



Régression linéaire locale pour variable fonctionnelle

Alain Berlinet, Abdallah Elamine, André Mas

► **To cite this version:**

Alain Berlinet, Abdallah Elamine, André Mas. Régression linéaire locale pour variable fonctionnelle. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386655

HAL Id: inria-00386655

<https://hal.inria.fr/inria-00386655>

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

REGRESSION LINEAIRE LOCALE POUR VARIABLES FONCTIONNELLES

Alain Berlinet, Abdallah Elamine, André Mas

*Département de Mathématiques
Université Montpellier II
Place Eugène Bataillon
34095 Montpellier Cedex 5
France*

Résumé. Dans ce travail, on s'intéresse à la régression non paramétrique locale pour des variables explicatives fonctionnelles. On propose tout d'abord un estimateur de l'opérateur de régression. La construction de cet estimateur est liée à la résolution d'un problème inverse linéaire. On établit des bornes de l'erreur quadratique moyenne en utilisant une méthode de décomposition classique. Cette EQM dépend de la fonction de petite boule de probabilité du régresseur au sujet de laquelle des hypothèses de type Gamma-variation sont posées.

Abstract. *In this paper, we address the problem of local non parametric regression with functional inputs. First, we propose an estimator of the unknown regression function. The construction of this estimator is related to the resolution of a linear inverse problem. Using a classical method of decomposition, we establish a bound for the mean square error. This bound depends on the small ball probability of the regressor which is assumed to belong to the class of gamma varying functions.*

Mots clés : Données fonctionnelles; Modèle de régression; Noyau; Erreur quadratique moyenne; Petite boule de probabilité; Problème inverse.

1 Introduction

Soit H un espace de Hilbert réel et séparable. On considère une suite de couples de variables aléatoires $(y_i, X_i)_{1 \leq i \leq n}$ de $\mathbb{R} \times H$, définies sur le même espace de probabilité. On note respectivement le produit scalaire et la norme de H par $\langle \cdot, \cdot \rangle$ et $\| \cdot \|$. On désigne respectivement par X et y les notations génériques des variables aléatoires X_i et y_i . On s'intéresse à la fonction de régression :

$$m(x_0) = \mathbb{E}(y|X = x_0)$$

où m est une application de H dans \mathbb{R} et x_0 un élément de l'espace de Hilbert H . Fan (1993) étudie l'optimalité des vitesses dans le cas des variables aléatoires X réelles.

Dans le cas de la dimension infinie ($\dim(H) = +\infty$), un estimateur à noyau est proposé par Ferraty, Vieu (2006). Récemment, Ferraty, Mas, Vieu (2006) ont montré la normalité asymptotique de cet estimateur.

Dans cet article, on souhaite proposer un estimateur de m par une autre méthode et, ensuite, établir des bornes de son erreur quadratique moyenne.

On propose un estimateur de m en considérant le problème de minimisation suivant :

$$\min_{a \in \mathbb{R}, \varphi \in H} \sum_{i=1}^n (Y_i - a - \langle \varphi, X_i - x_0 \rangle)^2 K \left(\frac{\|X_i - x_0\|}{h} \right) \quad (1)$$

où K est un noyau positif et h une fenêtre. Ce programme généralise celui de la dimension finie en considérant une approximation affine de la fonction de régression au voisinage de x_0 . La solution a^* du programme précédent est l'estimateur recherché et φ peut être vu comme un gradient. Des programmes alternatifs, dans ce cadre statistique sont considérés dans Barrientos-Marin, Ferraty, Vieu (2007) mais ne reprennent pas stricto sensu l'approche fini-dimensionnelle.

2 Construction de l'estimateur

On commence par introduire deux opérateurs linéaires sur H qui interviennent dans la définition de notre estimateur.

Definition 1 *L'opérateur de covariance local théorique de X en $x_0 \in H$ associé au noyau K et sa version empirique sont définis par*

$$\Gamma_K = \mathbb{E} \left[K \left(\frac{\|X - x_0\|}{h} \right) (X - x_0) \otimes (X - x_0) \right], \quad (2)$$

$$\Gamma_{n,K} = \frac{1}{n} \sum_{k=1}^n K \left(\frac{\|X_k - x_0\|}{h} \right) (X_k - x_0) \otimes (X_k - x_0). \quad (3)$$

Remarque 2 *L'opérateur Γ_K dépend de n à travers la suite ($h = h(n)$). L'inverse de l'opérateur $\Gamma_{n,K}$ n'existe pas puisque l'opérateur $\Gamma_{n,K}$ est presque sûrement de rang fini.*

Pour résoudre le programme de minimisation (1) on fait apparaître un problème linéaire mal posé qui fait intervenir les opérateurs de la définition précédente. On va en fait déterminer un inverse régularisé de l'opérateur $\Gamma_{n,K} : \Gamma_{n,K}^\dagger$ selon deux méthodes à savoir la pénalisation et la régularisation de Tikhonov. En désignant par α_n une suite de nombres réels positifs qui décroît vers 0 et par I l'opérateur identité, la première méthode permet d'écrire $\Gamma_{n,K}^\dagger = (\Gamma_{n,K} + \alpha_n I)^{-1}$ et la deuxième $\Gamma_{n,K}^\dagger = (\Gamma_{n,K}^2 + \alpha_n I)^{-1} \Gamma_{n,K}$. On peut maintenant donner une expression explicite de notre estimateur.

Definition 3 L'estimateur $\hat{m}_n(x_0)$ est défini par :

$$\hat{m}_n(x_0) = \frac{\sum_{i=1}^n y_i \omega_{i,n}}{\sum_{i=1}^n \omega_{i,n}} \quad (4)$$

où

$$\omega_{i,n} = K \left(\frac{\|X_i - x_0\|}{h} \right) \left(1 - \langle X_i - x_0, \Gamma_{n,K}^\dagger \bar{Z}_{n,K} \rangle \right)$$

avec

$$\bar{Z}_{n,K} = \frac{1}{n} \sum_{i=1}^n K \left(\frac{\|X_i - x_0\|}{h} \right) (X_i - x_0).$$

Remarque 4 L'estimateur $\hat{m}_n(x_0)$ dépend de la méthode utilisée pour déterminer $\Gamma_{n,K}^\dagger$.

Dans la suite, on suppose que les variables aléatoires X_i , $i = 1, \dots, n$, sont centrées.

3 Hypothèses

En dimension infinie, la densité de probabilité ne peut pas être définie car dans ce contexte il n'existe pas un équivalent de la mesure de Lebesgue. Pour contourner ce problème, on utilise une fonction appelée fonction de petite boule de probabilité définie par $F(h) = P(\|X - x_0\| \leq h)$. Pour plus d'informations, nous renvoyons à Li, Linde (1993) et Li, Linde (1999). Pour caractériser cette fonction, on donne la définition suivante.

Definition 5 Une fonction F qui s'annule en 0 est appelée fonction à variation Gamma en 0 si elle est mesurable croissante sur un voisinage de 0 et s'il existe une fonction continue et positive ρ qui s'annule aussi en 0 telle que pour tout $x \in \mathbb{R}$, on ait :

$$\lim_{h \rightarrow 0} \frac{F(h + \rho(h)x)}{F(h)} = \exp(x).$$

La fonction ρ est appelée fonction auxiliaire associée à la fonction F .

Dans le cas où X est le processus de Wiener, des résultats montrent que $\mathbb{P}(\|X - x_0\| \leq h) \sim C_1 h^\alpha \exp(-\frac{C_2}{h^\beta})$ au voisinage de 0, où C_1, C_2, α et β sont des constantes positives, pour différents choix du "shift" x_0 et pour la norme L^2 (voir Li, Shao (2001)). Cette fonction de petite boule de probabilité est une fonction à Gamma variation en 0 et sa fonction auxiliaire associée est définie par $\rho(s) = C_3 s^{1+\beta}$ avec C_3 une constante positive. La définition qui suit est utile pour caractériser la fonction auxiliaire.

Definition 6 Une fonction ρ est une fonction à variation régulière en 0 avec indice g si pour tout $x \in \mathbb{R}$, on a :

$$\lim_{h \rightarrow 0} \frac{\rho(hx)}{\rho(h)} = x^g.$$

Pour plus de détails sur les fonctions à variation régulière, le lecteur intéressé pourra consulter Bingham, Goldie, Teugels (1987), de Haan (1971) et de Haan (1974).

Voici maintenant nos hypothèses de travail.

H₀) Le noyau K est à support compact inclus dans $[0, 1]$ avec $K(1) > 0$. Sa dérivée K' existe, est non nulle, et appartient à l'espace $L^1([0, 1])$.

H₁) Soit Γ l'opérateur de covariance de X défini par $\Gamma = \mathbb{E}(X \otimes X)$. On désigne par $(e_i)_{i \in \mathbb{N}^*}$ la base de vecteurs propres de Γ et f_i la densité de probabilité de la variable aléatoire $\langle X, e_i \rangle$. On suppose que les variables aléatoires $(\langle X, e_i \rangle)_{1 \leq i \leq n}$ sont des variables aléatoires indépendantes. De plus, on suppose que les dérivées $f_i^{(p)}$, $p \in \{1, 2, 3\}$, de f_i existent, vérifient $f_i^{(p)}(\langle x_0, e_i \rangle) \neq 0$, et $f_i^{(3)}$ est continue.

H₂) La densité de la variable aléatoire

$$\sqrt{\sum_{i=1}^{\infty} \langle X - x_0, e_k \rangle^2} = \|X - x_0\|$$

existe dans un voisinage de 0, est notée $f_{\|X-x_0\|}$ et sa fonction de petite boule de probabilité F est une fonction à variation Gamma en 0 de fonction auxiliaire ρ .

H₃) Il existe un voisinage \mathcal{V}_0 de 0 tel que

$$\alpha_i = \sup_{u \in \mathcal{V}_0} \left| \frac{f_i(u) - f_i(-u)}{u(f_i(u) + f_i(-u))} \right| < +\infty$$

et

$$\sum_{i=1}^{+\infty} \alpha_i^2 < +\infty$$

H₄) La dérivée première $m'(x_0)$ de m au point x_0 est définie, non nulle, et il existe un voisinage $\mathcal{V}(x_0)$ de x_0 tel que

$$\sup_{x \in \mathcal{V}(x_0)} \|m''(x)\|_{\infty} < +\infty.$$

H₅) Soient $v(h) = \mathbb{E} \left[K \left(\frac{\|X-x_0\|}{h} \right) \|X - x_0\| \rho(\|X - x_0\|) \right]$ et respectivement Γ_K^\dagger et $\Gamma_{n,K}^\dagger$ les inverses régularisés de Γ_K et $\Gamma_{n,K}$. On suppose qu'il existe une suite (r_n) qui tend vers 0 telle que :

$$\max\{\|\Gamma_K^\dagger\|_{\infty}, \|\Gamma_{n,K}^\dagger\|_{\infty}\} \leq \frac{1}{r_n v(h)} \quad \text{p.s}$$

L'hypothèse **H₁** est toujours vraie si X est gaussienne et **H₂** est vérifiée pour une large classe de processus gaussien. L'hypothèse **H₃** est aussi satisfaite lorsque X est gaussienne. En effet, dans ce cas, la densité f_i est donnée par $f_i(u) = \frac{1}{\sqrt{2\pi\lambda_i}} \exp \left[-\frac{(u - \langle x_0, e_i \rangle)^2}{2\lambda_i} \right]$, où

$(\lambda_i)_{i \in \mathbb{N}^*}$ sont les valeurs propres de l'opérateur de covariance Γ , et l'hypothèse \mathbf{H}_3 est vérifiée, par exemple, lorsque la série $\sum_{i=1}^{\infty} \frac{\langle x_0, e_i \rangle^2}{\lambda_i}$ converge. La suite r_n qu'on voit dans l'hypothèse \mathbf{H}_5 est une suite qui peut tendre vers l'infini ou vers un nombre strictement positif et qui permet de contrôler l'absence de régularité des inverses Γ_K^\dagger et $\Gamma_{n,K}^\dagger$.

4 Résultats asymptotiques

Cette partie vise essentiellement à l'établissement des bornes de l'erreur quadratique de l'estimateur $\hat{m}_n(x_0)$. Pour alléger les résultats, on introduit les notations suivantes.

Soient a_n et b_n deux suites positives. On note $a_n \asymp b_n$ s'il existe deux constantes positives α et β telles que $\alpha b_n \leq a_n \leq \beta b_n$.

C désigne une constante générique qui peut varier d'une ligne à une autre.

Théorème 7 *Soit $x_0 \in H$ fixé. Lorsque les hypothèses $\mathbf{H}_0 - \mathbf{H}_5$ sont satisfaites et si $nF(h) \rightarrow +\infty$, alors*

$$\begin{aligned} \mathbb{E} (\hat{m}_n(x_0) - m(x_0))^2 &\leq C \left(\frac{h^6}{r_n^2} + h_n^4 + \frac{h^2}{nF(h)} + \frac{v^2(h)}{F^2(h)} \right) \\ &\quad + \frac{C}{nF(h)} \left(1 + \frac{h^2}{nr_n v(h)} + \frac{v(h)}{r_n F(h)} \right). \end{aligned}$$

Remarque 8 *La démonstration du résultat précédent est basée sur la décomposition classique biais plus variance. La première expression découle du terme de biais et la seconde du terme de variance.*

Des hypothèses supplémentaires vont nous permettre de simplifier considérablement l'expression ci-dessus.

Proposition 9 *Si la fonction auxiliaire ρ est à variation régulière d'indice $g \geq 1$ en 0 on a :*

$$v(h) \asymp h\rho(h)F(h).$$

Lorsque la condition $r_n \asymp h^{\frac{1}{4}}$ et si $\rho(s) \geq Cs^4$, alors

$$\mathbb{E} (\hat{m}_n(x_0) - m(x_0))^2 \leq C \left(h^4 + \frac{1}{nF(h)} \right).$$

La vitesse de décroissance de l'erreur quadratique moyenne dépend de la fenêtre optimale h^ donné par :*

$$(h^*)^4 F(h^*) = \frac{1}{n}.$$

Remarque 10 *Notre objectif n'est pas d'obtenir des vitesses optimales. Cependant, dans le cas réel ($F(h) \sim Ch$), les vitesses que l'on obtient sont égales aux vitesses optimales obtenues par Fan (1993) à constante près.*

Bibliographie

- [1] Barrientos-Marin J., Ferraty F., Vieu P. (2007) locally modelled regression and functional data, soumis
- [2] Berline, A. et Elamine, A. et Mas A. (2007) Local linear regression for functional data, soumis.
- [3] Bingham, N. H., Goldie, C. M. et Teugel, J. L. (1987) *Regular Variations*, Encyclopedia of Mathematics and its Applications, Cambridge University Press.
- [4] Fan, J. (1993) local linear regression smoothers and their minimax efficiencies. *Annals of Statistics*, 21, 196–216.
- [5] de Haan, L. (1971) A form of regular variation and its application to the domain of attraction of the double exponential distribution. *Wahrscheinlichkeitstheorie.verw. Geb*, 17, 241–258.
- [6] de Haan, L. (1974) Equivalence classes of regularly varying functions. *Stochastic Processes and their Applications*, 2, 243–259.
- [7] Li W.V., Linde W. (1993) Small ball problems for non-centered gaussian measures, *Prob. Math. Stat.*, 14, 231-251.
- [8] Li W.V., Linde W. (1999) Approximation, metric entropy and small ball estimates for Gaussian measures, *Annals of Probability*, 27, 1556-1578.
- [9] Li W.V., Shao Q.-M. (2001) Gaussian processes : Inequalities, small ball probabilities and applications, *Handbook of Statistics*, 19, 533-597.
- [10] Mas, A. (2007) Local Functional Principal Component Analysis. *Complex Analysis and Operator Theory*, 2, 135-167.