Tests minimax adaptatifs pour les modèles fonctionnels linéaires

Nadine Hilgert, André Mas, Nicolas N. Verzelen

Tests minimax adaptatifs pour les modèles fonctionnels linéaires

Nadine Hilgert [1] & André Mas [2] & Nicolas Verzelen [1]

[1] *INRA SUPAGRO, UMR 729 MISTEA, F-34060 Montpellier*
nadine.hilgert@supagrao.inra.fr & nicolas.verzelen@supagrao.inra.fr
[2] *I3M, Université Montpellier 2, F-34000 Montpellier*
andre.mas@univ-montp2.fr

Résumé. Nous présentons deux nouvelles procédures pour tester la nullité du paramètre de pente dans les modèles linéaires fonctionnels à sortie réelle. Les statistiques de test sont obtenues en combinant une approche par tests multiples et des projections aléatoires des covariables explicatives sous forme d'analyse en composante principale fonctionnelle. Les procédures sont totalement basées sur les données et ne requièrent pas de connaissance a priori sur la régularité du paramètre de pente ni sur celle des covariables fonctionnelles. Les niveaux et puissances par rapport à des alternatives locales sont étudiées dans un cadre non asymptotique. Nous montrons ainsi que les procédures sont minimax adaptatives à la régularité inconnue de la pente, à un terme multiplicatif $\log \log n$ près, inévitable. Comme résultat supplémentaire, nous déduisons les distances de séparation minimax de la pente pour une large gamme de classes de régularité. Les résultats présentés dans cette communication sont issus de *Hilgert, Mas and Verzelen, Minimax adaptive tests for the functional linear model, arXiv :1206.1194v1.*

Mots-clés. Régression linéaire fonctionnelle, analyse en composantes principales, tests adaptatifs, tests multiples

Abstract. We introduce two novel procedures to test the nullity of the slope function in the functional linear model with real output. The test statistics combine multiple testing ideas and random projections of the input data through functional Principal Component Analysis. Interestingly, the procedures are completely data-driven and do not require any prior knowledge on the smoothness of the slope nor on the smoothness of the covariate functions. The levels and powers against local alternatives are assessed in a nonasymptotic setting. This allows us to prove that these procedures are minimax adaptive (up to an unavoidable $\log \log n$ multiplicative term) to the unknown regularity of the slope. As a side result, the minimax separation distances of the slope are derived for a large range of regularity classes. The results presented here are developed in *Hilgert, Mas and Verzelen, Minimax adaptive tests for the functional linear model, arXiv :1206.1194v1.*

Keywords. Functional linear regression, principal component analysis, adaptive testing, multiple testing

1 Introduction

Nous considérons un modèle de régression linéaire générique dans un espace de dimension infinie :

$$Y = \langle X, \theta \rangle + \epsilon \ , \tag{1}$$

où Y est la variable à expliquer, réelle, X est une fonction aléatoire qui appartient à un espace de Hilbert séparable, noté $\mathcal{H}$, muni du produit scalaire $\langle ., . \rangle$, et ε est une variable aléatoire centrée de variance inconnue σ^2. Pour simplifier la présentation, nous supposons que X et Y sont centrés également. Enfin, θ est une fonction inconnue, représentant le paramètre de pente, qui appartient à $\mathcal{H}$. Dans la suite, nous notons $\mathbf{X}$ et $\mathbf{Y}$ les vecteurs composés de n observations i.i.d. X_i et Y_i $(1 \leq i \leq n)$, de même $\boldsymbol{\epsilon}$ représente un vecteur de taille n d'échantillon du bruit.

L'objectif de cette présentation est d'introduire des procédures automatiques de test de la nullité du paramètre de pente θ, dont les puissances sont optimales d'un point de vue non asymptotique. Cela revient à tester l'hypothèse :

$$H_0 : \text{ “}\theta = 0\text{”} \qquad \text{contre} \qquad H_1 : \text{ “}\theta \neq 0\text{”}$$

étant donné un échantillon i.i.d. $(\mathbf{X}, \mathbf{Y})$ issu du modèle (1). Les résultats présentés ici sont issus de [9].

2 Etat de l'art

La plupart des procédures de test de nullité de θ sont développés à partir de travaux sur l'estimation de θ. Une première classe de procédures est basée sur la minimisation d'un critère de type moindres carrés, pénalisé par un terme qui fixe la plausibilité de θ. Cela inclut par exemple les estimateurs à base de splines de lissage ou les méthodes basées sur les espaces de Hilbert à noyau reproduisant. Une seconde classe de procédures est basée sur l'analyse en composantes principales (ACP) de $\mathbf{X}$. Cela consiste à estimer θ dans une base de dimension finie engendrée par les k premières fonctions propres de l'opérateur de covariance empirique de $\mathbf{X}$. La différence principale avec la première approche est que la base de dimension finie est estimée à partir des observations du processus X. L'ouvrage [4] donne un aperçu de ces deux approches.

Les propriétés théoriques de ces deux classes d'estimateurs ont été étudiées dans le cadre de la prédiction ou de l'estimation de θ. Des vitesses de convergence optimales ont été établies et certaines des procédures d'estimation nommées ci-dessus permettent d'atteindre ces vitesses. Plus récemment, quelques résultats non asymptotiques ont été démontrés, pour des procédures d'estimation basées sur des bases fixes (base de splines par exemple, voir [5]). Une grande partie de ces procédures d'estimation reposent sur des paramètres de réglage dont les valeurs optimales dépendent de quantités telles que la variance du bruit ou la régularité de θ. Ces quantités ne sont pas connues en pratique, ce qui renforce le fossé entre théorie et pratique.

La littérature sur les tests dans les modèles linéaires fonctionnels est peu abondante. Dans [2], Cardot et al. introduisent une statistique de test basée sur les k premières composantes de l'ACP fonctionnelle de $\mathbf{X}$. Les auteurs déclinent la distribution limite sous H_0 et montrent que la puissance du test correspondant converge vers un sous H_1. Le principal inconvénient de cette démarche est que le nombre k de composantes impliquées dans la statistique doit être fixé. Comme dans le cadre de l'estimation de θ, fixer k est un problème difficile. Pour le contourner, il est possible d'appliquer une approche par permutation [3] ou d'utiliser du bootstrap [6, 7]. Même si les niveaux des tests correspondants sont contrôlés asymptotiquement, il n'y a toujours pas de garantie théorique sur les puissances.

3 Notre démarche

Notre démarche se fait en deux étapes. Nous introduisons des tests non-adaptatifs de type Fisher, $T_{\alpha,k}$, basés sur les projections de $\mathbf{Y}$ sur les k premières composantes principales de $\mathbf{X}$. Nous combinons ensuite ces tests avec des techniques de test multiple.

Nous notons Γ l'opérateur de covariance de X, et $\widehat{\Gamma}_n$ l'opérateur empirique correspondant. Nous utilisons dans la suite des procédures de test basées sur la décomposition de Karhunen-Loève (KL) de X, qui consiste à projeter X (et θ) dans la base des fonctions propres de Γ, notées $(V_j)_{j\geq 1}$.

3.1 Première étape : tests paramétriques

Soit k un entier positif plus petit que $n/2$. Nous considérons tout d'abord le test paramétrique des hypothèses

$$H_0 : \text{ “}\theta = 0\text{”} \quad \text{contre} \quad H_{1,k} : \text{ “}\theta \in \mathrm{Vect}[(V_j)_{j=1,\ldots,k}] \setminus \{0\}\text{”} \ . \tag{2}$$

Notons $\hat{k}^{KL}$ pour $k \wedge \mathrm{Rank}(\widehat{\Gamma}_n)$. Nous pouvons reformuler le modèle (1) par le modèle linéaire de dimension finie

$$\mathbf{Y} = \mathbf{W}\vartheta + \tilde{\boldsymbol{\epsilon}} \ .$$

où la matrice de design $\mathbf{W}$ est définie par $\mathbf{W}_{i,j} = \langle X_i, \widehat{V}_j\rangle$ pour $i = 1, \ldots n$, $j = 1, \ldots \hat{k}^{KL}$, le vecteur ϑ est défini par $\vartheta_j = \langle \theta, \widehat{V}_j\rangle$, $j = 1, \ldots \hat{k}^{KL}$, et le vecteur $\tilde{\boldsymbol{\epsilon}}$ par $\tilde{\epsilon}_i = \epsilon_i + \langle X_i, \theta\rangle - [\mathbf{W}\vartheta]_i$, pour $i = 1, \ldots n$.

Intuitivement, tester “$\vartheta = 0$” est relativement proche de tester H_0 contre $H_{1,k}$. Pour cette raison, nous introduisons tout d'abord la statistique suivante, de type Fisher :

$$\phi_k(\mathbf{Y}, \mathbf{X}) := \frac{\|\widehat{\boldsymbol{\Pi}}_k \mathbf{Y}\|_n^2}{\|\mathbf{Y} - \widehat{\boldsymbol{\Pi}}_k \mathbf{Y}\|_n^2/(n - \hat{k}^{KL})} \ , \tag{3}$$

où $\widehat{\mathbf{\Pi}}_k$ est la projection orthogonale dans $\mathbb{R}^n$ sur l'espace engendré par les $\hat{k}^{KL}$ colonnes de $\mathbf{W}$. La différence principale avec une statistique de Fisher classique vient du fait que la projection $\widehat{\mathbf{\Pi}}_k$ est aléatoire.

Nous rejetons H_0 contre $H_{1,k}$ quand la statistique

$$T_{\alpha,k} := \phi_k(\mathbf{Y}, \mathbf{X}) - \hat{k}^{KL} \bar{\mathcal{F}}^{-1}_{\hat{k}^{KL}, n-\hat{k}^{KL}}(\alpha) \ . \tag{4}$$

est positive.

Résultats :

Sous des hypothèses de moment pour ε et des hypothèses assez faibles sur Γ, nous avons montré que le niveau de ce test est plus petit qu'un seuil α donné à un terme supplémentaire $\log^{-1}(n)$ près. Nous avons également donné un contrôle fin de la puissance. Ces résultats sont comparables à des résultats standards en régression non paramétrique. La difficulté a été de contrôler l'aléa des composantes principales de $\mathbf{X}$. La plupart des résultats sur les tests basés sur la décomposition de KL sont étudiés dans un cadre asymptotique. Nos résultats ici sont non asymptotiques et reposent sur des hypothèses moins restrictives que celles habituellement utilisées. Nous avons étudié l'optimalité de $T_{\alpha,k}$ au sens minimax. Cette notion est relative à la distance de séparation d'un test sur une certaine classe de fonctions Θ. Intuitivement, la puissance d'un test raisonnable T_α doit être grande quand la norme de θ est grande. De même, la puissance de T_α doit être proche de α quand θ est proche de 0. Pour le problème du test H_0 : "$\theta = 0$" contre $H_{1,\Theta}$: "$\theta \in \Theta \setminus \{0\}$", la distance de séparation correspond à la plus petite distance ρ telle que T_α rejette H_0 avec une probabilité supérieure à $1-\beta$ pour tout $\theta \in \Theta$ dont la norme est plus grande que ρ. Plus la distance de séparation est petite, plus le test T_α est puissant. La distance de séparation minimax sur Θ est la plus petite distance de séparation qui est atteinte par un test de niveau α. Un test réalisant cette distance de séparation minimax est dit être minimax sur Θ. Les distances de séparation de nos procédures de test sont contrôlées non asymptotiquement. Nous avons déduit la distance de séparation minimax dans le modèle fonctionnel (1) pour une large classe d'ellipsoïdes. Nous avons montré que le test paramétrique $T_{\alpha,k}$ atteint le taux optimal de détection lorsque la dimension k est convenablement choisie.

3.2 Deuxième étape : procédure de test multiple

En pratique, la régularité de θ est inconnue. Mais le choix de k dépend de quantités inconnues telles que la régularité de X ou celle de θ. Supposer a priori que la fonction θ appartient à une classe particulière de régularité Θ et construire un test optimal sur Θ peuvent donc conduire à de mauvaises performances, par exemple si $\theta \notin \Theta$. Pour cette raison, un enjeu plus ambitieux est de construire une procédure de test adaptative minimax, c'est à dire, une procédure qui est simultanément minimax pour un large éventail de classes de régularité de θ. Des procédures de test minimax adaptatives ont déjà été étudiées dans le cadre de la régression non paramétrique, d'un point de vue asymptotique [10] et non asymptotique [1].

Nous combinons dans cette deuxième étape les tests paramétriques avec des techniques de tests multiples dans l'esprit de [1]. Dans la suite, nous notons $\mathcal{K}_n$ un ensemble "dyadique" de dimensions, défini par

$$\mathcal{K}_n = \{2^0, 2^1, 2^2, 2^3 \ldots, \bar{k}_n\} \ , \tag{5}$$

où $\bar{k}_n = 2^{\lfloor \log_2 n \rfloor - 1}$. Comme k ne peut pas être choisi a priori, nous évaluons la statistique $\phi_k(\mathbf{Y}, \mathbf{X})$ pour tous les k appartenant à $\mathcal{K}_n$. Nous rejetons H_0 : "$\theta = 0$" quand la statistique

$$T_\alpha := \sup_{k \in \mathcal{K}_n,\ k \leq \mathrm{Rank}(\widehat{\Gamma}_n)} \left[\phi_k(\mathbf{Y}, \mathbf{X}) - \hat{k}^{KL} \bar{\mathcal{F}}^{-1}_{\hat{k}^{KL}, n - \hat{k}^{KL}} \{\alpha_{\mathcal{K}_n}(\mathbf{X})\} \right] \tag{6}$$

est positive, où le poids $\alpha_{\mathcal{K}_n}(\mathbf{X})$ est choisi selon l'une des deux procédures suivantes :
$\mathbf{P_1}$: (Bonferroni) $\alpha_{\mathcal{K}_n}(\mathbf{X})$ est égale à $\alpha/|\mathcal{K}_n|$.
$\mathbf{P_2}$: Soit $\mathbf{Z}$ un vecteur Gaussien de taille n. Nous prenons $\alpha_{\mathcal{K}_n}(\mathbf{X}) = q_{\mathbf{X},\alpha}$, le quantile α de la distribution de la variable aléatoire

$$\inf_{k \in \mathcal{K}_n} \bar{\mathcal{F}}_{\hat{k}^{KL}, n - \hat{k}^{KL}} \left[\phi_k(\mathbf{Z}, \mathbf{X}) / \hat{k}^{KL} \right] \tag{7}$$

conditionnellement à $\mathbf{X}$.

Le test utilisant T_α associé à P_1 correspond à une procédure de test multiple de Bonferroni. Par contraste, celui associé à P_2 appréhende mieux la dépendance entre les statistiques ϕ_k, en utilisant un quantile $q_{\mathbf{X},\alpha}$ ad hoc.

Résultats :

Les procédures sont complètement basées sur les données : aucun paramètre de réglage n'est nécessaire, dont les valeurs optimales pourraient dépendre de θ, de la distribution de X ou de σ. Leurs niveaux et puissances sont analysés d'un point de vue non asymptotique. Nous avons montré que nos procédures de tests multiples sont simultanément minimax sur la même classe d'ellipsoïdes (à un facteur $\log \log n$ près) que celle de $T_{\alpha,k}$. Comme dans le cadre de l'estimation dans [8], les distances de séparation minimax impliquent la régularité commune de θ et X.

Bibliographie

[1] Baraud, Y., Huet, S., and Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *Ann. Statist.* **31**, 1, 225–251.

[2] Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Statist.* **30**, 1, 241–255.

[3] Cardot, H., Goia, A., and Sarda, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.* **33**, 1, 179–199.

[4] Cardot, H. and Sarda, P. (2010). Functional linear regression. In *Handbook of Functional Data Analysis*, F. Ferraty and Y. Romain, Eds. Oxford University Press, Oxford, 21–46.

[5] Comte, F. and Johannes, J. (2011). Adaptive functional linear regression.

[6] Cuevas, A. and Fraiman, R. (2004). On the bootstrap methodology for functional data. In *COMPSTAT 2004—Proceedings in Computational Statistics*. Physica, Heidelberg, 127–135.

[7] González-Manteiga, W. and Martínez-Calvo, A. (2011). Bootstrap in functional linear regression. *J. Statist. Plann. Inference* **141**, 1, 453–461.

[8] Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 1, 70–91.

[9] Hilgert, N., Mas, A., and Verzelen, N. (2013). Minimax adaptive tests for the functional linear model. *Ann. Statist. (A paraître)*

[10] Spokoiny, V. G. (1996). Adaptative hypothesis testing using wavelets. *Ann. Statist.* **24**, 2477–2498.