Prédiction en régression linéaire fonctionnelle avec variable d'intérêt fonctionnelle

Christophe Crambes, André Mas

Prédiction en régression linéaire fonctionnelle avec variable d'intérêt fonctionnelle

Christophe Crambes & André Mas

Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex, France

Résumé. Ce travail concerne l'étude de la prédiction dans le modèle linéaire fonctionnel lorsque la variable d'intérêt est elle aussi fonctionnelle. Nous introduisons un prédicteur basé sur les décompositions de Karhunen-Loève des courbes X (variable explicative) et Y (variable d'intérêt). Les résultats obtenus permettent de fournir un développement asymptotique de la moyenne quadratique de l'erreur de prédiction. Nous donnons également un résultat d'optimalité pour ces vitesses dans un sens minimax, ainsi qu'un théorème de la limite centrale du prédicteur.

Abstract. This work concerns the prediction problem in the functional linear model with functional output. We introduce a predictor based on Karhunen-Loève decompositions of the curves X (covariate) and Y (output). Our results give an asymptotic development of the mean square prediction error. We also give an optimality result for these rates of convergence in a minimax sense, as well as a central limit theorem for the predictor.

Mots clés. Modèle linéaire fonctionnel, variable d'intérêt fonctionnelle, décomposition de Karhunen-Loève, erreur de prédiction, vitesses optimales, théorème de la limite centrale.

1. Introduction

Les modèles de régression, permettant d'expliquer comment une variable d'intérêt Y est reliée à une variable explicative X, sont parmi les plus utilisés en statistique. Nous nous plaçons dans ce cadre de travail, en supposant que les variables X et Y sont à valeurs dans l'espace $L^2(I)$ des fonctions de carré intégrable sur un intervalle I, qui sera considéré comme $[0,1]$ pour simplifier. Ce type de variables aléatoires dites fonctionnelles permet de prendre en compte de nombreuses situations pratiques où les observations sont par nature des courbes (fonctions du temps par exemple). Ces données étant très présentes dans de nombreuses applications, les travaux concernant l'étude des données fonctionnelles se multiplient actuellement à très grande vitesse. Les ouvrages de référence actuels en la matière sont les monographies de Ramsay et Silverman (2002, 2005), qui donnent une vue d'ensemble sur ce champ de recherche, tandis que la monographie de Ferraty et Vieu

(2006) recense les principaux résultats obtenus dans un contexte non-paramétrique sur les données fonctionnelles.

On considère dans la suite un modèle qui s'écrit sous la forme

$$Y(t) = \int_0^1 \mathcal{S}(s,t) X(s) ds + \varepsilon(t), \quad \mathbb{E}(\varepsilon|X) = 0, \tag{1}$$

où $\mathcal{S}$ est un noyau intégrable. Ce modèle, encore peu étudié, a fait l'objet de quelques travaux, parmi lesquels Chiou, Müller et Wang (2004), Yao, Muller et Wang (2005) qui proposent une estimation de $\mathcal{S}$ basée sur une analyse en composantes principales des courbes X et Y. Une des premières études est due à Cuevas, Febrero et Fraiman (2002). Récemment, Antoch *et al.* (2008) ont étudié un estimateur spline de $\mathcal{S}$ tandis que Aguilera, Ocaña and Valderrama (2008) en ont proposé un estimateur à base d'ondelettes. Le modèle (1) peut s'écrire sous la forme $Y(t) = S(X)(t) + \varepsilon(t)$ où l'opérateur S est défini par $Sf(t) = \int_0^1 \mathcal{S}(s,t) f(s) ds$ pour toute fonction f de L^2.

Dans la suite, on considère un échantillon $(X_i, Y_i)_{i=1,\ldots,n}$ d'observations indépendantes et de même loi que (X, Y) sur lequel on se base pour construire notre prédicteur.

2. Construction du predicteur

On introduit les notations suivantes. Le produit scalaire usuel de L^2 est noté $\langle ., . \rangle$ et défini par $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$ pour toutes fonctions f et g de L^2. Le produit tensoriel entre deux fonctions f et g de L^2 est défini par $f \otimes g = \langle g, . \rangle f$ et associe à toute fonction h de L^2 la fonction $\langle g, h \rangle f$. Partant du modèle (1), il vient

$$\mathbb{E}[Y \otimes X] = \mathbb{E}[S(X) \otimes X] + \mathbb{E}[\varepsilon \otimes X].$$

En notant

$$\Delta = \mathbb{E}[Y \otimes X], \quad \Gamma = \mathbb{E}[X \otimes X],$$

on en déduit $\Delta = S\Gamma$. En introduisant les versions empiriques des opérateurs Δ et Γ par

$$\Delta_n = \frac{1}{n} \sum_{i=1}^n Y_i \otimes X_i, \quad \Gamma_n = \frac{1}{n} \sum_{i=1}^n X_i \otimes X_i,$$

un estimateur naturel de S est donné par $\widehat{S}_n$ vérifiant $\Delta_n = \widehat{S}_n \Gamma_n$. Le problème est que l'opérateur Γ_n ne peut pas être directement inversé. Une solution classique consiste à considérer un inverse régularisé. Pour cela, on note $\left(\widehat{\lambda}_j, \widehat{e}_j\right)$ les éléments propres de Γ_n (les valeurs propres étant rangées par ordre décroissant). De façon analogue, (λ_j, e_j) désignent les éléments propres de Γ. L'opérateur Γ_n s'écrit alors $\Gamma_n = \sum_j \widehat{\lambda}_j (\widehat{e}_j \otimes \widehat{e}_j)$ et son inverse régularisé est donné par

$$\Gamma_n^\dagger = \sum_{j=1}^{k} \widehat{\lambda}_j^{-1} \left(\widehat{e}_j \otimes \widehat{e}_j\right), \tag{2}$$

où $k = k_n$ est le nombre de composantes principales choisies. De cette décomposition se déduit une expression de l'estimateur de $\mathcal{S}$ par

$$\widehat{\mathcal{S}}_n(s,t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} \frac{\int X_i \widehat{e}_j}{\widehat{\lambda}_j} Y_i(t)\, \widehat{e}_j(s).$$

L'estimateur $\widehat{S}_n$ de S est défini par $\widehat{S}_n = \Delta_n \Gamma_n^\dagger$ et le prédicteur associé est donné par $\widehat{Y}_{n+1} = \widehat{S}_n(X_{n+1}) = \Delta_n \Gamma_n^\dagger(X_{n+1})$ pour une nouvelle observation X_{n+1}.

3. Résultats asymptotiques

3.1. Hypothèses

Les hypothèses permettant d'établir nos résultats sont les suivantes.

(H.1) On suppose que S est un opérateur de Hilbert-Schmidt: pour toute base $(e_j)_{j\in\mathbb{N}}$ de H, on a

$$\sum_{j,\ell} \langle S(e_\ell), e_j\rangle^2 < +\infty.$$

(H.2) Considérons la décomposition de Karhunen-Loève de X qui s'écrit

$$X = \sum_{j=1}^{+\infty} \sqrt{\lambda_j}\xi_j e_j \quad p.s.,$$

où les ξ_j sont des variables aléatoires centrées réduites et non corrélées. On suppose que, pour $j, \ell \in \mathbb{N}$, il existe une constante b telle que

$$\mathbb{E}\left(|\xi_j|^\ell\right) \leq \frac{\ell!}{2} b^{\ell-2} \cdot \mathbb{E}\left(|\xi_j|^2\right).$$

(H.3) Soit λ la fonction définie par $\lambda(j) = \lambda_j$ pour tout entier j (les λ_j étant les valeurs propres de l'opérateur Γ). On interpole cette fonction de façon continue entre j et $j+1$ telle que

$$x \to \lambda(x) \text{ est convexe.}$$

L'hypothèse (H.1) équivaut à supposer que le noyau $\mathcal{S}$ est doublement intégrable. Remarquons qu'en dehors de cette hypothèse, on ne suppose rien d'autre sur S, en particulier aucune hypothèse de régularité n'est requise. L'hypothèse (H.2) a comme conséquence

de faire une hypothèse de moment (d'ordre 4) sur X. Cette hypothèse est par exemple vérifiée lorsque X est un processus gaussien ou encore un processus borné p.s. L'hypothèse $(H.3)$ est un hypothèse de décroissance sur les valeurs propres de Γ. Elle est vérifiée pour une large classe d'opérateurs, dont les valeurs propres sont à décroissance arithmétique, exponentielle ..., y compris pour des processus X très irréguliers.

3.2. Erreur de prédiction en moyenne quadratique

On note $\Gamma_\varepsilon = \mathbb{E}(\varepsilon \otimes \varepsilon)$ l'opérateur de covariance du bruit et $\sigma_\varepsilon^2 = \mathrm{tr}\Gamma_\varepsilon$. On a alors, sous les hypothèses précédentes

$$\mathbb{E}\left\|\widehat{S}_n(X_{n+1}) - S(X_{n+1})\right\|^2 = \sigma_\varepsilon^2 \frac{k}{n} + \sum_{j=k+1}^{+\infty} \lambda_j \|S(e_j)\|^2 + A_n + B_n, \tag{3}$$

où $A_n \leq C_A \frac{k^2 \lambda_k}{n} \|S\|_{\mathcal{L}_2}$ et $B_n \leq C_B \frac{k^2 (\log k)}{n^2}$, les constantes C_A et C_B ne dépendant pas de k, n ou S.

3.3. Optimalité

Notre estimateur est construit de façon très proche de celui de Yao, Muller et Wang (2005). Notre apport concerne un résultat, énoncé ci-dessous, donnant des vitesses optimales de convergence de l'estimateur. Pour toute fonction $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ de classe C^1 et décroissante telle que $\sum_{j=1}^{+\infty} \varphi(j) = 1$, on note $\mathcal{L}_2(\varphi, L)$ la classe des opérateurs linéaires de H dans H définie par

$$\mathcal{L}_2(\varphi, L) = \left\{T \in \mathcal{L}_2, \|T\|_{\mathcal{L}_2} \leq L : \|T(e_j)\| \leq L\sqrt{\varphi(j)}\right\}.$$

Si on note de plus $L = \left\|S\Gamma^{1/2}\right\|_{\mathcal{L}_2}$, $\varphi(j) = \lambda_j \|S(e_j)\|^2 / L^2$ et k_n^* la partie entière de la solution en x de l'équation

$$\frac{1}{x}\int_x^{+\infty} \varphi(x)\, dx = \frac{1}{n}\frac{\sigma_\varepsilon^2}{L^2},$$

on a alors (comme conséquence de la sous-section précédente):

$$\limsup_{n\to+\infty} \frac{n}{k_n^*} \sup_{S\Gamma^{1/2} \in \mathcal{L}_2(L,\varphi)} \mathbb{E}\left\|\widehat{S}_n(X_{n+1}) - S(X_{n+1})\right\|^2 = 2\sigma_\varepsilon^2.$$

Le résultat d'optimalité peut à présent être énoncé:

$$\inf_{\widehat{S}_n} \sup_{S \in \mathcal{L}_2(\varphi, C)} \mathbb{E}\left\|\widehat{S}_n(X_{n+1}) - S(X_{n+1})\right\|^2 \asymp \frac{k_n^*}{n}.$$

3.4. Convergence faible

Notre principal résultat est donné ci-dessous. Sous les hypothèses précédentes et sous la condition que $(k \log k)^2 / n \to 0$, alors

$$\sqrt{\frac{n}{k}} \left[\widehat{S}_n (X_{n+1}) - S\Pi_k (X_{n+1}) \right] \xrightarrow{w} \mathcal{G}_\varepsilon$$

où $\mathcal{G}_\varepsilon$ est une variable aléatoire gaussienne à valeurs dans H, centrée et d'opérateur de covariance Γ_ε. Sous certaines conditions, ce résultat peut alors s'écrire

$$\sqrt{\frac{n}{k}} \left[\widehat{S}_n (X_{n+1}) - S (X_{n+1}) \right] \xrightarrow{w} \mathcal{G}_\varepsilon.$$

Une des conséquences de ce résultat est que, pour un choix de $H = W_0^{2,1}([0,1]) = \{f \in L^2([0,1]) : f(0) = 0, f' \in L^2([0,1])\}$, on obtient, pour un t_0 fixé dans $[0,1]$

$$\mathbb{P}\left(Y_{n+1}^* (t_0) \in \left[\widehat{Y}_{n+1} (t_0) \pm \sqrt{\frac{k}{n}} \sigma_{t_0} q_{1-\alpha/2} \right] \right) = 1 - \alpha,$$

avec $\sigma_{t_0}^2 = \Gamma_\varepsilon (t_0, t_0)$.

Bibliographie

[1] Aguilera, A., Ocaña, F. and Valderrama, M. (2008). Estimation of functional regression models for functional responses by wavelet approximation. *International Workshop on Functional and Operatorial Statistics 2008 Proceedings, Functional and operatorial statistics*, Dabo-Niang and Ferraty (Eds.), Physica-Verlag, Springer.

[2] Antoch, J., Prchal, L., De Rosa, M. and Sarda, P. (2008). Functional linear regression with functional response: application to prediction of electricity consumption. *International Workshop on Functional and Operatorial Statistics 2008 Proceedings, Functional and operatorial statistics*, Dabo-Niang and Ferraty (Eds.), Physica-Verlag, Springer.

[3] Chiou, J-M., Müller, H-G. and Wang J-L. (2004). Functional response models. *Statistica Sinica*, **14**, 659-677.

[4] Cuevas, A., Febrero, M. and Fraiman, R. (2002). Linear functional regression : the case of a fixed design and a functional response. *Canadian Journal of Statistics*, **30**, 285-300.

[5] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: methods, theory, applications and implementations*. Springer-Verlag, London.

[6] Ramsay, J.O. and Silverman, B.W. (2002). *Applied functional data analysis*. Springer-Verlag.

[7] Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis* (Second Ed.). Springer, New York.

[8] Yao F., Müller H-G. and Wang, J-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.*, **33**, 2873-2903.