressogramme et histogramme

Cet estimateur a été introduit par Tukey (1961) et consiste à estimer la fonction de régression par une fonction constante par morceaux. Pour fixer les idées donnons sa définition dans le cas univarié $d = 1$. On découpe l'intervalle d'estimation en k_n sous-intervalles disjoints et généralement de même amplitude. Sur chaque sous-intervalle l'estimateur r_n est constant et égal à la moyenne des $Y_i, i = 1, \dots, n$ tels que X_i appartienne à ce sous-intervalle (on fixe cette constante égale à 0 si aucun X_i n'appartient au sous-intervalle). Cette définition se généralise de manière évidente au cas d'une subdivision en intervalles d'amplitudes distinctes et au cas $d \geq 2$.

Plusieurs auteurs ont étudié le régressogramme d'un point de vue asymptotique. Parmi eux citons Bosq (1970) et Geffroy (1975) qui obtiennent respectivement la convergence presque sûre uniforme et une condition nécessaire et suffisante de convergence uniforme presque sûre. Collomb (1978) obtient l'écriture asymptotique du biais et de la variance ponctuelles en fonction de $r'(x)$, $f(x)$ et $v(x)$ ainsi que la convergence en loi.

L'analogie entre le régressogramme et l'**histogramme**, pour estimer une densité f est évidente. ce dernier étant construit sur le même principe en notant que la constante dans chaque subdivision est égale à la proportion de X_i tombant dans cette subdivision.

Malgré leur facilité de calcul, régressogramme et histogramme ont les mêmes défauts : manque de régularité (fonctions en escaliers), mauvais comportement du biais sur les bords de chaque subdivision (*cf.* Collomb, 1978, pour le régressogramme) et enfin problème du choix de la position des subdivisions.

4.3 Estimateurs à noyau

L'estimateur à noyau a été introduit par Watson (1964) et Nadaraya (1964) de manière indépendante et simultanée. Etant donné un réel positif h_n et un **noyau** de \mathbb{R}^d , c'est-à-dire une fonction K bornée sur \mathbb{R}^d , symétrique et telle que $|y|^d K(y) \rightarrow_{|y| \rightarrow \infty} 0$, l'**estimateur à noyau** r_n de la régression r est défini en un point x de \mathbb{R}^d comme une moyenne pondérée des Y_i :

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

On adopte la convention $r_n(x) = 0$, lorsque le dénominateur s'annule. L'estimateur s'adapte donc localement aux données. L'estimateur $r_n(x)$ est en effet

solution du problème de minimisation :

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

c'est-à-dire qu'il revient à ajuster localement (autour du point x) une constante à l'aide du critère des moindres carrés pondérés.

Parmi les résultats asymptotiques concernant cet estimateur notons ceux obtenus par Schuster (1972) qui montre la normalité asymptotique, Collomb (1976 et 1977) qui donne une évaluation asymptotique du biais (qui dépend de r'' , f' et v) et de la variance ou encore Devroye (1979) qui établit une condition suffisante de convergence en norme L^q .

D'autres estimateurs à noyau de la régression ont été introduits par la suite. La généralisation porte sur le degré du polynôme ajusté autour du point x : voir à ce sujet le livre de Fan et Gijbels (1996). Le plus étudié est celui reposant sur une **régression linéaire locale** c'est-à-dire un polynôme de degré 1. Il a été introduit par Cleveland (1979) (qui en donne une version robuste) puis étudié par la suite par divers auteurs dont principalement Fan et Gijbels (voir leur ouvrage, 1996, pour les références à ce sujet). La popularité de cet estimateur vient de ce que l'expression asymptotique de son biais est plus simple que celle de l'estimateur de Nadaraya-Watson puisqu'elle ne dépend que de la dérivée seconde de r . Par ailleurs, il n'est pas nécessaire de corriger la forme du noyau sur les bords du domaine d'estimation contrairement à l'estimateur de Nadaraya-Watson (voir Gasser et Müller, 1979).

On définit l'**estimateur à noyau** de la densité par :

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

Cet estimateur a été introduit par Rosenblatt (1956) puis étudié par Parzen (1962) qui établit entre autres la normalité asymptotique. L'estimateur $f_n(x)$ apparaît au dénominateur de l'estimateur de Nadaraya-Watson ci-dessus et ainsi les auteurs qui établissent des propriétés pour r_n montrent généralement des résultats analogues pour l'estimateur f_n . On peut donc trouver dans les travaux ci-dessus des résultats concernant l'estimateur à noyau de la densité.

4.4 Autres estimateurs

De nombreux autres estimateurs non paramétriques de la régression (ou de la densité) ont été introduits dans la littérature. Outre les estimateurs à noyau, les **estimateurs splines** sont également très populaires. En particulier, les estimateurs reposant sur les **splines de régression** ont été principalement étudiés par

Wahba : on peut se reporter à Wahba (1990) pour une revue de ses principaux articles et résultats (voir également Eubank, 2000). L'estimateur de la régression s'obtient, dans le cas univarié et pour un échantillon (t_i, Y_i) , $i = 1, \dots, n$ où $t_1 < \dots < t_n$ sont fixés, en minimisant le critère :

$$\sum_{i=1}^n (Y_i - g(t_i))^2 + \lambda \int_{t_1}^{t_n} g^{(m)}(t)^2 dt,$$

parmi les fonctions g ayant $m - 1$ dérivées absolument continues et une dérivée d'ordre m intégrable. Lorsque $n \geq m$, la solution unique est un polynôme de degré $2m - 1$ sur $[t_i, t_{i+1}]$ et $2m - 2$ fois dérivable sur $[t_1, t_n]$, c'est-à-dire une fonction spline d'ordre $2m$ avec des noeuds placés aux points t_1, \dots, t_n . Les estimateurs splines ont été principalement étudiés dans le cas univarié (voir cependant Gu, 2000, pour une bibliographie sur le cas multivarié) et dans le cas d'un plan fixe.

Un autre type d'estimateur spline est également étudié dans la littérature quoique de manière moins importante que le précédent : il s'agit des estimateurs **splines de régression**. Dans le cas univarié, cet estimateur s'écrit à l'aide d'une fonction spline d'ordre q avec $k + 1$ noeuds (généralement équidistants sur l'intervalle d'estimation). L'ensemble de ces fonctions est un espace vectoriel de dimension $k + q$ dont les fonctions B-splines en constituent une base. On cherche alors une combinaison linéaire de B-splines minimisant le critère des moindres carrés. Agarwal et Studen (1980) donnent l'expression asymptotique de l'erreur quadratique moyenne intégrée. Stone (1985) utilise cet estimateur dans le cadre du modèle additif de régression (*cf.* ci-dessous) et généralise dans un article ultérieur cet estimateur au cas multivarié (*cf.* Stone, 1994).

Parmi les autres estimateurs de la régression, citons ceux reposant sur les ondelettes : voir Nason et Silverman (2000) pour un article de synthèse sur le sujet.

5 Choix du paramètre de lissage

Les propriétés asymptotiques des estimateurs autant que leur mise en oeuvre montrent que leur qualité dépend de manière cruciale du choix du paramètre de lissage : amplitude k_n des subdivision pour le régressogramme, largeur de fenêtre pour un estimateur à noyau, paramètre λ pour les splines de lissage ou encore nombre de noeuds pour les splines de régression. Le problème du choix du paramètre de lissage occupe une grande partie de la littérature depuis le début des années quatre-vingts.

Parmi les méthodes de sélection du paramètre de lissage, celles reposant sur la **validation croisée** ont donné lieu à une abondante littérature. Le principe de ces méthodes est de choisir le paramètre de lissage minimisant un critère des

moindres carrés dans lequel Y_i est ajustée par son estimation construite à l'aide de l'échantillon privé de la $i^{\text{ème}}$ observation. Dans le cadre des estimateurs à noyau de la régression citons l'article de Härdle et Marron (1985) qui obtiennent l'optimalité asymptotique de la largeur de fenêtre choisie par validation croisée. Des techniques analogues ont été étudiées pour le choix de largeur de fenêtre dans un estimateur à noyau de la densité: voir par exemple Rudemo (1982), Bowman (1984), Stone (1984) et Marron (1987). Craven et Wahba (1979) introduisent un critère de validation croisée pour le choix du paramètre λ dans une spline de lissage. Un critère de validation croisée généralisée est le plus souvent utilisé dans ce contexte dont les propriétés d'optimalité asymptotique ont été étudiées entre autres par Li (1985), Utreras (1988) et Wahba (voir Wahba, 1990).

D'autres techniques de sélection du paramètre de lissage ont été introduites: pour un estimateur à noyau citons les méthodes de réinjection (*cf.* Ruppert *et al.*, 1993 et Biau, 1999) ou encore celles reposant sur le bootstrap (*cf.* Mammen, 2000). 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) pour les estimateurs à noyau et celle de van der Linde (2000) pour les splines de lissage.

6 Réduction de la dimension

Les vitesses de convergence optimales pour un estimateur non paramétriques de la régression ou de la densité sont, on l'a vu plus haut, liées entre autres à la dimension du vecteur des covariables ou du vecteur aléatoire dont on estime la densité. Ainsi, ces vitesses se "dégradent" lorsque la dimension est grande. Il s'agit du **fléau de la dimension** (curse of dimensionality) qui se traduit par la "raréfaction" des observations dans un voisinage d'un point x de \mathbb{R}^d (voir à ce sujet 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. Stone (1985) passe en revue les principaux modèles de régression de ce type: modèle additif (étudié dans le même article de Stone), modèle d'analyse de la variance avec composantes fonctionnelles (Stone, 1994), régression sur projections

révélatrices ou projection pursuit regression (Friedman et Stuetzle, 1981) et modèle additif généralisé (Winsberg et Ramsay, 1980, Hastie et Tibshirani, 1990). Stone (1985) compare ces modèles du point de vue de leur **flexibilité, dimensionnalité** et **interprétabilité**. Un des modèles les plus populaires est le modèle de régression additive qui bien que moins flexible que ses concurrents est très attractif par sa simplicité d'écriture et sa facilité d'interprétation. Dans ce cas, la fonction de régression s'écrit :

$$r(x) = \mu + r_1(x_1) + \dots + r_d(x_d), \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

avec :

$$\mathbb{E}r_j(X_j) = 0, \quad j = 1, \dots, d.$$

Différents estimateurs ont été proposés dans le cadre des modèles réduisant la dimensionnalité. Parmi les plus utilisés citons ceux reposant sur l'algorithme **back-fitting** (voir par exemple Breiman et Friedman, 1985) et, dans une moindre mesure, ceux reposant sur les splines de régression (Stone, 1985, Burman, 1991, entres autres). Nous renvoyons sur ce premier aspect à la revue bibliographique sélective de Pelegrina *et al.* (1996) présentant les principaux modèles de régression et les méthodes d'estimation s'y rattachant.

Un autre aspect de l'estimation non paramétrique multidimensionnelle est celui du choix de modèle. Il s'agit d'introduire des méthodes de sélection de modèles, non paramétriques purs ou non paramétriques de dimensionnalité inférieure par exemple : voir entre autres sur ce sujet Vieu (1994) et Sperlich *et al.* (2000).

7 Quelques sujets de recherche actuels

7.1 Estimation non paramétrique

Les estimateurs non paramétriques de la régression ou de la densité sont construits, de manière implicite, pour estimer des fonctions continues. Même si des résultats de convergence sont démontrés lorsque la continuité n'est pas assurée (voir par exemple Stone, 1977, Devroye et Wagner, 1979 et 1980), il est souvent nécessaire d'introduire des méthodes *ad hoc* pour estimer les points de discontinuités éventuels d'une courbe et/ou estimer une courbe discontinue. Godtliebsen *et al.* (1997) utilisent un estimateur construit par médianes locales pour estimer le contour (discontinu) d'une image. Wu et Chu (1993) adapte, dans le contexte de la régression, les estimateurs à noyau afin d'estimer un point de discontinuité ainsi que la hauteur de la discontinuité : la méthode 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. Couallier *et al.* (1997)

proposent un estimateur analogue d'un point de discontinuité (et de la hauteur du saut) pour l'intensité d'un processus de Poisson non homogène.

Pour plus de détails sur l'estimation non paramétrique de points de discontinuités, domaine qui donne lieu à l'heure actuelle à de nombreuses recherches et pour lequel plusieurs méthodes (dont l'utilisation d'ondelettes) sont étudiées, nous renvoyons aux exposés de J. Antoch, G. Grégoire et V. Couallier ci-dessous.

7.2 Estimation multidimensionnelle

On l'a vu ci-dessus, parmi les estimateurs les plus utilisés dans le cadre de modèles de réduction de la dimension (comme le modèle de régression additive par exemple) figurent principalement ceux reposant sur l'algorithme backfitting. Cependant, un des désavantages de ces estimateurs est qu'ils se prêtent assez difficilement à une analyse de leurs propriétés statistiques. Ainsi, peu de résultats théoriques existent pour ces estimateurs (notons cependant un travail récent de Mammen *et al.*, 1999, dans lequel des propriétés de convergence sont données pour une version de l'estimateur). On peut préférer alors utiliser un estimateur "direct" (tel que celui introduit par Stone, 1985) pour lequel on possède entre autres les vitesses de convergence.

Toujours dans le cadre du modèle additif, d'autres estimateurs directs des fonctions r_j reposant cette fois-ci sur la méthode du noyau ont été proposés vers le milieu des années quatre-vingt-dix. Ces estimateurs sont basés 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). L'étude de ces estimateurs (propriétés asymptotiques, choix de paramètres de lissage, *etc*) est un des centres d'intérêt actuels de l'estimation fonctionnelle. Leur utilisation pour des processus vérifiant des conditions de mélange (et donc dans le cas des séries chronologiques) a donné lieu à la thèse de C. Camlong (2000). De nombreux problèmes ouverts dans ce domaine sont autant de sujets pour le futur. Notons également que l'estimation fonctionnelle sous hypothèse de mélange est au centre de nombreux travaux actuels (voir à ce sujet Rio, 2000).

7.3 Variables fonctionnelles

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. Dans ce contexte, plusieurs méthodes utilisant une approche "discrète", adaptées le plus souvent du modèle linéaire, ont été abondamment utilisées dans la pratique (voir à ce

sujet Frank et Friedman, 1993). Il existe cependant une littérature concernant une approche “fonctionnelle” : cf. par exemple Hastie et Mallows (1993), Ramsay et Silverman (1997). D’autres auteurs ont montré l’utilité de cette seconde approche : voir Marx et Eilers (1999).

L’étude de cette approche est un des thèmes que nous souhaitons privilégier au sein du groupe de travail STAPH, comme cela a été souligné dans l’introduction. Elle fait appel à l’estimation fonctionnelle et à l’environnement fonctionnel de la statistique et se trouve ainsi au carrefour entre deux thématiques très présentes au Laboratoire de Statistique et Probabilités. Nous ne développons pas plus en détail cet aspect et renvoyons à l’exposé de H. Cardot et F. Ferraty (modèles pour variables fonctionnelles) et à celui de Y. Romain (approche fonctionnelle de la statistique).

Références

- Agarwal, G.G. et Studden, W.J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines. *Ann. Statist.*, **8**, 1307-1325.
- 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.
- Biau, G. (1999). Estimateurs à noyaux itérés : synthèse bibliographique. *Journal de la Société Française de Statistique*, **140**, 41-67.
- Bosq, D. (1970). Contribution à la théorie de l’estimation fonctionnelle. *Pub. Inst. Statist. Univ. Paris*, **19**, Fasc. 2 et 3.
- 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.W. et Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Oxford University Press, Oxford.
- Breiman, L. et Friedman, J.H. (1985). Estimating Optimal Transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.*, **80**, 580-619.
- Burman, P. Rates of convergence for the estimates of the optimal transformations of variables, *Ann. Statist.*, **19**, 702-723.

- 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.
- 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. (1977). Quelques propriétés de la méthode du noyau pour l'estimation non-paramétrique de la régression en un point fixé. *C. R. A. S.*, **285**, 289-292.
- Collomb, G. (1978). Estimation non paramétrique de la régression : régressogramme et méthode du noyau. *Publications du Laboratoire de Statistique et probabilités de l'Université de Toulouse*, **07-78**, 1-59.
- Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahr. und verw. Geb.*, **66**, 441-460.
- Collomb, G. et Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis. *Stoch. Proc. Appl.*, **23**, 77-89.
- 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.
- Craven, P. et Waha, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.*, **31**, 377-403.
- Devroye, L. (1979). The uniform convergence of the Nadaraya-watson regression function estimate. *Can. J. Statist.*, **6**, 179-191.
- 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 discrimination and regression function estimation. *Ann. Statist.*, **8**, 231-239.
- Eubank, R.L. (2000). Spline Regression. in *Smoothing and Regression, Approaches, Computation and Application*, M.C. Schimek ed., Wiley, New York.
- Fan, J. et Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.

- Frank, I.E. et Friedman, J.H. (1993). A statistical view of some chemometric regression tools (with doiscussion). *Technometrics*, **35**, 109-148.
- Friedman, J.H. et Stuetzle, W. (1993). Projection Pursuit Regression. *J. Amer. Statist. Assoc.*, **76**, 817-823.
- Gasser, T. et Müller, H.G. (1979). Kerne estimation of regression functions. in *Smoothing techniques for curve estimation*, Lecture Notes in Mathematics 757, New York: Springer, 23-68.
- Geffroy, J. (1975). Etude de la convergence du régressogramme. *Séminaire de l'Institut de Mathématiques de l'Université de Paris*.
- Godtliebsen, F., Spjøtvoll, E. et Marron, J.S. (1997). A nonlinear gaussian filter applied to images with discontinuities, *Nonparametric Statist.*, **8**, 21-43.
- Gu, C. (2000). Multivariate Spline Regression. in *Smoothing and Regression, Approaches, Computation and Application*, M.C. Schimek ed., Wiley, New York.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Härdle, W. (1991). *Smoothing Techniques with implementation in S*. Springer-Verlag.
- Härdle, W. et Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression estimation. *Ann. Statist.*, **13**, 1465-1481.
- 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.
- Hastie, T.J. et Tibshirani, R.J. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Huber, P.J. (1985). Projection Pursuit. *Ann. Statist.*, **13**, 435-475.
- Li, K.C. (1985). From Stein's unbiased risk estimates to the method of generalized cross-validation. *Ann. Statist.*, **13**, 1352-1377.
- van der Linde, A. (2000). Variance Estimation and Smoothing-Parameter Selection for Spline Regression. in *Smoothing and Regression, Approaches, Computation and Application*, M.C. Schimek ed., Wiley, New York.
- Linton, O.B. et Nielsen, J.B. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82**, 93-100.

- Loader, C.L. (1999). *Local Regression and Likelihood*, Springer-Verlag, New-York.
- Mammen, E. (2000). Resampling Methods for Nonparametric Regression. in *Smoothing and Regression, Approaches, Computation and Application*, M.C. Schimek ed., Wiley, New York.
- Mammen, E., Linton, O. et 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.
- 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.
- Nadaraya, E.A. (1964). On estimating regression. *Theory Probab. Appl.*, **9**, 141-142.
- Nason, G.P. et Siverman, B.W. (2000). Wavelets for Regression and Other Statistical Problems. in *Smoothing and Regression, Approaches, Computation and Application*, M.C. Schimek ed., Wiley, New York.
- Parzen, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.*, **35**, 1065-1076.
- Pelegriana, L., Sarda, P. et Vieu, P. (1996). On multidimensional nonparametric regression, **COMPSTAT**, *Proceedings in Computational Statistics*, Ed. A. Prat.
- Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis*. Springer-Verlag, Berlin.
- Rio, E. (2000). Théorie asymptotique de processus aléatoires faiblement dépendants. *Mathématiques et applications*, **31**, Springer.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.*, **27**, 832-837.

- Rudemo, M. (1982). Empirical choices of histograms and kernel density estimates. *Scand. J. Statist.*, **9**, 65-78.
- Ruppert, D., Sheather, S.J. et Wand, M.P. (1993). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, **90**, 1257-1270.
- Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Sperlich, S., Tjøstheim, D. et Yang L. (2000). Nonparametric estimation and testing of interaction in additive models. Preprint.
- 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 estimates. *Ann. Statist.*, **12**, 1285-1297.
- Stone, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, **13**, 689-705.
- Stone, C.J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.*, **22**, 118-184.
- Tjøstheim, D. et 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.
- Utreras, F. (1988). Convergence rates for multivariate smoothing spline functions. *J. Approx. Theor.*, **52**, 1-27.
- 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.

- Wand, M.P. et Jones, M.C. (1995). *Kernel Smoothing*, Chapman & Hall, London.
- Watson, G.S. (1964). Smooth regression analysis. *Sankhya, Ser. A*, **26**, 359-472.
- Winsberg, S. et Ramsay, J.O. (1980). Monotonic transformations to additivity using splines. *Biometrika*, **67**, 669-674.
- Wu, J.S. et Chu, C.K. (1993). Kernel-type estimators of jump points and values of a regression function. *Ann. Statist.*, **21**, 1545-1566.

Modélisation pour variables fonctionnelles dans un contexte explicatif

Hervé CARDOT[†] et Frédéric FERRATY[‡]

[†] INRA Toulouse

Unité Biométrie et Intelligence Artificielle

31326 Castanet-Tolosan

cardot@toulouse.inra.fr

[‡] GRIMM et LSP

Université Paul Sabatier

118, Route de Narbonne

31062, Toulouse Cédex

ferraty@cict.fr

10/01/00

Ce compte-rendu présente une revue bibliographique des méthodes statistiques pour données fonctionnelles en insistant sur les modèles où la variable explicative est fonctionnelle.

1 Variables fonctionnelles

Les données fonctionnelles sont des données pour lesquelles l'individu statistique est

- une courbe (de croissance, de température, ...),
- une fonction (champs de pression,)

considéré comme la réalisation d'une variable aléatoire à valeurs dans un **espace fonctionnel** de dimension **infinie** tel que $L^2[0, 1]$, l'espace des fonctions de carré intégrable définies sur $[0, 1]$ de carré intégrable. La terminologie "données fonctionnelles" semble être due à Jim Ramsay.

Plus généralement on considérera un espace de fonctions H supposé de Hilbert (espace complet muni d'un produit d'un produit scalaire) séparable (et qui admet une base dénombrable).

1.1 Exemples

On trouve les premiers exemples de données considérées comme des fonctions en climatologie: Obhukov (1960), Holmstrom (1963), Buell (1971), ... étudient les grandes composantes (les grands modes de variabilité) des champs de pression sur

la surface de la Terre à l'aide des EOF (Empirical Orthogonal Functions). Petit à petit, tous les domaines scientifiques où apparaissent des trajectoires semblent avoir utilisé ces méthodes: la génétique (Molenar & Boomsma 1987, Kirkpatrick & Heckman 1989, ...), l'économie (Deville, 1974), la biologie, la chimie, ...

Le livre récent "Functional Data Analysis" de Ramsay & Silverman (1997) illustre de manière quasi-exhaustive l'ensemble des méthodes statistiques utilisées pour traiter ce type de données. Chronologiquement, ces méthodes ont d'abord été des extensions naturelles des la statistique multidimensionnelle :

- ACP (Dauxois & Pousse 1976, Dauxois, Pousse et Romain 1982,
- Classification (Michel-Briand & Escouffier 1994)
- Analyse Canonique (Leurgans *et al.* 1993)
- Analyse Discriminante (Hastie, Buja, Tibshirani, 1995)

Plus récemment, des extensions des modèles classiques de la statistique ont également été proposées: modèle linéaire fonctionnel, processus autorégressifs hilbertiens, modèle linéaire généralisé fonctionnel,

2 La décomposition de Karhunen-Loeve

L'approche statistique descriptive présentée aujourd'hui repose essentiellement sur la décomposition de Karhunen-Loeve qui permet d'obtenir une *représentation linéaire optimale* au sens de la variance par projection de X sur un espace de dimension finie. L'intérêt d'une telle décomposition est quelle reste valide sous des hypothèses "minimales" sur X (pas d'hypothèses liées à la stationnarité notamment). L'aspect "fréquence" sera présenté dans une prochaine séance de travail par A. Boudou. Les données qualitatives ne seront pas abordées.

Soient T un intervalle compact de \mathbb{R} et $X(\omega) = \{X(t, \omega), t \in T\}$ une fonction aléatoire définie sur l'espace probabilisé (Ω, A, P) et à valeur dans $L^2(T)$. On note $\mu(t)$ la moyenne ($\mu(t) = E(X(t))$) et

$$\gamma(s, t) = \mathbb{E} \{(X(t) - \mu(t))(X(s) - \mu(s))\}$$

la fonction de covariance. On suppose $\mathbb{E} \|X\|^2 < +\infty$, c-à-d. $\int \gamma(t, t) dt < +\infty$.

La decomposition de K-L est la représentation linéaire optimale de X suivante :

$$X(w, t) = \mu(t) + \sum_{j=1}^{+\infty} \eta_j(w) \phi_j(t), \quad \text{en moyenne quadratique} \quad (1)$$

uniformément en t si les trajectoires de X sont continues. Si de plus X est gaussien, l'égalité est presque sûre. Les fonctions ϕ_j sont choisies orthonormées et sont les solutions de l'équation intégrale :

$$\int_T \gamma(s, t) \phi_j(t) dt = \lambda_j \phi_j(s) \quad (2)$$

où les valeurs propres sont classées par ordre décroissant $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq \dots \geq 0$. Les variables aléatoires réelles

$$\eta_j = \int_T (X(t) - \mu(t)) \phi_j(t) dt \quad (3)$$

sont centrées, non corrélées deux à deux et de variance $Var(\eta_j) = \lambda_j$. Les valeurs propres λ_j représentent donc la variance de la projection de X sur l'espace engendré par la fonction ϕ_j . Ces fonctions sont appelées "mode of variation" ou "empirical orthogonal function" par les météorologues. La **représentation linéaire optimale** de X dans un **espace affine de dimension q** au sens de la variance est donc:

$$X_q(w, t) = \mu(t) + \sum_{j=1}^q \eta_j(w) \phi_j(t). \quad (4)$$

Cette décomposition se généralise pour des variables X à valeurs dans un espace de Hilbert H séparable muni du produit scalaire $\langle \cdot, \cdot \rangle$ (Dauxois et Pousse, 1976). La v.a X est dite variable aléatoire hilbertienne. On va supposer qu'elle est du second ordre. L'espérance de X , notée μ , est alors définie par :

$$\mathbb{E} \langle X, x \rangle = \langle \mu, x \rangle, \quad \forall x \in H,$$

et l'opérateur de covariance vérifie :

$$\langle \Gamma x, y \rangle = \mathbb{E} (\langle X - \mu, x \rangle \langle X - \mu, y \rangle), \quad \forall (x, y) \in H \times H.$$

Supposons X centré ($\mu = 0$), l'opérateur Γ s'écrit

$$\Gamma = \mathbb{E} \langle X, \cdot \rangle X \quad (5)$$

$$= \sum_{j=1}^{+\infty} \lambda_j \langle \phi_j, \cdot \rangle \phi_j \quad (6)$$

C'est un opérateur auto-adjoint, non négatif ($\lambda_j \geq 0$) et nucléaire ($\sum_{j=1}^{+\infty} \lambda_j < +\infty$).

Dans le cas particulier où $H = \mathbb{R}^p$, la décomposition de KL est l'ACP classique.

3 Estimation dans l'Analyse en Composantes Principales

On observe un échantillon X_1, \dots, X_n de v.a.h i.i.d de même loi que la v.a.h X . À partir de maintenant, X est supposée centrée.

L'ACP de X repose sur l'estimation des éléments propres de Γ , dont l'estimateur naturel est :

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^n \langle X_{i,\cdot}, \cdot \rangle X_i \quad (7)$$

Ensuite on en déduit

$$\hat{\phi}_1 = \text{Argmax} \langle \Gamma_n \phi, \phi \rangle, \quad \text{s.c. } \|\phi\| = 1$$

et

$$\hat{\lambda}_1 = \langle \Gamma_n \hat{\phi}_1, \hat{\phi}_1 \rangle.$$

Le processus est ensuite itéré sous contrainte d'orthogonalité entre les $\hat{\phi}_j$.

Il existe de très nombreux articles sur l'ACP de variables fonctionnelles et d'autres manières d'estimer les éléments propres de Γ (les sieves par exemple, Antoniadis & Beder, 1989).

3.1 Convergence

Deville (1974)	convergence des éléments propres de l'approximation de l'opérateur de covariance empirique sur une base de Fourier
Dauxois, Pousse, Romain (1982)	Formalisation dans un espace de Hilbert général, lois asymptotiques des éléments propres et des projecteurs.
Besse (1991)	Convergence <i>p.s</i> de l'approximation spline en présence de trajectoires discrétisées.
Bosq (1991)	vitesse de convergence (mq, p.s) pour données corrélées
...	...

Les outils sont ceux de la théorie des opérateurs (compacts), de l'analyse fonctionnelle et les outils probabilistes pour v.a.h (loi des grands nombre, théorème central limite et inégalités exponentielles).

3.2 Extensions

3.2.1 Pénalisation

L'estimation des vecteurs propres en pénalisant à l'aide d'un critère de pénalité noté J a été initialement proposée afin d'obtenir des estimateurs "régulier" (esthétiques). Si par exemple $J(\phi) = \|\phi''\|^2$ on pénalise l'écart à une fonction affine et donc les oscillations de l'estimateur.

Rice & Silverman (1991) ont proposé l'estimateur

$$\max_{\|\phi\|=1} \langle \Gamma_n \phi, \phi \rangle - \rho J(\phi) \quad (8)$$

Une autre approche basée sur une modification de la norme a été proposée par Silverman (1996) :

$$\max_{\|\phi\|_\rho=1} \langle \Gamma_n \phi, \phi \rangle, \quad \|\phi\|_\rho^2 = \|\phi\|^2 + \rho J(\phi). \quad (9)$$

Pourquoi pénaliser ?

- Critère esthétique (Besse & Ramsay, 1986).
- Biais positif systématique de l'estimation de la première valeur propre si on ne pénalise pas (Rice & Silverman, 1991). D'après l'inégalité de Jensen :

$$\lambda_1 = \|\mathbb{E}\Gamma_n\| \leq \mathbb{E}\hat{\lambda}_1 = \mathbb{E}\|\Gamma_n\|$$

- Développement asymptotique de l'erreur quadratique d'estimation qui montre que le lissage améliore (sous certaines hypothèses) cette erreur (Pezzulli & Silverman 1993, Silverman 1996, Cardot 2000).

3.2.2 Effets paramétriques

On peut distinguer deux grands types d'effets "paramétriques" (ce terme n'est pas pris au sens de l'exposé de P. Sarda, j'entend par effet paramétrique un effet qui n'est pas une fonction). Les premiers ont été introduits par Besse & Ramsay (1986) et avaient pour objectif d'éliminer de l'analyse les composantes triviales des trajectoires X_i telles que la saisonnalité ou la tendance par l'application au préalable d'opérateurs différentiels qui jouent le rôle de filtres. Si on applique par exemple l'opérateur de dérivation $F = D^2$ on élimine la tendance linéaire aux trajectoires X_i , si on applique $F = D + D^3$ on élimine la composante périodique $a + \cos(t) + \sin(t)$ des trajectoires. L'ACP est ensuite réalisée sur les données filtrées $FX_i, i = 1, \dots, n$.

Une seconde approche proposée par Silverman (1995) consiste à introduire des effets paramétriques afin de "recaler" les courbes par rapport à l'origine. En supposant les trajectoires périodiques, le modèle s'écrit alors :

$$X(t + \tau(\omega), \omega) = \mu(t) + \sum_{j=1}^{+\infty} \eta_j(\omega) \phi_j(t).$$

Il a ensuite généralisé cette écriture afin de pouvoir tenir compte de covariables. Le modèle s'écrit alors :

$$G_{\theta}^{-1} X = \mu + \sum_{j=1}^{+\infty} \eta_j \phi_j.$$

L'opérateur G est indexé par le vecteur de paramètre d'intérêt θ .

3.3 Discrétisation

La discrétisation est un problème important car les courbes observées sont toujours discrétisées et il faut donc approximer les trajectoires ou l'opérateur de covariance. On suppose à partir de maintenant que les trajectoires X_i sont observées en des points de mesure $0 \leq t_{i1} < t_{i2} < \dots < t_{in_i} \leq 1$, qui peuvent être différents d'une courbe à l'autre.

Trois grandes familles de méthodes existent. La plus ancienne est basée sur une approximation de l'équation intégrale (??). Elle suppose que la discrétisation est la même pour toutes les courbes. Une autre approche repose sur l'approximation des trajectoires dans un espace de fonctions puis du calcul de l'opérateur de covariance empirique des trajectoires approximées. Enfin plus récemment, des méthodes basées sur l'interprétation de l'équation (??) comme un modèle linéaire à effets aléatoires sont apparues.

3.3.1 Par approximation de l'équation intégrale

(Buell 1971, Castro *et al.* 1986, Rice & Silverman 1991, ...)

L'approximation est basée sur une méthode de quadrature :

$$\int_T \gamma(t_j, t) \phi_j(t) dt \approx \sum_k w_k \gamma(t_j, t_k) \phi_j(t_k)$$

3.3.2 Par approximation des trajectoires

On prend une base de fonction $\{B_1, \dots, B_k\}$ et on décompose

$$\hat{X}_i = \sum_{j=1}^q \beta_{ij} B_j(t)$$

afin d'obtenir une représentation X_i dans un espace de fonctions. On estime ensuite les couples (λ_j, ϕ_j) à partir de

$$\hat{\Gamma}_n = \frac{1}{n} \sum_{i=1}^n \langle \hat{X}_i, \cdot \rangle \hat{X}_i.$$

- Deville (1974), séries de Fourier,
- Besse & Ramsay (1986), fonction splines,
- Jones & Rice (1991), méthode du noyau,
- Besse, Cardot, Ferraty (1997), splines de régression avec pénalisation
- ...

3.3.3 Modèles linéaires mixtes: EM

(James & Hastie 1999, Rice & Wu 1999).

$$X_i(t_{ij}) \approx \sum_{m=1}^p \mu_k B_m(t_{ij}) + \sum_{m=1}^q \eta_{im} B_m(t_{ij}). \quad (10)$$

Les premiers termes du modèles sont les effets fixes (la moyenne), les seconds les effets aléatoires associés à la décomposition de KL.

Cette dernière approche semble pouvoir traiter les cas les plus défavorables tels que les trajectoires partiellement observées en basant l'estimation sur l'algorithme EM sous contraintes.

4 Analyse Canonique et Analyse Discriminante pour données fonctionnelles

Des extensions de l'analyse canonique (Leurgans, Moyeed, Silverman 1993) et de l'analyse discriminante (Hastie, Buja, Tibshirani 1995) ont également été proposées.

La grande nouveauté est que la **pénalisation** dans l'estimation n'est plus d'ordre esthétique, elle **devient nécessaire** pour que les estimateurs soient consistants (voir deuxième partie de l'exposé). En effet, pour ces problèmes, l'estimateur fait intervenir l'inverse (généralisé) de Γ_n et la norme de cet opérateur tend très rapidement vers l'infini avec n (Γ nucléaire implique $\lambda_n = o(n^{-1})$). Il devient donc nécessaire de contrôler la norme de cet opérateur inverse, ce contrôle pouvant être exercé

- par le paramètre de lissage associé à la pénalisation.
- par le choix de la dimension de l'ACP.

Le sections suivantes concernent les différentes modélisations statistiques recensées dans la littérature lorsqu'on dispose d'une variable réponse et d'une ou plusieurs variable(s) explicative(s) avec au moins l'une d'entre elles de nature fonctionnelle (c'est-à-dire une variable aléatoire fonctionnelle). Bien entendu, sur un sujet aussi vaste, il est prétentieux et illusoire de prétendre à une quelconque exhaustivité. Néanmoins, nous nous sommes efforcés de présenter le spectre le plus large possible des méthodes rentrant dans ce cadre d'étude.

5 Modèles linéaires pour variables fonctionnelles

L'approche fonctionnelle que nous aborderons ultérieurement s'inspire fortement de ce qui a été proposé de manière heuristique concernant le traitement des données fonctionnelles. Ici, le mot heuristique n'est pas péjoratif; il signifie simplement que les méthodes employées n'ont pas été justifiées théoriquement alors que d'un point de vue pratique, elles offrent des possibilités intéressantes. Ainsi, dans un premier temps, nous allons présenter les méthodes statistiques adaptées au cas de données fonctionnelles discrétisées pour pouvoir ensuite faire le lien avec la modélisation purement fonctionnelle.

5.1 Analyse de variables fonctionnelles discrétisées

L'individu statistique est une courbe (croissance, température, spectrométrie, ...) que l'on observe aux p points de mesure t_1, t_2, \dots, t_p ; on dispose de n trajectoires discrétisées $\mathbf{X}_i = \{X_i(t_1), \dots, X_i(t_p)\}$ pour $i = 1, \dots, n$, considérées comme étant n réalisations d'un vecteur aléatoire $\mathbf{X} = {}^t(X(t_1), \dots, X(t_p))$. Intéressons-nous alors au modèle de régression multiple d'une variable aléatoire réelle (v.a.r.) Y sur les p régresseurs $X(t_1), \dots, X(t_p)$. Le modèle s'écrit :

$$\mathbb{E}(Y/\mathbf{X}) = \sum_{j=1}^p \alpha_j X(t_j) \iff \mathbb{E}(Y/\mathbf{X}) = \sum_{j=1}^p \alpha(t_j) X(t_j) \iff \mathbb{E}(Y/\mathbf{X}) = {}^t \boldsymbol{\alpha} \mathbf{X}$$

où $\boldsymbol{\alpha} = {}^t(\alpha(t_1), \dots, \alpha(t_p))$.

Estimations de $\boldsymbol{\alpha}$ (Frank et Friedman, 1993)

On dispose de n observations $(X_i, Y_i)_{i=1, \dots, n}$ identiquement et indépendamment distribuées suivant la loi du couple (\mathbf{X}, Y) . Posons $\boldsymbol{\Gamma} = \mathbb{E}(\mathbf{X}^t \mathbf{X})$ et $\boldsymbol{\Delta} = \mathbb{E}(Y \mathbf{X})$. On a alors :

$$\boldsymbol{\Delta} = \boldsymbol{\Gamma} \boldsymbol{\alpha} \implies \boldsymbol{\alpha} = \boldsymbol{\Gamma}^{-1} \boldsymbol{\Delta}.$$

De cette équation, on peut déduire les méthodes d'estimation de $\boldsymbol{\alpha}$ suivantes.

* Méthode 1: *Ordinary Least Squares*

Il suffit de remplacer $\boldsymbol{\Delta}$ (resp. $\boldsymbol{\Gamma}$) par $\boldsymbol{\Delta}_n = \frac{1}{n} \sum_{i=1}^n Y_i \mathbf{X}_i$ (resp. $\boldsymbol{\Gamma}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^t \mathbf{X}_i$) pour en déduire l'estimateur noté $\boldsymbol{\alpha}_{OLS}$ de $\boldsymbol{\alpha}$ défini par :

$$\boldsymbol{\alpha}_{OLS} = \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\Delta}_n,$$

et on a :

$$\alpha_{OLS} = \arg \min_{\alpha \in \mathbb{R}^p} \left\{ \sum_{i=1}^n (Y_i - {}^t \alpha \mathbf{X}_i)^2 \right\}$$

Problème: Γ (et donc Γ_n) est souvent mal conditionnée, ce qui nous conduit à la méthode suivante, version “robustifiée” de la méthode *OLS*.

* Méthode 2: Ridge Regression

Pour améliorer α_{OLS} , il suffit d’avoir un meilleur conditionnement de Γ_n . Un nouvel estimateur de α , noté α_{RR} est alors défini par

$$\alpha_{RR} = (\Gamma_n + \lambda \mathbf{I})^{-1} \Delta_n,$$

et on a :

$$\alpha_{RR} = \arg \min_{\alpha \in \mathbb{R}^p} \left\{ \sum_{i=1}^n (Y_i - {}^t \alpha \mathbf{X}_i)^2 + \lambda {}^t \alpha \alpha \right\}$$

Ainsi, Γ_n est remplacé par $\Gamma_n + \lambda \mathbf{I}$, en contre-partie, il y a introduction d’un paramètre λ .

* Méthode 3: Principal Component Regression

Cette méthode est basée sur la décomposition spectrale de Γ_n ; soit \mathbf{U}_K la matrice contenant les K vecteurs propres de Γ_n associés à ses K plus grandes valeurs propres :

$$\alpha_{PCR} = ({}^t \mathbf{U}_K \Gamma_n \mathbf{U}_K)^{-1} {}^t \mathbf{U}_K \Delta_n$$

et on a :

$$\alpha_{RR} = \arg \min_{\alpha \in \mathbb{E}^K} \left\{ \sum_{i=1}^n (Y_i - {}^t \alpha \mathbf{X}_i)^2 \right\}$$

avec \mathbb{E}^K sous-espace vectoriel de dimension K de \mathbb{R}^p .

Remarque : introduction du paramètre K ($\ll p$).

5.2 Analyse de variables fonctionnelles non discrétisées

5.2.1 Régression d’une v.a. réelle Y sur une v.a. fonctionnelle \mathcal{X} : “Functional linear models for scalar response”

Soit H un espace (fonctionnel) de dimension infinie et \mathcal{B}_H une tribu de H . De manière formelle, \mathcal{X} est dite variable aléatoire fonctionnelle si et seulement si :

$$\mathcal{X} : \begin{cases} (\Omega, \mathcal{A}, P) & \longrightarrow (H, \mathcal{B}_H) \\ \omega & \longrightarrow \mathcal{X}(\omega, \cdot) \end{cases}$$

a) Modélisation :

- Ecriture 1 (Hastie et Mallows, 93, Ramsay et Silverman, 97, Cardot *et al.*, 2000) :

$$\mathbb{E}(Y / \{\mathcal{X}(t) ; t \in T\}) = \int_T \alpha(t) \mathcal{X}(t) dt \quad (\mathcal{E}_1)$$

ou bien

$$Y = \int_T \alpha(t) \mathcal{X}(t) dt + \varepsilon$$

ε étant centré, indépendant de \mathcal{X} .

- Ecriture 2 (Cardot *et al.*, 99) :

$$\mathbb{E}(Y / \{\mathcal{X}(t) ; t \in T\}) = A(\mathcal{X}) \quad (\mathcal{E}_2)$$

où A est un opérateur linéaire continu de H (espace de Hilbert séparable) dans \mathbb{R} .

- Une équation importante : soit $\Gamma = \mathbb{E}(\langle \mathcal{X}, \cdot \rangle \mathcal{X})$ et $\Delta = \mathbb{E}(\langle \mathcal{X}, \cdot \rangle Y)$. Alors, il est facile de voir que l'on a :

$$\Delta = A\Gamma \quad (\Leftrightarrow \quad \forall x \in H, \Delta(x) = \langle \alpha, \Gamma(x) \rangle).$$

Considérons maintenant un échantillon de taille n , $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$, et donnons les versions empiriques de Γ et Δ :

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^n \langle \mathcal{X}_i, \cdot \rangle \mathcal{X}_i \quad \text{et} \quad \Delta_n = \frac{1}{n} \sum_{i=1}^n \langle \mathcal{X}_i, \cdot \rangle Y_i$$

b) Estimation :

b.1) **Estimation du coefficient fonctionnel α (modèle \mathcal{E}_1)**

- Splines de lissage (Hastie et Mallows, 93, Ramsay et Silverman, 97) :

$$\alpha_{SL} = (\Gamma_n + \lambda \mathbf{N})^{-1} \Delta_n$$

et on a :

$$\alpha_{SL} = \arg \min_{\alpha \in \mathcal{W}_m} \left\{ \sum_{i=1}^n \left(Y_i - \sum_{j=1}^p \alpha(t_j) \mathcal{X}_i(t_j) \right)^2 + \lambda \int_T [\alpha^{(m)}(s)]^2 ds \right\}$$

où $\mathcal{W}_m = \left\{ f \in H / f \in \mathcal{C}^{m-1} \text{ et } \int_T [f^{(m)}(t)]^2 dt \right\}$ et $\alpha_{SL} = {}^t(\alpha_{SL}(t_1), \dots, \alpha_{SL}(t_p))$.

- B-splines : on pose $\alpha(t) = \sum_{l=1}^{k+q} a_l B_l(t) = {}^t \mathbf{B} \mathbf{a}$ et on définit un nouvel estimateur noté α_{BS} par :

$$\alpha_{BS} = \arg \min_{\alpha \in \mathcal{S}_{k,q}} \left\{ \sum_{i=1}^n (Y_i - \langle \alpha, \mathcal{X}_i \rangle)^2 \right\}$$

où $\mathcal{S}_{k,q}$ est le s.e.v. de H engendré par $\{B_1, \dots, B_{k+q}\}$.

- B-splines pénalisées :
 - * Marx et Eilers (1996) définissent l'estimateur α_{BSP1} par :

$$\alpha_{BSP1} = {}^t \mathbf{B} \mathbf{a}_{BSP1}$$

avec

$$\mathbf{a}_{BSP1} = \arg \min_{\mathbf{a} \in \mathbb{R}^{k+q}} \left\{ \sum_{i=1}^n (Y_i - \langle {}^t \mathbf{B} \mathbf{a}, \mathcal{X}_i \rangle)^2 + \lambda {}^t \mathbf{a} \mathbf{P}_m \mathbf{a} \right\}$$

- * Cardot, Ferraty et Sarda (2000) proposent α_{BSP2} qui est construit de la manière suivante :

$$\alpha_{BSP2} = {}^t \mathbf{B} \mathbf{a}_{BSP2}$$

où

$$\mathbf{a}_{BSP2} = \arg \min_{\mathbf{a} \in \mathbb{R}^{k+q}} \left\{ \sum_{i=1}^n (Y_i - \langle {}^t \mathbf{B} \mathbf{a}, \mathcal{X}_i \rangle)^2 + \lambda \int_T [({}^t \mathbf{B} \mathbf{a})^{(m)}]^2 \right\}$$

Ajoutons que pour α_{BSP2} , on obtient des résultats asymptotiques.

- b.2) **Estimation de l'opérateur A (modèle \mathcal{E}_2).** La construction de l'estimateur est basée sur la relation

$$\Delta = A\Gamma$$

- Régression fonctionnelle sur composantes principales :

H_K s.e.v de H engendré par les K fonctions propres associées aux K plus grandes valeurs propres de Γ_n :

$$A_n = \Delta_n \Pi_K (\Pi_K \Gamma_n \Pi_K)^{-1}$$

où Π_K est la projection orthogonale sur H_K .

- Régression fonctionnelle lisse sur composantes principales :
 - * Etape 1 : lissage des courbes. On pose pour $i = 1, \dots, n$:

$$\hat{\mathcal{X}}_i = \arg \min_{\tilde{\mathcal{X}} \in \mathcal{S}_{qk}} \left\{ \int_T (\tilde{\mathcal{X}}(t) - \mathcal{X}_i(t))^2 dt \right\}$$

- * Etape 2 : on réalise la régression fonctionnelle à partir des $\widehat{\mathcal{X}}_i$.
- Dans les deux cas, on obtient des résultats asymptotiques.

5.2.2 Régression d'une v.a. fonctionnelle \mathcal{Z} sur une ou plusieurs v.a. fonctionnelles

Soit $I_{\mathcal{Z}} \subset \mathbb{R}$ et \mathcal{Z} telle que

$$\mathcal{Z} : \begin{cases} (\Omega, A, P) & \longrightarrow L^2(I_{\mathcal{Z}}) \\ \omega & \longrightarrow Z(\omega, \cdot) \end{cases}$$

- a) “Functional linear models for functional response” (Ramsay et Silverman, 97)

- Régression sur **une** v.a. fonctionnelle \mathcal{X} : soit $I_{\mathcal{X}} \subset \mathbb{R}$ et \mathcal{X} telle que

$$\mathcal{X} : \begin{cases} (\Omega, A, P) & \longrightarrow L^2(I_{\mathcal{X}}) \\ \omega & \longrightarrow \mathcal{X}(\omega, \cdot) \end{cases}$$

Dans ce contexte, on peut s'intéresser au modèle suivant :

$$\mathbb{E}(\mathcal{Z}(s) / \{\mathcal{X}(t) ; t \in I_{\mathcal{X}}\}) = \int_{I_{\mathcal{X}}} \beta(s, t) \mathcal{X}(t) dt.$$

- Régression sur **p** v.a. fonctionnelles $\{\mathcal{X}^j\}_{j=1, \dots, p}$: le modèle précédent se généralise facielement à cette situation :

$$\mathbb{E}\left(\mathcal{Z}(s) / \bigcap_{j=1}^p \{\mathcal{X}^j(t) ; t \in I_{\mathcal{X}^j}\}\right) = \sum_{j=1}^p \int_{I_{\mathcal{X}^j}} \beta_j(s, t) \mathcal{X}^j(t) dt.$$

- b) Functional linear models for longitudinal data (“Varying time coefficient model”, Fan et Zhang, 99, Hoover *et al.*, 98). Soit $\mathbf{X}(t) = {}^t(\mathcal{X}^1(t), \dots, \mathcal{X}^p(t))$ et $\boldsymbol{\beta}(t) = {}^t(\beta_1(t), \dots, \beta_p(t))$. Lorsqu'on veut expliquer \mathcal{Z} par \mathbf{X} , par analogie avec le modèle de régression linéaire multiple classique, on peut poser le modèle suivant :

$$Y(t) = {}^t \mathbf{X}(t) \boldsymbol{\beta}(t) + \varepsilon(t)$$

Remarque :

- On dit “longitudinal” car certaines trajectoires correspondent à des variables qualitatives,
- Il existe quelques résultats asymptotiques.

c) Processus ARH (Bosq, 91, Besse et Cardot, 96)

Objectif : prédire une v.a. fonctionnelle \mathcal{X}_{n+1} à partir des v.a. fonctionnelles observées $\mathcal{X}_1, \dots, \mathcal{X}_n$ qui correspondent à un découpage en n morceaux d'une même trajectoire.

Modèle (ARH(1)) :

$$\mathcal{X}_j = A(\mathcal{X}_{j-1}) + \varepsilon_j, \quad j \in \mathbb{Z}.$$

Remarque :

- on a l'équation de base :

$$\Delta = A\Gamma$$

avec $\Gamma = \mathbb{E}(\langle \mathcal{X}_j, \cdot \rangle \mathcal{X}_j)$ et $\Delta = \mathbb{E}(\langle \mathcal{X}_j, \cdot \rangle \mathcal{X}_{j+1})$.

- Obtention de propriétés asymptotiques pour l'estimateur de A .

5.2.3 Décomposition d'une v.a. fonctionnelle X : ANOVA fonctionnelle

Parmi les modèles d'analyse de la variance fonctionnels (Fan et Zhang, 2000, Luo, 98, Luo *et al.*, 98, Ramsay et Silverman, 97), on peut donner comme exemple le modèle dit "Nested Functional ANOVA", lequel se formule de la façon suivante :

$$\mathcal{Z}_{ijk}(t) = a_i(t) + b_{ij}(t) + \mathcal{E}_{ijk}(t)$$

6 Modèles non-paramétriques pour variables fonctionnelles

- Régression non-paramétrique d'une v.a.r. Y sur une v.a. fonctionnelle \mathcal{X} (Ferraty et Vieu, 2000) :

$$A(x) = \mathbb{E}(Y/\mathcal{X} = x)$$

$$\hat{A}_{h_n}(x) = \frac{\sum_{i=1}^n y_i K\left(\frac{\|x_i - x\|_*}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\|x_i - x\|_*}{h_n}\right)}$$

où $\|\cdot\|_*$ désigne une semi-norme. Dans ce contexte.

Remarque : obtention de résultats asymptotiques.

- Régression non-paramétrique d'une v.a. fonctionnelle sur une autre v.a. fonctionnelle : processus ARH (Bosq, 83, Besse *et al.*, 99)

$$A(x) = \mathbb{E}(\mathcal{X}_i / \mathcal{X}_{i-1} = x)$$

$$\widehat{A}_{h_n}(x) = \frac{\sum_{i=1}^{n-1} x_{i+1} K\left(\frac{\|x_i - x\|_{L^2}}{h_n}\right)}{\sum_{i=1}^{n-1} K\left(\frac{\|x_i - x\|_{L^2}}{h_n}\right)}$$

en rajoutant éventuellement un lissage sur les \mathcal{X}_i .

Références

- Antoniadis, A. and Beder, J. (1989). Joint estimation of the mean and the covariance of a banach valued gaussian vector. *The Annals of Statistics*, 20:73–93.
- Besse, P. (1991). Approximation spline de l'analyse en composantes principales d'une variable aléatoire hilbertienne. *Annales de la Faculté des sciences de Toulouse*, 12:329–346.
- Besse, P. (1992). Pca stability and choice of dimensionality. *Statistics & Probability Letters*, 13:405–410.
- Besse, P. and Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics*, **24**, 467-487.
- Besse, P., Cardot, H., and Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Computational Statistics and Data Analysis*, **24**, 255-270.
- Besse, P., Cardot, H., and Stephenson, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scand. J. of Statist.*, **27**, 673-687.
- Besse, P. and Ramsay, J. (1986). Principal component analysis of sampled curves. *Psychometrika*, 51:285–311.
- Bosq, D. (1991). Modelization, non-parametric estimation and prediction for continuous time processes. In Roussas, G., editor, *Nonparametric Functional Estimation and Related Topics*, pages 509–529. Nato, Asi series.

- Bosq, D. (1983). Sur la prédiction non paramétrique de variables aléatoires et de mesures aléatoires. *Z. Wahr. verw. Geb.*, **64**, 541-553.
- Boudou, A. and Dauxois, J. (1994). Principal component analysis for a stationary random function defined on a locally compact group. *Journal of Multivariate Analysis*, 51:1-16.
- Bouhaddou, O., Obled, C., and Dinh, T. P. (1987). Principal component analysis and interpolation of stochastic processes: methods and simulation. *Journal of applied statistics*, 14(3):251-267.
- Buell, C. (1971). Integral equation representation for factor analysis. *Journal of Atmospheric Science*, 28:1502-1505.
- Buell, C. (1978). The number of significant proper functions of two-dimensional fields. *Journal of Applied Meteorology*, 17:717-722.
- Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional Linear Model. *Statist. & Prob. Letters*, **45**, 11-22.
- Cardot, H., Ferraty, F. et Sarda, P. (2000). Étude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel. *C. R. Acad. Sci. Paris*, t. 330, Série I, 501-504.
- Cardot, H. (2000). Nonparametric estimation of the smoothed principal components analysis of sampled noisy functions. *Nonparametric Statistics*, **12**, 503-538.
- Castro, P., Lawton, W., and Sylvestre, E. (1986). Principal modes of variation for processes with continuous samples curves. *Technometrics*, 28:329-337.
- Dalzell, C. and Ramsay, J. (1993). Computing reproducing kernels with arbitrary constraints. *SIAM, J. Sci. Comput.*, 14:511-518.
- Dauxois, J. et Pousse, A. (1976). *Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique*. Thèse de Doctorat, Université Paul Sabatier, Toulouse.
- Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function. *Journal of Multivariate Analysis*, 12:136-154.
- Deville, J.C. (1974). *Méthodes statistiques et numériques de l'analyse harmonique*. Ann Insee, 15.
- Fan, J. and Zhang, W. (1999). Statistical Estimation in Varying coefficient models. *Ann. Stat.*, **27**, 5, 1491-1518.

- Fan, J. and Zhang, W. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. Roy. Statist. Soc. Ser. B*, **62**. To appear.
- Ferraty, F. Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi normés. Note CRAS, t330, I, 139-142.
- Frank, I.E. and Friedman, J.H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, **35**, 109-148.
- Fukunaga, K. and Koontz, W. (1970). Representation of random processes using the finite karhunen-loeve expansion. *Information and Control*, 16:85–101.
- Hastie, T. and Mallows, C. (1993). A discussion of “A Statistical View of Some Chemometrics Regression Tools” by I.E. Frank and J.H. Friedman. *Technometrics*, **35**, 140-143.
- Hastie, T., Buja, A. and Tibshirani, R (1995). Penalized Discriminant Analysis. *Ann. Statist.*, **23**, 73-102.
- Holmström, I. (1963). On a method for parametric representation of the state of the atmosphere. *Tellus*, 15:127–149.
- Hoover, D., Rice, J., Wu, O. and Yang, L. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 4, 809-822.
- James, G. and Hastie, T. (1999). Principal components models for sparse functional data. Preprint.
- Jones, M. and Rice, J. (1992). Displaying the important features of large collections of similar curves. *The American Statistician*, 46:140–145.
- Kirkpatrick, M. and Heckman, N. (1989). A quantitative genetic model for growth, shape, reaction norms and other infinite-dimensional characters. *Journal of Mathematical Biology*, 27:429–450.
- Leurgans, S., Moyeed, R. and Silverman, B. (1993). Canonical Correlation Analysis when the Data are Curves. *Journal of the Royal Statistical Society, B*, **55**, 725–740.
- Loeve, M. (1945). Fonctions aléatoires du second ordre. *Comptes Rendus à l'Académie des Sciences, Série I Mathématique*, 220:469.
- Luo, Z. (1998). Backfitting in Smoothing Spline ANOVA. *Ann. Stat.*, **26**, 5, 1733-1759.

- Luo, Z., Wahba, G. and Johnson, D. (1998). Spatial-Temporal Analysis of Temperature Using Smoothing Spline ANOVA. *J. Climate*, **11**, 18-28.
- Marx, B.D. and Eilers P.H. (1996). Generalized Linear Regression on Sampled Signals with Penalized Likelihood. In: A. Forcina, G.M. Marchetti, R. Hatzinger, G. Galmacci (eds.) *Statistical Modelling. Proceedings of the 11th International workshop on Statistical modelling, Orvieto*.
- Michel-Briand, C. and Escouffier, Y. (1994). Segmentation d'un ensemble de courbes. *Revue de Statistique Appliquée*, 4:5-24.
- Molenaar, P. and Boomsma, D. (1987). The genetic analysis of repeated measures ii. the karhunen-loeve expansion. *Behavior Genetics*, 17:229-242.
- Obhukov, A. (1960). The statistically orthogonal expansion of empirical functions. *American Geophysical Union*, pages 288-291.
- Pezzulli, S. and Silverman, B. (1993). On smoothed principal components analysis. *Computational Statistics*, 8:1-16.
- Preisendorfer, R. and Mobley, C. (1988). *Principal Components Analysis in Meteorology and Oceanography*. Elsevier : Amsterdam.
- Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, B*, 53:539-572. with discussion.
- Ramsay, J.O and Silverman, B.W (1997). *Functional Data Analysis*. Springer-Verlag.
- Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, B*, 53:233-243.
- Rice, J. and Wu, C. (1999). Nonparametric mixed effects models for unequally sampled noisy curves. Preprint.
- Silverman, B. (1995). Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society, B*, **57**, 673-689.
- Silverman, B.W. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, **24**, 1-24.
- Yuan, S. and Bertram, H. (1992). Statistical data analysis of magnetic recording noise mechanisms. *IEEE transactions on Magnetics*, 28:84-92.

Sur et pour une approche fonctionnelle en Statistique

Yves ROMAIN

Laboratoire de Statistique et Probabilités

Université Paul Sabatier

118, Route de Narbonne

31062, Toulouse Cédex

romain@cict.fr

24/01/00 et 31/01/00

Dans ces exposés, on s'appuie sur trois exemples pour illustrer une approche "fonctionnelle stochastique" en Statistique :

- la Statistique Factorielle considérée dans un cadre opératoire (*cf.* les travaux de l'équipe fondée par Dauxois et Pousse),

- la Statistique Inférentielle sous formalisme hilbertien (*cf.* l'ouvrage de Small et Mac Leish),

- la Statistique Quantique (ou non commutative) (*cf.*, par exemple, l'article synthétique de Malley et Hornstein).

En s'interrogeant sur les notions clés et pivotales d'une telle approche, on dégage l'importance fondamentale des deux outils-concepts que sont d'une part la projection orthogonale, d'autre part la structure d'algèbre de von Neumann. Comme ces algèbres d'opérateurs sont engendrées par les projecteurs qu'elles contiennent, on fixe notre intérêt sur cette notion clé et on développe alors une argumentation pour une "Statistique par les projecteurs".

En guise de résumé sur ces propos, nous reprenons ci-dessous un extrait de Romain (1997) qu'on consultera pour les éléments bibliographiques cités :

"En fait, qu'elle soit uni ou multidimensionnelle, "la Statistique par les projecteurs" pourrait bien être un titre pour un cours moderne sur la Statistique tant cette notion intervient de façon primordiale dans cette discipline. Tout d'abord, remarquons que l'on possède maintenant un large éventail de connaissances des propriétés mathématiques de l'ensemble des projecteurs dans $\mathcal{L}(H)$ (ou, ce qui est équivalent, des sous-espaces fermés dans H) :

i) Sur le plan algébrique, on dispose de tout l'arsenal des propriétés dues à la structure de treillis (voir Nördstrom et von Rosen (1987) pour une première référence en dimension finie avec applications en Statistique, et Holland (1970) et Kalmbach (1983), déjà cités, pour des propriétés générales). Par exemple, le treillis est modulaire si, et seulement si, H est de dimension finie (il est orthomodulaire en dimension infinie) ; de plus, l'orthomodularité caractérise les espaces complets parmi les espaces préhilbertiens... (pour E, F (contenant E) et G sous-

espaces de H , la loi modulaire implique $E \wedge (F \vee G) = F \vee (E \wedge G)$ alors que la loi orthomodulaire implique $E \vee (E^\perp \wedge F) = F$.

ii) Sur le plan géométrique, l'étude de la position relative de deux sous-espaces a été menée sur plusieurs fronts complémentaires. Depuis les travaux précurseurs de Jordan (1896) dans \mathbb{R}^n , elle a donné lieu aux développements fondamentaux de Dixmier (1948), Davis (1958) et Halmos (1969) pour le cas de deux sous-espaces (puis aux prolongements au cas de n sous-espaces par Sunder (1988)). Ensuite, se sont greffés à cette étude comparative de deux projecteurs nombre de résultats connexes. Citons, par exemple, Lenard (1972) sur l'étendue numérique conjointe, Bhatia (1997) sur les liens avec la CS décomposition et les équations de Sylvester, Sano et Watatani (1994) sur l'étude des angles entre deux sous-Facteurs (de type II_1), ou encore Paskiewicz (1986) sur les définitions de divers types de convergences stochastiques dans les \mathcal{W}^* -algèbres. Enfin, et parallèlement, il y a eu aussi la contribution des statisticiens depuis Hotelling (1936) jusqu'à Dauvois et Pousse (1975), en passant par Gel'fand et Yaglom (1957) et Hannan (1961) (les travaux sur ce domaine concerne, bien sûr, l'Analyse Canonique).

(iii) Sur le plan stochastique, on sait depuis les travaux de Gleason (1957) probabiliser l'espace des projecteurs de $\mathcal{L}(H)$ (i.e. toute mesure μ finie σ -additive sur cet espace est de la forme $\mu_T(\pi) = \text{tr}(\pi T)$, où T est un opérateur à trace finie et π est un projecteur de $\mathcal{L}(H)$). De plus, il est important de rappeler le rôle essentiel que joue l'espérance conditionnelle en tant que projecteur en Statistique. Même si tout projecteur n'est pas, bien sûr, une espérance conditionnelle (voir, par exemple, le cas des fonctions estimantes), l'inverse est en revanche vrai et on dispose d'un grand nombre d'applications diverses plus ou moins classiques de cet opérateur (voir, par exemple, le traitement des données incomplètes à partir des vraisemblances conditionnelles dans Small et Mc Leish (1994)). Enfin, on sait obtenir les lois asymptotiques des suites des projecteurs propres des approximations par échantillonnage d'un opérateur compact autoadjoint (voir Dossou-Gbete et Pousse (1991)) ce qui permet de nombreuses applications tant théoriques (voir, par exemple, Romain (1979), Larrère (1994) ou Dauvois et al. (1993) pour des problèmes de tests ou de comparaison de deux populations...) que pratiques (voir, par exemple, Jeffries et al. (1985), Paulraj et al. (1993) en traitement du signal...).

(iv) Toutes ces propriétés, bien que sommairement abordées ici, doivent enfin être complétées sur un plan méthodologique. Nous concluons cette partie sur ce point. Ce qui nous semble le plus simple, voire pédagogique, à souligner ici est l'approche commune à de très nombreux cas en Statistique qui se ramènent à un processus de décomposition d'un espace en plusieurs composantes orthogonales et donc à considérer les projections associées naturellement à cette décomposition. Pour illustrer ce fait bien ancré dans la culture statistique, mais peut-être pas toujours appréhendé de façon fonctionnelle, citons quelques exemples variés tels

que :

— l'élimination des paramètres fantômes en estimation paramétrique et la réduction de la sensibilité des estimateurs par projection,

— la stratégie d'inférence conditionnelle à une statistique en projetant sur le sous-espace engendré par cette statistique,

— la décomposition d'un espace initial en deux sous-espaces, l'un dit informatif (ou d'intérêt), l'autre, orthogonal, dit non informatif (ou résiduel) ; on a de nombreux exemples en modèle linéaire, en traitement du signal multidimensionnel,...

— l'approximation d'un opérateur (ou d'un tableau de données) par une combinaison linéaire de projecteurs "suffisamment informative",

— ... Chacun pourrait continuer cette liste dans de nombreuses directions. En résumé, la projection orthogonale est une notion clef qui permet de simplifier tout un ensemble de problèmes statistiques dans une pensée algébrique, géométrique et fonctionnelle et qui prépare naturellement à des extensions très variées (tant sur le plan de la dimension infinie que dans le contexte "inhabituel" de la Statistique non commutative). "

Quelques références de base :

Dauxois, J. et Pousse, A. (1976). Les Analyses factorielles en calcul des Probabilités et en Statistique: essai d'étude synthétique. Thèse de Doctorat d'Etat, Université Paul Sabatier, Toulouse.

Malley, J.D. et Hornstein, J. (1993). Quantum statistical inference. *Statist. Sci.* 8, 4 433-457.

Romain, Y. (1997). Une introduction à l'approche fonctionnelle stochastique de la Statistique. *Publi. Labo. Stat. Proba.*, 02-97, 1-28, Université Paul Sabatier, Toulouse.

Small, C. G. et McLeish, D. L. (1994). *Hilbert Space Methods In Probability & Statistical Inference*. Wiley, New-York.

Produit de convolution de mesures spectrales

Alain BOUDOU

Laboratoire de Statistique et Probabilités

Université Paul Sabatier

118, Route de Narbonne

31062, Toulouse Cédex

boudou@cict.fr

14/02/00 et 21/02/00

Nous appuyant sur le fait que toute fonction aléatoire continue stationnaire est la transformée de Fourier d'une mesure aléatoire, qui la définit d'une façon biunivoque, nous nous proposons d'exprimer en fonction de la mesure aléatoire associée à une série stationnaire donnée $(Y_n)_{n \in \mathbb{Z}}$ la mesure aléatoire correspondant à une fonction aléatoire continue stationnaire $(X_t)_{t \in \mathbb{R}}$ telle que $X_n = Y_n$ pour tout n de \mathbb{Z} .

Pour cela nous développons des outils que nous nommons, par analogie avec la théorie classique des probabilités, "produit" et "produit de convolution" de mesures spectrales. Nous rappelons que tout opérateur unitaire peut s'exprimer comme intégrale stochastique d'une mesure spectrale et que "l'opérateur de décalage" est un opérateur unitaire qui joue un grand rôle dans l'étude de celui-ci.

L'introduction du concept de "famille de mesures aléatoires stationnairement corélées" permet d'associer une mesure spectrale à un ensemble de mesures aléatoires. Étant donné un opérateur unitaire U nous pouvons ainsi expliciter, d'une façon originale, la correspondance "opérateur unitaire - mesure spectrale" mentionnée plus haut en considérant la famille des mesures aléatoires associées aux séries stationnaires $(U^n X)_{n \in \mathbb{Z}}$.

Ensuite étant donné deux opérateurs unitaires U_1 et U_2 , qui commutent, on considère la famille des séries $(U_1^n \circ U_2^m(X))_{(n,m) \in \mathbb{Z}^2}$. De la commutativité de U_1 et U_2 on déduit que ces séries sont stationnaires et stationnairement corélées ce qui nous permet de définir une "famille de mesures aléatoires stationnairement corélées". La mesure spectrale associée à cette dernière, compte tenu de ses propriétés, est appelée "produit" de ε_1 et ε_2 , mesures spectrales respectivement associées à U_1 et à U_2 . Bien entendu l'image de ce produit par l'application "somme" est nommée "produit de convolution" de ε_1 et ε_2 . On montre que ce "produit de convolution" est la mesure spectrale correspondant à l'opérateur unitaire $U_1 \circ U_2$.

Ces concepts de "produit" et de "produit de convolution" concernent des mesures spectrales définies sur la tribu borélienne du groupe $[-\pi, \pi[$, nous les étendons à des mesures spectrales définies sur des tribus boréliennes de groupes

plus généraux.

Enfin, étant donné une mesure spectrale ε et f un homomorphisme de groupe nous définissons toutes les mesures spectrales dont l'image par f est ε . La solution générale de cette équation combine "solution particulière" et "solution de l'équation homogène".

Ces différents résultats nous permettent de résoudre le problème d'interpolation initial.

Références

- Azencot, R.; Dacunha-Castelle, D. (1984). *Séries d'observations irrégulières*. Masson, Paris.
- Boudou, A. (2000). Produit et produit de convolution de mesures spectrales. *Publi. labo. Stat. Proba. Univ. P. Sabatier*, Toulouse.
- Birman, M.; Solomyak, M. (1996). Tensor product of a finite number of spectral measures is always a spectral measures. *Integral Equations Operator Theory*, **24**, 179-187.
- Dacunha-Castelle, D.; Duflo, M. (1990). *Probabilités et statistiques 1. Problèmes à temps fixe*. Masson, Paris.
- Riesz, F.; Nagy, B.,SZ. (1968). *Leçons d'analyses fonctionnelle*. Gauthier-Villars, Paris.

The Geometrical Theory of Estimating Functions

Christopher SMALL
Waterloo University
cgsmall@uwaterloo.ca
28/02/00

By an *estimating function*, we shall mean a function

$$g : \Theta \times \mathcal{Y} \rightarrow \mathbb{R}^k$$

with finite first and second moments, in the sense that $E_\theta g(\theta, Y)$ and $E_\theta [g(\theta, Y)]^2$ are both finite for all $\theta \in \Theta$. We suppose that the estimating function is *unbiased* in the sense that

$$E_\theta g(\theta, Y) = 0 \tag{1}$$

for all $\theta \in \Theta$.

For example, suppose that $t(Y)$ is an estimator for θ . Defining $g(\theta, Y) = t(Y) - \theta$, we can see that the estimating function g is unbiased if and only if the estimator $t(Y)$ is unbiased. While every unbiased estimator for θ can be represented as the root of an unbiased estimating function, it is not the case that the root of a general unbiased estimating function is itself unbiased. The reason for this is that most estimating functions do not separate as a difference of the form $t(Y) - \theta$. The theory of estimating functions provides an extension of the methodology of best unbiased estimation to models which do not admit a best unbiased estimator.

Suppose θ is a real parameter, and that $g(\theta, Y)$ is a real-valued function of θ and Y that is differentiable with respect to θ . Denote the derivative of g with respect to θ by \dot{g} . We define the *Godambe efficiency* of g by

$$\text{eff}_\theta(g) = \frac{[E_\theta \dot{g}(\theta, Y)]^2}{E_\theta [g(\theta, Y)]^2}. \tag{2}$$

An estimating function g^* is said to be *Godambe efficient* if $\text{eff}_\theta(g^*) \geq \text{eff}_\theta(g)$ for all differentiable g and for all θ . The concept of Godambe efficiency was first proposed in Godambe (1960) and Durbin (1960). It is the extensive promotion and elaboration of the consequences of this definition by V. P. Godambe which leads us attach his name to the definition.

We shall simplify the notation $g(\theta, Y)$ by denoting it as $g(\theta)$ with the data Y suppressed. Let us consider the relationship between Godambe efficient estimating functions and the theory of best unbiased estimation. Suppose $t(Y)$ is

an unbiased estimator for θ , so that $g(\theta) = t(Y) - \theta$ is an unbiased estimating function. Then $\dot{g}(\theta) = -1$ and

$$\text{eff}_\theta(g) = \frac{1}{\text{Var}_\theta t(Y)}$$

So if $g^*(\theta) = t^*(Y) - \theta$ is Godambe efficient, it will follow that $\text{Var}_\theta t^*(Y) \leq \text{Var}_\theta t(Y)$ for any unbiased estimator $t(Y)$. We can conclude that $t^*(Y)$ is a best unbiased estimator in the classical sense.

References

- Chandrasekar, B. (1983). Contributions to the theory of unbiased statistical estimation functions. Ph.D. thesis, Univ. of Poona, Pune-7, India.
- Chandrasekar, B. and Kale, B. K. (1984). Unbiased statistical estimation functions in presence of nuisance parameter. *J. Statist. Plan. Inf.* **9** 45-54.
- Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econ. Theory* **2** 305-330.
- Durbin, J. (1960). Estimation of parameters in time series regression models. *J. Roy. Statist. Soc. Ser. B* **22** 139-53.
- Ferreira, P. E. (1982a). Multiparametric estimating equations. *Ann. Statist. Math.* **34A** 423-431.
- Ferreira, P. E. (1982b). Estimating equations in the presence of prior knowledge, *Biometrika*, **69**, 667 – 669.
- Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Camb. Phil. Soc.* **22** 700-725.
- Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208-1212.
- Greene, W. (1990). Multiple roots of the Tobit log-likelihood. *J. Econometrics* **46** 365-380.
- Hanfelt, J.J. and Liang, K.-Y. (1995). Approximate likelihood ratios for general estimating functions. *Biometrika* **82** 461-477.
- Heyde, C. C. (1997). *Quasi-Likelihood And Its Application*. Springer-Verlag, New York.
- Heyde, C. C. and Morton, R. (1998). Multiple roots in general estimating equations. *Biometrika* **85** 954-959.

- Huzurbazar, V. S. (1948). The likelihood equation, consistency and the maxima of the likelihood function. *Ann. Eugen.* **14** 185-200.
- Jensen, J. L. and Wood, A. T. A. (1999). Large deviation results for minimum contrast estimators. *Ann. Inst. Statist. Math.*. To appear.
- Kale (1962) An extension of Cramer-Rao inequality for statistical estimation functions. *Skand. Aktur.* **45**, 60-89.
- Kendall, M. G. (1951). Regression, structure and functional relationship – I. *Biometrika* **38**, 11-25.
- Li, B. (1993). A deviance function for the quasi-likelihood method. *Biometrika* **80** 741-753.
- Li, B. and McCullagh, P. (1994). Potential functions and conservative estimating functions. *Ann. Statist.* **22** 340-356.
- McCullagh, P. (1991). Quasi-likelihood and estimating functions. In *Statistical Theory and Modelling, in honour of Sir David Cox*. Edited by D. V. Hinkley, N. Reid, E. J. Snell. Chapman & Hall, London, 267-286.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall, London.
- McLeish, D. L. and Small, C. G. (1988). *The Theory and Applications of Statistical Inference Functions*. Springer Lecture Notes in Statistics **44**, Springer-Verlag, New York.
- McLeish, D. L. and Small, C. G. (1992). A projected likelihood function for semiparametric models. *Biometrika.* **79** 93-102.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Phil. Trans. Roy. Soc. London A* **185**, 71-110.
- Singh, A. C. and Mantel, H. J. (1998). Minimum chi-square estimating function and the problem of choosing among multiple roots. *Proc. Amer. Statist. Assoc.* 1998. To appear.
- Skovgaard, I. M. (1990). On the density of minimum contrast estimators. *Ann. Statist.* **18** 779-789.
- Small, C. G. and McLeish, D. L. (1994). *Hilbert Space Methods in Probability and Statistical Inference*. Wiley, New York.
- Small, C. G. and Yang, Z. (1999). Multiple roots of estimating functions. *Can. J. Statist.* **27**. To appear.

- Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74** 703-716.
- Wang, J. (1999). Nonconservative estimating functions and approximate quasi-likelihoods. *Ann. Inst. Statist. Math.* **51**. To appear.
- Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and Gauss-Newton method. *Biometrika* **61** 439-447.

Inférence statistique pour des estimateurs de discontinuités dans un cadre non paramétrique

Vincent COUALLIER

Laboratoire de Statistique et Probabilités

Université Paul Sabatier

118, Route de Narbonne

31062, Toulouse Cédex

coual@cict.fr

13/03/00

L'exposé traite des méthodes d'estimation non paramétriques pour des fonctions pouvant présenter des discontinuités. La localisation de ces sauts présente un intérêt en lui-même et les principaux résultats concernent divers types de convergence pour des estimateurs à noyau adaptés aux fonctions discontinues.

Nous présentons, de manière didactique, le problème de l'estimation non paramétrique de discontinuités pour les principales fonctions d'intérêt en statistique (densité, régression, intensité, hasard). Un résultat sur l'estimation d'une discontinuité dans la fonction d'intensité d'un processus de Poisson non homogène avec une application sur des données météorologiques est ensuite établi. Sont ensuite présentés des résultats de convergence presque complète et de normalité asymptotique d'un estimateur à noyau pour la localisation de discontinuité dans une fonction de densité à partir de données indépendantes. Puis nous adaptons nos résultats dans le cadre de la fonction de hasard de données de survie. Ces résultats sont étendus au cadre de processus strictement stationnaires α -mélangeant. La principale nouveauté réside dans le fait que ce problème d'estimation non paramétrique n'impose pas au paramètre de lissage les conditions usuelles pour l'obtention de résultats optimaux ou quasi-optimaux. Nous nous intéressons ensuite à l'influence de l'étape de localisation d'un saut sur les mesures d'erreurs quadratiques intégrées d'un estimateur à noyau d'une fonction de densité. Diverses simulations numériques de nos estimateurs de sauts sont présentées pour lesquels le choix du paramètre de lissage est discuté.

Références

- 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.

Couallier, V. (1999), Estimation non paramétrique d'une discontinuité dans une densité, *Comptes Rendus de l'Académie des Sciences, I*, **t.329**, 633-636.

Couallier, V. (1999), Nonparametric change point estimation for discontinuous density functions under dependence conditions, en révision pour *Statistical Inference for Stochastic Processes*.

Nonparametric estimation in null recurrent time series

Dag TJØSTHEIM

Department of Mathematics

University of Bergen

Johs. Brunsgate 12

5008 Bergen, Norvège

dagt@mi.uib.no

27/03/00

Work on nonparametric estimation has so far with very few exceptions been carried out in a stationary strongly mixing framework (see e.g. Robinson, 1983, Masry and Tjøstheim, 1995, and references therein). Recently asymptotics for processes with long range dependence have been covered (Robinson, 1997), but still no systematic theory exists for a nonstationary situation.

The main purpose of this work is to try to fill this gap by establishing a nonparametric estimation theory that can be used in a nonstationary environment, more precisely in the framework of null recurrent Markov chains, or possibly regime models including null recurrent states (see Karlsen and Tjøstheim, 2001). An essential tool is the split chain, which makes it possible to decompose the times series under consideration in independent and identical parts. A tail condition on the distribution of the recurrence time is introduced. This condition makes it possible to prove weak convergence results for sums of functions of the process depending on a smoothing parameter. These limit results are subsequently used to obtain consistency and asymptotic normality for local density estimators and for estimators of the conditional mean and the conditional variance. In contrast to the parametric case, the convergence rate is slower than in the stationary case, and it is directly linked to the tail behavior of the recurrence time.

With the single exception of the work by Yakowitz (1993) on consistency of nearest neighbour estimates, as far as we know, the estimation theory of null recurrent processes has been confined to the parametric case.

Independent of our work, however, nonparametric estimation has been considered in a random walk situation by Phillips and Park (1998), who use local time argument to derive asymptotic distribution and by Xia (1998), who in his doctoral thesis gives a proof of consistency in a transfer function case.

Asymptotics of parametric null recurrent (usually non-time series) models have been treated by Höpfner (1990, 1994), Höpfner *et al.* (1990), Kasahara (1982, 1984, 1985), Touati (1990), and we exploit some of their techniques. For two early contributions in the field we refer to Darling and Kac (1957) and Kallianpur and Robbins (1954).

In nonparametric situations series with large or very large sample sizes are required. Long series are becoming increasingly available, e.g. in finance and econometrics. There is therefore also a practical motivation behind our work. The particulars of this motivation are much the same as for the stationary case: it is desirable to have greater flexibility in the initial stage of modelling than that offered by a fixed parametric or semiparametric model, for example using nonparametric estimates as a guide in choosing a parametric (linear or nonlinear) model (see Myklebust *et al.*, 2000, for some more examples). We would like to mention very briefly potential implications for econometric time series modelling, though, since such series are often thought to be nonstationary. The kind of nonstationarity that has been built into the *parametric* econometric modelling has overwhelming been of linear unit-root type leading to ARIMA models and, in the multivariate case, to linear cointegration models. For such models a very considerable body of literature exists (*cf.* the review paper by Stock, 1994, Watson, 1994 and the book by Johansen, 1995). Asymptotic distributions are typically non-normal and the parameter estimates are super-efficient (Dickey and Fuller, 1979, Johansen, 1995). The need for models combining features of non-linearity and nonstationarity has been emphasized (see e.g. Granger and Hallman, 1991, Granger, 1995, Aparicio and Escribano, 1997), but no systematic estimation theory exists. Again, we believe that the class of null recurrent processes constitutes an adequate framework for posing such problems. Finally, it should be mentioned that there are challenging and interesting connections to attempts having been made to construct a nonlinear cointegration theory. We look at some of these in Karlsen *et al.* (2000).

References

- Aparicio, F.M. and Escribano, A. (1997). Searching for linear and nonlinear cointegration: a new approach, Working paper 97-65, Universidad Carlos III de Madrid, Statistics and Economics Series.
- Darling, D.A. and Kac, M. (1957). On occupation times for Markoff processes. *Transactions Amer. Math. Soc.*, **84**, 444-458.
- Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimation for autoregressive time series with unit root. *Econometrica*, **49**, 1057-1072.
- Granger, C.W.J. and Hallman, J. (1991). Long memory series with attractors. *Oxford Bulletin of Economics and Statistics*, **53**, 11-26.
- Granger, C.W.J. (1995). Modelling nonlinear relationships between extended-memory variables. *Econometrica*, **63**, 265-279.

- Höpfner, R. (1990). Null recurrent birth and death processes, limits for certain martingales and local asymptotic mixed normality. *Scand. J. Statist.*, **17**, 201-215.
- Höpfner, R. (1994). Estimating a parameter in a birth-and-death process model. *Statistics and Decisions*, **12**, 149-160.
- Höpfner, R., Jacod, J. and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Th. Rel. Fields*, **86**, 105-129.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegrated vector Autoregressive Models*. Oxford University Press, Oxford.
- Kallianpur, G. and Robbins, R. (1954). The sequence of sums of independent random variables. *Duke Math. J.*, **21**, 285-307.
- Karlsen, H.A., Myklebust, T. and Tjøstheim, D. (2000). Nonparametric estimation in a nonlinear cointegration type model. *Sonderforschungsbereich 373*, **33**, Humboldt-Universität zu Berlin.
- Karlsen, and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. To appear in *Ann. Statist.*
- Kasahara, Y. (1982). A limit theorem for slowly increasing occupation times. *Ann. Prob.*, **10**, 728-736.
- Kasahara, Y. (1984). Limit theorems for Lévy processes and Poisson point processes and their applications to Brownian excursions. *J. Math. Kyoto Univ.*, **24**, 521-538.
- Kasahara, Y. (1985). A limit theorem for sums of random number of i.i.d. random variables and its application to occupation times of Markov chains. *J. Math. Soc. Japan*, **37**, 197-205.
- Masry, E. and Tjøstheim, D. (1995), Nonparametric estimation and identification of nonlinear ARCHtimes series. *Econometric Theory*, **13**, 214-252.
- Myklebust, T., Karlsen, H.A. and Tjøstheim, D. (2000), A Markov chain characterization of unit-root processes and the problem of nonlinear cointegration. In preparation.
- Phillips, P.C.B. and Park, J. (1998), Nonstationary density estimation and kernel autoregression. *Cowles Foundation Discussion Paper No. 1181*, Yale University.

- Robinson, P.M. (1983), Nonparametric estimators for time series. *J. Time Ser. Anal.*, **4**, 185-207.
- Robinson, P.M. (1997), Large sample inference for nonparametric regression with dependent errors. *Ann. Statist.*, **25**, 2054-2083.
- Stock, J.H. (1994). Unit roots, structural breaks and trends. In *Handbook of Econometrics*, **4**, 2740-2843. Engle, R.F. and McFadden, D.L. editors. Elsevier, AMsterdam.
- Touati, P.A. (1990). Loi fonctionnelle du logarithme itéré pour les processus de Markov récurrents. *Ann. Prob.*, **18**, 140-159.
- Watson, M.W. (1994). Vector autoregression and cointegration. In *Handbook of Econometrics*, **4**, 2740-2843. Engle, R.F. and McFadden, D.L. editors. Elsevier, AMsterdam.
- Xia, Y. (1998), Doctoral thesis University of Hong Kong.
- Yakowitz, S. (1993), Nearest neighbour regression estimation for null-recurrent Markov times series. *Stoch. proc. Appl.*, **48**, 311-318.

Analyse en Composantes Principales (ACP) de fonctions de densité. Application aux données climatiques.

Théodora ANTONIADOU, P. BESSE, A.L. FOUGÈRES,

D. STEPHENSON et F. ABDELLANI

Laboratoire de Statistique et Probabilités

Université Paul Sabatier

118, Route de Narbonne

31062, Toulouse Cédex

02/05/00

L'analyse en composantes principales (ACP) de fonctions de densité est un sujet de recherche relativement récent sur lequel deux travaux sont connus, ceux de Boumaza R. (1999) et Kneip A. & Utikal K.J. (1999). Dans les deux travaux, les auteurs étudient la question du même point de vue; ils supposent, entre autres, que les fonctions de densité f_1, \dots, f_N forment une base de l'espace, hypothèse qui conduit à l'analyse spectrale d'une matrice dont les éléments sont les produits scalaires $\int f_i f_j, i, j = 1, \dots, N$. Le problème se réduit donc en un problème d'estimation du produit scalaire de deux densités. Kneip A. & Utikal K.J. (1999) proposent alors des estimateurs non paramétriques à noyau tandis que Boumaza R. (1999) considère le cas de densités gaussiennes en présentant des estimateurs des deux types, paramétriques et non paramétriques.

Pour notre part, nous abordons le problème d'un autre point de vue qui consiste à ramener l'ACP fonctionnelle à sa plus simple version multidimensionnelle. Ceci implique la discrétisation des fonctions de densité. La méthode proposée comporte trois étapes:

1. Estimation non paramétrique des fonctions de densité f_1, \dots, f_N .
2. Discrétisation des fonctions selon la même grille.
3. Analyse spectrale de la matrice dont les éléments sont les valeurs des estimateurs aux points de discrétisation.

Cette méthode a été mise en oeuvre dans le cadre du projet européen STOEC (Storm-Track upper Ocean interaction and the impact on European Climate) pour étudier l'évolution de la distribution annuelle de température en Angleterre Centrale. Les données sont les valeurs journalières de température des 175 dernières années. Les fonctions de densité des 175 années sont estimées par noyau gaussien. D'autres méthodes d'estimation de densité sont présentées dans le livre de Silverman B.W. (1986).

Il est intéressant de noter que la suite des fonctions propres obtenues par l'analyse des courbes de densité de température ressemble beaucoup à la suite des fonctions d'Hermite (produit d'une densité gaussienne avec les polynômes d'Hermite).

Afin d'expliquer ce résultat, nous avons présenté une troisième approche qui est basée sur le développement d'une fonction de densité suivant des polynômes orthogonaux. Cette approche est citée dans le livre de Ramsay J. & Silverman B.W. (1997) pour le cas général de l'ACP de fonctions. L'idée est de pouvoir exprimer les fonctions de densité f_1, \dots, f_N comme combinaison linéaire des polynômes orthogonaux (polynômes d'Hermite, Laguerre, Jacobi, etc). Lorsque ceci est possible le problème peut se réduire à l'analyse spectrale d'une matrice $C^t C$, C étant la matrice contenant les coefficients du développement des densités.

Références

- Boumaza, R. (1999) Analyses factorielles des distributions marginales de processus, These de doctorat, *Universite de Grenoble*.
- Kneip, A. & Utikal, K.J. (1999) *Inference for density families using functional principal component analysis*, Preprint.
- Ramsay, J & Silverman B.W. (1997) *Functional data analysis*, Springer-Verlag.
- Silverman B.W. (1986) *Density estimation for statistics and data analysis*, Chapman and Hall.

Modèle non linéaire fonctionnel : une approche par régression inverse

Anne-Françoise YAO[†] et Louis FERRÉ[‡]

[†]Laboratoire de Statistique et Probabilités

Université Paul Sabatier

118, Route de Narbonne

31062, Toulouse Cédex

yao@cict.fr

[‡] LSP et GRIMM

Université Toulouse Le Mirail

31058 Toulouse Cédex

loferre@univ-tlse2.fr

09/05/00

Le modèle le plus utilisé pour expliquer une variable aléatoire réelle par un processus du second ordre, (X_t) , est le modèle linéaire:

$$Y = \int_0^1 \theta(t) X(t) dt + \epsilon,$$

où θ est un paramètre fonctionnel et ϵ une variable aléatoire réelle indépendante de (X_t) . Les méthodes classiques d'estimation du paramètre θ sont répertoriées dans l'article de Franck and Friedman (1993).

Cardot *et al.* (1999) ont proposé une forme plus théorique du modèle précédent en se basant sur le fait que (X_t) est un processus du second ordre donc élément d'un espace de Hilbert H et grâce à la représentation de Riesz, il réécrivent le modèle précédent sous la forme :

$$Y = \Theta(X) + \epsilon,$$

où Θ est une forme linéaire continue sur H à estimer.

Bien entendu, il n'y a aucune raison pour que dans certaine situation, X dépende linéairement de Y . On cherche alors de façon naturelle à étendre les méthodes non-linéaire multidimensionnelles classiques. C'est ainsi que Ferraty et Vieu (2000) proposent une régression non-paramétrique :

$$Y = R(X) + \epsilon$$

où $R(\cdot) = E(Y|X = \cdot)$ est estimé par une méthode des noyaux.

Cet exposé est basé essentiellement sur l'article de Ferré et Yao (2000) dans lequel nous proposons une approche semi-paramétrique. Le modèle considéré est le suivant:

$$Y = f(\langle \theta_1, X \rangle, \dots, \langle \theta_K, X \rangle, \epsilon).$$

où $\theta_1, \dots, \theta_K$ sont des vecteurs linéairement indépendants de H et ϵ défini comme précédemment. Par le théorème de représentation de Riesz, pour $K = 1$ on a soit un modèle non paramétrique avec $R = f(\langle \theta_1, \cdot \rangle)$ soit le modèle de Cardot *et al.* (1999) lorsque f est la fonction identité. Ce modèle apparaît alors comme la version fonctionnelle du modèle introduit par Li (1991) qui a conduit à la méthode de régression inverse par tranches (SIR). Nous avons montré que si X suit une loi elliptique hilbertienne, les θ_i sont les vecteurs propres Γ -orthonormés associés aux valeurs non nulles de l'opérateur $\Gamma^{-1}\text{var}(E(X|Y))$ (où $\Gamma = \text{var}(X)$). Ainsi, les propriétés conduisant à la SIR restent vraies dans le cadre hilbertien mais Γ^{-1} étant un opérateur non-borné, l'estimation de ces paramètres n'est pas aussi simple qu'en dimension finie. Nous proposons donc de procéder comme Bosq (1991), par une suite de projection des données sur des espaces de dimension finie convergeant vers l'espace entier. À partir de l'estimateur obtenue par tranchage du support de Y de l'opérateur $\text{var}(E(X|Y))$, et sous des hypothèses similaires à Cardot *et al.* (1999), nous obtenons des estimateurs convergeant de l'espace des paramètres θ_i .

Références

- 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, pp. 509-529.
- Cardot, H. , Ferraty. F and Sarda, P. (1999) Functional linear model *Statist. Probab. Let.* **45** 11-22.
- Ferraty, F. and Vieu P. (2000). Functional nonparametric model for scalar response. Preprint.
- Ferré, L. and Yao, A.F. (2000) Sliced inverse functional analysis. Preprint.
- Li, K. C. (1991). Sliced Inverse Regression for dimension reduction *J. Amer. Statist. Assoc.*, **86**, 316-342.
- Ramsay, J. O. and Silverman, B. W. (1997). Functional Data Analysis Springer-Verlag.

Estimation bayésienne de l'intensité d'un processus de Cox non-homogène par une méthode MCMC à saut réversible

Michel GOULARD

INRA Toulouse

Unité Biométrie et Intelligence Artificielle

31326 Castanet-Tolosan

goulard@toulouse.inra.fr

29/05/00

Dans l'étude de peuplements forestiers, on étudie généralement les interactions entre arbres par des techniques à l'ordre 2 basées sur la théorie des processus ponctuels spatiaux. Dans ce cas on fait l'hypothèse que le processus sous-jacent est stationnaire. Dans un certain nombre de situations cette hypothèse est manifestement inadéquate. On peut utiliser le cadre des processus de Cox inhomogènes pour décrire ces répartitions. Dans ce cadre on suppose l'existence d'une intensité $\lambda(x)$, dépendant du point géographique x , qui décrit localement le nombre moyen de points par unité de surface. C'est cette quantité que l'on cherche alors à estimer. Pour aller plus avant dans la compréhension des phénomènes biologiques sous-jacents, on peut supposer que cette intensité dépend d'une variable explicative Z (par exemple la topographie, la déclivité, la nature du sol ...) sous la forme $\lambda_0(x) \gamma(Z(x))$ et on va estimer λ_0 et γ . Si le premier problème peut-être résolu en faisant appel aux techniques d'estimation non-paramétrique de densités, il n'en est pas de même du second. Pour celui-ci une méthode d'estimation bayésienne basée sur l'algorithme de Métropolis à sauts réversibles de Green (1995) a été proposée par Heikkinen (1997).

Durant l'exposé la méthode proposée par Heikkinen sera explicitée pour le cas simple de l'estimation de $\lambda(x)$ (Heikkinen & Arjas 1998). Cela nous permettra de revenir sur l'algorithme de Green qui peut intéresser la communauté de STAPH (voir un exemple d'utilisation de cet algorithme pour l'estimation d'une régression (Denison, Mallick & Smith 1998)). Cet algorithme est présenté sous une forme plus détaillée que le papier de Green dans (Waagepetersen & Sorensen 2000). L'extension à l'estimation de λ_0 et γ sera ensuite esquissée. L'utilisation de la méthode sur des données de positionnement d'arbres dans un peuplement en Guyanne sera ensuite présenté. Cet exemple montre la difficulté du choix des paramètres à spécifier dans la méthode et sa lourdeur en temps calcul. S'il est clair que la méthode ne semble pas compétitive pour le premier problème d'estimation, par exemple par rapport à un estimateur à noyau, c'est la seule approche existante

pour le deuxième.

Références

- R. Waagepetersen & D. Sorensen, 2000. A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping.
Preprint sur <http://www.statslab.cam.ac.uk/mcmc/>.
- D.G.Denison, B.K. Mallick & A.F.M. Smith, 1998. Automatic bayesian curve fitting. *JRSS B*, **60**, 333-350.
- P.J. Green, 1995. Reversible jump Markov chain Monte Carlo computation and bayesian model determination. *Biometrika*, **82**, 711-732.
- J. Heikkinen & E. Arjas, 1998. Non-parametric bayesian estimation of a spatial Poisson intensity. *Scand.J Stat.*, **25**, 435-450.
- J. Heikkinen, 1997. Bayesian smoothing and step functions in the nonparametric estimation of curves and surfaces. Jyväskylä studies in computer science, economics and statistics, vol 40.

Permutation tests in change point analysis

Jaromír ANTOCH et Marie HUŠKOVÁ

Charles University, Department of Statistics

Sokolovská 83, CZ-186 75

Praha 8 – Karlín, Czech Republic

jaromir.antoch@karlin.mff.cuni.cz

marie.huskova@karlin.mff.cuni.cz

09/06/00

For simplicity, consider the location model with a change after an unknown time point m , i.e.

$$X_i = \mu + \delta_n I\{i > m\} + e_i, \quad i = 1, \dots, n, \quad (1)$$

where $1 \leq m \leq n$, μ and $\delta_n \neq 0$ are unknown parameters and $I\{A\}$ denotes the indicator of a set A . Assume, moreover, that

$$e_1, \dots, e_n \text{ are independent identically distributed random variables (iid rv's) with } E e_i = 0, \quad 0 < \text{var } e_i < \infty \text{ and } E |e_i|^{2+\Delta} < \infty \text{ with some } \Delta > 0. \quad (2)$$

We are interested in the testing problem

$$H_0 : m = n \quad \text{against} \quad H_1 : m < n. \quad (3)$$

In this talk it will be shown that the permutation tests can also be used to get asymptotically correct approximations for critical values.

The permutation test suggested below motivated us to develop along the same lines a variety of permutation tests related to other test statistics used in change point analysis, too. This means that the same principle can be applied to other test statistics (including M -tests) for the testing problem (3) in the model (1) as well as to the case of multiple changes in location models. The crucial point is that the test statistic must be expressible through the partial sums of residuals and that under H_0 these residuals are exchangeable random variables.

We apply the permutation arguments to the test based on the statistic

$$T_{n1} = \max_{1 < k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{1}{\hat{\sigma}_n} \left| \sum_{i=1}^k (X_i - \bar{X}_n) \right| \right\}, \quad (4)$$

where

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (5)$$

This test statistic is closely related to the likelihood ratio test when the error terms e_i 's have normal distribution. The large values indicate that the null hypothesis is not true and therefore the null hypothesis H_0 is rejected for large values of T_{n1} .

The permutation distribution of T_{n1} can be described as the conditional distribution (given X_1, \dots, X_n) of

$$T_{n1}(\mathbf{R}) = \max_{1 < k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{1}{\widehat{\sigma}_n} \left| \sum_{i=1}^k (X_{R_i} - \bar{X}_n) \right| \right\}, \quad (6)$$

where $\mathbf{R} = (R_1, \dots, R_n)$ is a random permutation of $(1, \dots, n)$. This permutation distribution, denoted $F_P(x; T_{n1})$, can be expressed as

$$F_P(x; T_{n1}) = \frac{1}{n!} \# \left\{ \mathbf{r} \in \mathcal{R}_n; T_{n1}(\mathbf{r}) \leq x \right\}, \quad (7)$$

where \mathcal{R}_n is the set of all permutations of $\{1, \dots, n\}$ and $\#\{A\}$ denotes the cardinality of a set A . Denote by $x_{1-\alpha, n}$ the $100(1-\alpha)\%$ quantile of the permutation distribution $F_P(\cdot; T_{n1})$. Then the critical region with the level α of the permutation test based on T_{n1} has the form

$$T_{n1} \geq x_{1-\alpha, n}. \quad (8)$$

The permutation test can be described as follows:

1. we calculate T_{n1} according to (4) and the quantile $x_{1-\alpha, n}$;
2. the null hypothesis is rejected if (8) holds true.

Now then study the permutation distribution function of T_{n1} , more precisely, we derive the conditional limit distribution of $T_{n1}(\mathbf{R})$ given X_1, \dots, X_n . It is important to realize that $T_{n1}(\mathbf{R})$ given X_1, \dots, X_n can be viewed as a functional of a simple linear rank statistic and theorems on rank statistics for change point can be applied. Our main assertion states:

Theorem. *Let the observations X_1, \dots, X_n follow the model (1), the assumptions (2) be satisfied and let $|\delta_n| \leq D_0$ with some $D_0 > 0$. If $n \rightarrow \infty$, then for all $-\infty < y < \infty$ we have*

$$\begin{aligned} P\left(\sqrt{2 \log \log n} T_{n1}(\mathbf{R}) \leq y + 2 \log \log n + \frac{1}{2} \log \log \log n - \frac{1}{2} \log \pi \mid X_1, \dots, X_n\right) \\ \rightarrow \exp\{-2 \exp\{-y\}\}, [P] - \text{a.s.} \end{aligned}$$

Remark. Paper is now in print in *Statistics and Probability Letters* journal.

Inférence statistique pour la localisation d'une discontinuité par régression linéaire locale

Gérard GRÉGOIRE

LMC/IMAG

Université Joseph Fourier

BP 53

38041 Grenoble Cédex 9

Gerard.Gregoire@imag.fr

09/06/00

L'exposé présenté concerne l'inférence statistique pour les discontinuités dans un modèle de régression non paramétrique. Le thème est centré autour d'un travail qui a fait l'objet de la thèse soutenue à Grenoble par Zouhir Hamrouni (Janvier 1999). C'est un sujet qui a été l'occasion de travaux de plus en plus nombreux ces dernières années. Les applications concernent des secteurs aussi divers que l'économie, la biostatistique, la fiabilité, le traitement du signal ... On observe les valeurs bruitées d'une fonction de régression $m(\cdot)$ présentant un certain nombre de discontinuités. Les problèmes que l'on est habituellement amenés à traiter concernent les tests d'existence de discontinuités (détection) l'estimation du nombre de discontinuités, de leurs localisations et des amplitudes.

Le modèle de régression auquel nous intéressons ici est un modèle à plan d'expérience aléatoire. Nous observons des couples (X_i, Y_i) , $i = 1, \dots, n$ tels que

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i, \quad i = 1, \dots, n,$$

où les ϵ_i sont i.i.d., indépendants des X_i de moyenne nulle et de variance unité. Les X_i sont à valeurs dans $[0, 1]$. Nous nous limitons dans le travail présenté à la situation simple où nous savons qu'une discontinuité unique est présente. La discontinuité est un saut de la fonction $m(\cdot)$; en d'autres termes on suppose

$$m(x) = m_0(x) + \gamma I_{[\tau, 1]}(x),$$

où typiquement m_0 est de classe \mathcal{C}_2 . Bien que l'extension ne soit pas tout à fait immédiate, nous pensons possible d'étendre les résultats présentés à des situations plus générales où le nombre de discontinuités serait supérieur à un, ou telles que les discontinuités concerneraient les dérivées...

Nous avons recours au principe maintenant bien établi de l'estimation de l'amplitude d'une discontinuité éventuelle au point x par la différence des lissages à droites et à gauche

$$\hat{\gamma}(x) = \hat{m}_+(x) - \hat{m}_-(x)$$

où $\hat{m}_+(x)$ (respt. $\hat{m}_-(x)$) n'utilise que les observations à droite (respt. à gauche) de x . $\hat{m}_+(x)$ et $\hat{m}_-(x)$ sont des estimateurs des limites à droites et à gauche $m_+(x)$ et $m_-(x)$ obtenus par régression linéaire locale en ayant recours à des noyaux unilatéraux K_+ et K_- . Suivant en cela Loader, et Wu et Chu, nous supposons que $K_+(0) > 0$ et $K_-(0) > 0$. Cette hypothèse se révèle déterminante pour la vitesse de convergence de l'estimateur de la localisation.

Le résultat central présenté dans l'exposé concerne le comportement asymptotique du processus $x \rightarrow \hat{\gamma}(x)$. Nous montrons que le processus $\mathcal{Z}(z)$ de déviation locale construit en effectuant une opération de dilatation-translation autour du point τ , contrôlée par la taille de l'échantillon et la largeur de fenêtre h_n :

$$\mathcal{Z}_n(z) = \alpha(n, h_n) \left(\hat{\gamma}\left(\tau + z \frac{h_n}{\beta(n, h_n)}\right) - \hat{\gamma}(\tau) \right)$$

converge en loi vers un processus de Poisson composé avec dérive. Une conséquence immédiate est que $\hat{z} = \arg \sup \mathcal{Z}_n(z)$ converge en loi. Par suite il en est de même de $\hat{\tau} = \tau + \hat{z} \frac{h_n}{\beta(n, h_n)}$. Il s'ensuit en particulier que l'on obtient la convergence en loi de $\hat{\tau}$ vers τ à la vitesse exacte de n^{-1} alors que les résultats standards conduisent à des vitesses en $n^{-1+\varepsilon}$. La preuve du résultat fondamental repose sur l'approximation de $\mathcal{Z}(z)$ par la somme des variables d'une ligne d'un tableau triangulaire pour lequel on peut appliquer des résultats de convergence relatifs dans un contexte de lois indéfiniment divisibles.

Pour terminer nous présentons aussi des expériences numériques mettant en évidence le comportement asymptotique du processus de déviation locale et les performances de l'estimateur de la localisation ainsi que de celui qui en découle pour l'amplitude de la discontinuité. Nous traitons aussi les données bien connues du débit du Nil et montrons que nos résultats sont tout à fait cohérents avec ceux obtenus par d'autres méthodes.

Références

- Antoniadis, A. and Grégoire, G. (1993). Nonparametric estimation in change-point hazard rate models for censored data: a counting process approach. *Journal of Nonparametric Statistics*, **3**, 135-154.
- V. Couallier, P. Sarda et P. Vieu. (1997). Estimation non paramétrique de discontinuités d'une fonction d'intensité, *Revue de statistique appliquée*, XLV, 89-106.
- V. Couallier (2000). Inférence statistique pour des estimateurs de discontinuités dans un cadre non paramétrique. Thèse de l'Université Paul Sabatier. Juin 2000.

- J. Fan and I. Gijbels. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Grégoire, G. and Hamrouni, Z. (1999). Local Linear Smoothing For Change Point Estimation. Technical report, Laboratoire LMC/IMAG, Université Joseph Fourier.
- Grégoire, G. and Hamrouni, Z. (2000). Two non-parametric tests for change-point problems. *Journal of Nonparametric Statistics*, to appear.
- Hall, P., Gijbels, I. and Kneip, A. (1996). On the estimation of jump points in smooth curves. Technical report, Institut de Statistique, Université Catholique de Louvain.
- Hamrouni, Z. (1999). Inférence statistique par lissage linéaire local pour une fonction de régression présentant des discontinuités. Unpublished thesis, Laboratoire LMC/IMAG, Université Joseph Fourier, 1999.
- Loader, C.R. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics*, **24**, 1667-1678 .

Non causalité et discrétisation fonctionnelle, théorèmes limites pour un processus ARHX(1)

Serge GUILLAS

Université Paris VI et Ecole des Mines de Douai

guillas@ensm-douai.fr

19/06/00

A partir de deux processus à trajectoires continues $(x_t, t \in \mathbb{R}^+)$, $(z_t, t \in \mathbb{R}^+)$ on peut définir les processus à temps discret mais à valeurs dans un espace fonctionnel $X_k(t) = x_{k\delta+t}$, $0 \leq t \leq \delta$, $k = 0, 1, 2, \dots$ et $Z_k(t) = z_{k\delta+t}$, $0 \leq t \leq \delta$, $k = 0, 1, 2, \dots$. Cela permet de manipuler des portions de trajectoires. Des modèles autorégressifs banachiques ou hilbertiens ont été étudiés dans Bosq (2000) et une application se trouve dans Besse *et al.* (2000).

Nous généralisons à des processus à valeurs dans un espace de Banach et utilisons ici les notions de non causalité en temps discret ou continu issues de Florens et Fougère (1996) (historiquement introduites par Granger (1969) et Sims (1972)). L'information contenue dans (X_0, \dots, X_{n-1}) est la même que celle contenue dans $(x_u, 0 \leq u \leq n\delta)$, ce qui permet d'établir que la non-causalité forte ou faible globale est stable par discrétisation fonctionnelle, ce qui est faux pour la non-causalité instantannée.

Nous pouvons aussi remarquer que la connaissance des échantillons $\{X_i(\frac{j}{k}), 1 \leq j \leq k\}$, $\{Z_i(\frac{j}{k}), 1 \leq j \leq k\}$, $i \in \mathbb{N}$, $k \in \mathbb{N}^*$ permet d'obtenir des propriétés asymptotiques de non causalité pour les processus à trajectoires continus sous-jacents.

On peut considérer sur un espace probabilisé (Ω, \mathcal{A}, P) le modèle suivant, noté ARHX(1):

$$X_n = \rho(X_{n-1}) + a(Z_n) + \varepsilon_n \quad (11)$$

avec ρ et a opérateurs linéaires et bornés de l'espace de Hilbert réel et séparable H vers H , (ε_n) un bruit blanc fort hilbertien et (Z_n) la suite de variables aléatoires exogènes. Les hypothèses de base sont les suivantes: (Z_n) faiblement stationnaire de moyenne nulle, $(\varepsilon_n) \perp (Z_n)$ et $\sum_{n=0}^{\infty} \|\rho^n\| < \infty$.

On considèrera deux cas particuliers. Le premier suppose la propriété très exigeante d'indépendance du présent de (X_n) et du futur de (Z_n) . Le deuxième cas est celui où (Z_n) est un ARH(1) de moyenne nulle. On démontre pour ces modèles et avec parfois des hypothèses complémentaires l'existence et l'unicité d'une solution stationnaire, des propriétés de Markov, la loi des grands nombres en précisant la vitesse et le théorème central-limite. Nous obtenons également

des résultats voisins de ceux de Bosq (2000) concernant l'opérateur de covariance empirique.

Références

- P. Besse, H. Cardot and D. Stephenson, 2000. Autoregressive forecasting of some functional climatic variations, *Scandinav. J. Statist.*, à paraître.
- Bosq, 2000. *Linear processes in function spaces*, Springer Verlag.
- J.P. Florens, D. Fougère, 1996. Noncausality in continuous time, *Econometrica*, **64**, No 5.
- C.W.J. Granger, 1969. Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37**, 424-459.
- S. Guillas, 2000. Non-causalité et discrétisation sonctionnelle, théorèmes limites pour un processus ARHX(1), *C. R. Acad. Sci. Paris*, t. 331, Série I, p. 91-94
- C.A. Sims, 1972. Money, income and causality, *American Economic Review*, **62**, 540-552.

Data Exploration Using Piecewise Polynomial Regression Trees

Probal CHAUDHURI
Indian Statistical Institute, Calcutta
probal@isical.ac.in
26/06/00

Tree structured regression based on adaptive recursive partitioning of the data can be blended with the idea of maximum likelihood estimation in generalized regression problems and quantile regression techniques to lead to some powerful non-parametric function estimation procedures. The function estimate can be obtained in a piecewise polynomial form with each piece determined by the data in a terminal node of a binary decision tree. The tree structure takes care of much of the model complexity and interactions among covariates. We shall demonstrate using several examples how piecewise polynomial generalized regression trees and quantile regression trees can help in exploring important statistical features in large and high dimensional noisy data. This is joint work with Professor Wei-Yin Loh of University of Wisconsin at Madison, U.S.A.

References

- Chaudhuri, P., 2000, Asymptotic Consistency of Median Regression Trees, To appear in *Journal of statistical Planning and Inference*.
- Chaudhuri, P., Huang, M.C., Loh, W.Y. and Yao, R., 1994, Piecewise-Polynomial Regression Trees, *Statistica Sinica*, **4**, 143-167.
- Chaudhuri, P., Lo, W.D., Loh, W.Y. and Yang, C.C., 1995, Generalized Regression Trees, *Statistica Sinica*, **5**, 641-666.
- Chaudhuri, P. and Loh, W.Y. Nonparametric estimation of conditional quantiles using quantile regression trees, 2000, *Preprint*.