

Estimation consistante des paramètres d'un modèle non linéaire pour des données fonctionnelles discrétisées aléatoirement

Fabrice Rossi, Brieuc Conan-Guez

▶ **To cite this version:**

Fabrice Rossi, Brieuc Conan-Guez. Estimation consistante des paramètres d'un modèle non linéaire pour des données fonctionnelles discrétisées aléatoirement. Comptes rendus de l'Académie des sciences. Série I, Mathématique, 2005, 340 (2), pp.167-170. 10.1016/j.crma.2004.11.029 . inria-00000223

HAL Id: inria-00000223
https://inria.hal.science/inria-00000223

Submitted on 15 Sep 2005

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.

inria-00000223, version 1 - 15 Sep 2005

Estimation consistante des paramètres d'un modèle non linéaire pour des données fonctionnelles discrétisées aléatoirement ⋆

Consistent estimation of parameters in a nonlinear model for functional data with randomly chosen evaluation points

Fabrice ROSSI [a,b] Brieuc CONAN-GUEZ [b]

[a] *CEREMADE, UMR CNRS 7534,*

Université Paris-IX Dauphine,

Place du Maréchal de Lattre de Tassigny,75016 Paris, France

[b] *Projet AxIS, INRIA Rocquencourt,*

Domaine de Voluceau, Rocquencourt, B.P. 105

78153 Le Chesnay Cedex, France

⋆ Version finale parue dans *Comptes rendus de l'Académie des Sciences - Série I*, 340(2) :167–170, Janvier 2005

http://dx.doi.org/10.1016/j.crma.2004.11.029

Abstract

We study a class of nonlinear regression models for scalar or vectorial response when the explanatory variable is a function. We introduce a consistent estimator of the parameters of models in this class when functions are evaluated at randomly chosen observation points.

Résumé

Nous étudions une classe de modèles de régression non linéaires quand la variable explicative est fonctionnelle alors que la variable réponse est scalaire ou vectorielle. Nous construisons un estimateur consistant des paramètres des modèles de cette classe pour le cas de fonctions discrétisées à des positions aléatoires.

1 Introduction

Dans de nombreux domaines, les individus à traiter sont naturellement décrits par une ou plusieurs fonctions régulières (météorologie, économie, reconnaissance de la parole, etc..). Le but de l'Analyse de Données Fonctionnelles (A.D.F.) [4] est d'adapter les techniques classiques de l'analyse de données traditionnelles à ce type de description.

Nous présentons dans cette Note une classe de modèles de régression non linéaires d'une variable aléatoire (v.a.) vectorielle sur une v.a. à valeurs dans un espace fonctionnel. Ces modèles généralisent le modèle linéaire fonctionnel

Email addresses: Fabrice.Rossi@inria.fr (Fabrice ROSSI),
Brieuc.Conan-Guez@inria.fr (Brieuc CONAN-GUEZ).

[2]. En A.D.F., on considère, en accord avec la pratique expérimentale, que les fonctions observées sont discrétisées. Dans cette Note, nous considérons que les points de discrétisation des fonctions observées sont choisis aléatoirement, ce qui généralise le modèle déterministe habituellement utilisé en A.D.F.

2 Une classe de modèles fonctionnels paramétriques

Dans la suite de cette Note, on suppose que toutes les v.a. sont définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$. Soient $\mathcal{K}$ une partie compacte de $\mathbb{R}^u$ et μ une mesure de Borel sur $\mathcal{K}$. On désigne par $C(\mathcal{K}, \mathbb{R})$ l'ensemble de fonctions continues de $\mathcal{K}$ dans $\mathbb{R}$ muni de la norme uniforme. Soient X une v.a. à valeurs dans $C(\mathcal{K}, \mathbb{R})$ et Y une v.a. à valeurs dans $\mathbb{R}^t$. On s'intéresse à une classe de modèles de régression de Y en X de la forme suivante :

$$E(Y|X) = u\left(\theta_0, \int f_1(\theta_1, .)X \, \mathrm{d}\mu, \ldots, \int f_k(\theta_k, .)X \, \mathrm{d}\mu\right),$$

où k est un entier strictement positif supposé connu et où (hypothèses H_a) :

(1) pour $0 \leq l \leq k$, le vecteur θ_l est élément de Θ_l, une partie compacte de $\mathbb{R}^{v_l}$ (on note $\Theta = \Theta_0 \times \Theta_1 \times \ldots \times \Theta_k$ et $\theta = (\theta_0, \theta_1, \ldots, \theta_k)$)

(2) pour $1 \leq l \leq k$, f_l est une fonction de $\Theta_l \times \mathcal{K}$ vers $\mathbb{R}$ telle que :

  (a) pour chaque $z \in \mathcal{K}$, $f_l(., z)$ est continue ;

  (b) pour chaque $\theta_l \in \Theta_l$, $f_l(\theta_l, .)$ est mesurable ;

  (c) il existe $d_l \in L^q(\mu)$ telle que $\forall\ (\theta_l, z) \in \Theta_l \times \mathcal{K}$, $|f_l(\theta_l, z)| \leq d_l(z)$ (q ne dépend pas de l) ;

(3) u est une fonction bornée uniformément continue de $\Theta_0 \times \mathbb{R}^k$ vers $\mathbb{R}^t$.

Outre k, les ensembles Θ_l, les fonctions f_l et la fonction u sont supposés connus. Pour un modèle particulier de la classe, les seuls paramètres à estimer

sont donc les coordonnées du vecteur $\theta \in \Theta$. Pour simplifier la suite de cette Note, on notera $h(\theta, X) = u\left(\theta_0, \int f_1(\theta_1, .)X \, \mathrm{d}\mu, \ldots, \int f_k(\theta_k, .)X \, \mathrm{d}\mu\right)$.

Un choix approprié pour u permet de s'intéresser au cas particulier des modèles de régression basés sur le perceptron multi-couches fonctionnel [5]. Ce dernier est un approximateur universel, ce qui motive l'utilisation du modèle en pratique : toute fonction continue d'un compact de $C(\mathcal{K}, \mathbb{R})$ dans $\mathbb{R}$ peut être approchée uniformément sur ce compact à une précision arbitraire par une fonction calculable exactement par un tel perceptron. La classe de modèles proposée n'est donc pas limitée *a priori*, contrairement au modèle linéaire par exemple.

Les fonctions f_l peuvent être choisies en pratique dans diverses classes. On peut par exemple fixer un système libre de fonctions de $L^q(\mu)$, les $(\phi_i)_{1 \leq i \leq v_l}$, et poser $f_l(\theta_l, z) = \sum_{i=1}^{v_l} \theta_{l,i} \phi_i(z)$ ($\theta_{l,i}$ désigne la i-ème coordonnée de θ_l). On peut aussi utiliser un perceptron multi-couches classique.

3 Cas de fonctions parfaitement connues

On considère maintenant un modèle fixé dans la classe décrite dans la section précédente (k, u et les f_l sont donc supposés connus). On cherche à estimer les paramètres du modèle (i.e., θ) à partir des n premiers éléments, $(X_i, Y_i)_{1 \leq i \leq n}$, d'une suite i.i.d. de v.a. distribuées comme (X, Y). Pour ce faire, on se donne une fonction continue c sur $\mathbb{R}^t \times \mathbb{R}^t$, comme par exemple la distance quadratique, qui mesure l'adéquation entre le modèle et les observations. On définit ainsi l'erreur théorique réalisée par le modèle pour le paramètre θ par $\lambda(\theta) = E(c(Y, h(\theta, X)))$ et on note Θ^* l'ensemble des minimiseurs de $\lambda(\theta)$.

En général, le modèle n'est pas identifiable et l'estimation de ses paramètres consiste donc seulement à trouver un θ dans Θ^*.

On estime $\lambda(\theta)$ par l'erreur empirique réalisée par le modèle, $\widehat{\lambda}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} c(Y_i, h(\theta, X_i))$. On désigne par $\widehat{\theta}_n$ un minimiseur de cette erreur empirique. On démontre alors que l'estimateur $\widehat{\lambda}_n$ converge presque sûrement uniformément sur Θ vers $\lambda(\theta)$ (la preuve est basée sur [1]), et que, par conséquent, presque toute suite de paramètres optimaux $\widehat{\theta}_n$ converge vers Θ^*. En plus de la continuité de c, ce résultat nécessite l'existence d'une function mesurable c_{max} de $\mathbb{R}^t$ dans $\mathbb{R}$ telle que (hypothèses H_b) :

(1) $\forall \theta \in \Theta, x \in C(\mathcal{K}, \mathbb{R}), y \in \mathbb{R}^t, c(y, h(\theta, x)) \leq c_{max}(y)$;

(2) $E(c_{max}(Y)) < \infty$.

Il est important de noter qu'on ne fait pas l'hypothèse qu'il existe θ tel que $E(Y|X) = h(\theta, X)$: Y est simplement une v.a. à valeurs dans $\mathbb{R}^t$ qui vérifie H_b-(2).

Par contre, ce premier résultat de consistance impose que les fonctions observées soient connues parfaitement et que les calculs d'intégrales soient réalisés de façon exacte. Dans la pratique cependant, les fonctions à traiter sont discrétisées. La section suivante définit un estimateur de l'erreur théorique tenant compte de la discrétisation.

4 Prise en compte de la discrétisation

Dans les sections précédentes, la mesure d'intégration μ était une mesure de Borel quelconque définie sur $\mathcal{K}$. Nous considérons maintenant que $\mu = P_Z$, la probabilité induite sur $\mathcal{K}$ par une variable aléatoire Z (à valeurs dans $\mathcal{K}$). Nous

supposons que toutes les fonctions observées sont discrétisées en des points choisis selon μ. Ceci permet de tenir compte de variabilités dans le processus d'acquisition des données fonctionnelles considérées (cf [4] pour des exemples concrets). Comme dans la section précédente, $(X_i, Y_i)_{i\in\mathbb{N}}$ est une suite i.i.d. de v.a. distribuées comme (X, Y), mais nous n'avons plus une connaissance exacte des X_i. Plus précisément, nous faisons les hypothèses suivantes (hypothèses H_c) :

(1) $(Z_{i,j})_{i\in\mathbb{N},j\in\mathbb{N}}$ est une suite bi-indicée de v.a. à valeurs dans $\mathcal{K}$, toutes distribuées comme Z ;

(2) pour tout i, les v.a. $(Z_{i,j})_{j\in\mathbb{N}}$ sont indépendantes ;

(3) $(\mathcal{E}_{i,j})_{i\in\mathbb{N},j\in\mathbb{N}}$ est une suite bi-indicée de v.a. à valeurs dans $\mathbb{R}$, identiquement distribuées ;

(4) pour tout i, les v.a. $(\mathcal{E}_{i,j})_{j\in\mathbb{N}}$ sont indépendantes ;

(5) $E(\mathcal{E}_{i,j}) = 0$ et $E(|\mathcal{E}_{i,j}|^p) < \infty$ avec p et q exposants conjugués (voir l'hypothèse H_a-(2)-(c)).

Pour tout i, les $(Z_{i,j})_{j\in\mathbb{N}}$ sont les points de discrétisation de la fonction X_i, alors que les $(\mathcal{E}_{i,j})_{j\in\mathbb{N}}$ sont des bruits d'observations. En d'autres termes, nous observons la suite bi-indicée de v.a. $(V_{i,j})_{i\in\mathbb{N},j\in\mathbb{N}}$ définie par

$$V_{i,j} = X_i(Z_{i,j}) + \mathcal{E}_{i,j}.$$

On peut alors définir un nouvel estimateur pour l'erreur théorique :

$$\widehat{\lambda}_{n,m}(\theta) = \frac{1}{n}\sum_{i=1}^{n} c\left(Y_i, u\left(\theta_0, \frac{1}{m_i}\sum_{j=1}^{m_i} f_1(\theta_1, Z_{i,j})V_{i,j}, \ldots, \frac{1}{m_i}\sum_{j=1}^{m_i} f_k(\theta_k, Z_{i,j})V_{i,j}\right)\right),$$

où $m = \inf_{1\le i\le n} m_i$. En notant $\widehat{\theta}_{n,m}$ un minimiseur de $\widehat{\lambda}_{n,m}$, on a le résultat de consistance suivant :

Théorème 1 *Sous les hypothèses H_a,H_b et H_c, on a presque sûrement :*

$$\lim_{n\to\infty}\lim_{m\to\infty} d(\widehat{\theta}_{n,m},\Theta^*)=0,$$

où $d(a,B)$ désigne la distance de $a\in\Theta$ à $B\subset\Theta$, c'est-à-dire $\inf_{b\in B}\|a-b\|$, où $\|.\|$ désigne la norme euclidienne sur Θ.

Le résultat de convergence ci-dessus est séquentiel : le nombre de points d'observation (m) nécessaires à l'obtention d'une précision donnée dépend du nombre de fonctions (n). On remarque aussi que les hypothèses sur les fonctions observées (continues d'un compact dans $\mathbb{R}$) sont assez faibles et comparables aux hypothèses classiquement utilisées en Analyse de Données Fonctionnelles (par exemple des fonctions hölderiennes dans [3]), sans pour autant nécessiter une discrétisation déterministe.

Résumé de la preuve :

Pour tout $x\in C(\mathcal{K},\mathbb{R})$ et $\theta_l\in\Theta_l$, on définit $M^m_{i,l}(x,\theta_l)(\omega)=\frac{1}{m}\sum_{j=1}^m f_l(\theta_l,Z_{i,j}(\omega))\left(x(Z_{i,j}(\omega))+\mathcal{E}_{i,j}(\omega)\right)$ et de même $M_l(x,\theta_l)=\int f_l(\theta_l,z)x(z)\,\mathrm{d}\mu(z)$. La démonstration se décompose en 3 étapes.

Première étape : on montre que l'ensemble $B_{i,l}=\{\omega\in\Omega\mid\forall x\in C(\mathcal{K},\mathbb{R}),\lim_{m\to\infty}\sup_{\theta_l\in\Theta_l}\left|M^m_{i,l}(x,\theta_l)(\omega)-M_l(x,\theta_l)\right|=0\}$ est de probabilité 1. Pour ce faire, en utilisant une loi forte des grands nombres uniforme, on montre que pour tout g fixé, $B_{i,l}(x)=\{\omega\in\Omega\mid\lim_{m\to\infty}\sup_{\theta_l\in\Theta_l}\left|M^m_{i,l}(x,\theta_l)(\omega)-M_l(x,\theta_l)\right|=0\}$ est de probabilité 1. Soit alors une suite $(h_t)_{t\in\mathbb{N}}$ dense dans $C(\mathcal{K},\mathbb{R})$ (qui est séparable), et soit l'ensemble $A_{i,l}=\cap_{t\in\mathbb{N}}B_{i,l}(h_t)$. Par intersection dénombrable, $A_{i,l}$ est de probabilité 1. En utilisant la régularité de $M^m_{i,l}(.,\theta_l)(\omega)$ et de $M_l(.,\theta_l)$, on montre que pour tout $\omega\in A_{i,l}$, la propriété de convergence uniforme au point h_t peut

être étendue à un voisinage de h_t. L'ensemble de ces voisinages formant un recouvrement de $C(\mathcal{K}, \mathbb{R})$, on conclut que $A_{i,l} = B_{i,l}$.

Deuxième étape : on montre, en appliquant de nouveau une loi forte des grands nombres uniforme, que l'ensemble $C = \{\omega \in \Omega \mid \sup_{\theta\in\Theta} \left|\widehat{\lambda}_n(\theta)(\omega) - \lambda(\theta)(\omega)\right| \rightarrow_{n\rightarrow\infty} 0\}$ est de probabilité 1 (il s'agit en fait de la première partie du résultat de consistance énoncé en fin de section 3).

Troisième étape : on note $D = C \cap \bigcap_{i\in\mathbb{N} l\in\{1,\ldots,k\}} A_{i,l}$. Soit $\omega \in D$ et $\varepsilon > 0$, alors il existe un rang N et $n > N$ tel que $\sup_{\theta\in\Theta} \left|\widehat{\lambda}_n(\theta)(\omega) - \lambda(\theta)(\omega)\right| < \varepsilon$. On cherche à présent à obtenir le même type de majoration pour $\sup_{\theta\in\Theta} \left|\widehat{\lambda}_{n,m}(\theta)(\omega) - \widehat{\lambda}_n(\theta)(\omega)\right|$. Il suffit de montrer que pour tout $i \leq n$ et pour tout $\theta \in \Theta$, $|c(Y_i(\omega), u(\theta_0, M_{i,1}^m(X_i(\omega), \theta_1)(\omega), \ldots, M_{i,k}^m(X_i(\omega), \theta_k)(\omega))) - c(Y_i(\omega), u(\theta_0, M_1(X_i(\omega), \theta_1), \ldots, M_k(X_i(\omega), \theta_k)))|$ peut être majoré par ε. Ceci est une conséquence directe de l'uniforme continuité en (θ_0, x) de la fonction $c(y, u(\theta_0, x))$, et de l'uniforme convergence de $M_{i,l}^m(X_i(\omega), \theta_l)(\omega)$ vers $M_l(X_i(\omega), \theta_l)$ (cf la première étape de la preuve). La convergence uniforme de $\widehat{\lambda}_{n,m}(\theta)$ vers $\lambda(\theta)$ permet enfin d'établir la convergence de presque toute suite $\widehat{\theta}_{n,m}$ vers l'ensemble Θ^*.

Références

[1] Donald W. K. Andrews. Consistency in nonlinear econometric models : A generic uniform law of large numbers. *Econometrica*, 55(6) :1465–1471, November 1987.

[2] Hervé Cardot, Frédéric Ferraty, and Pascal Sarda. Functional linear model. *Statist. & Prob. Letters*, 45 :11–22, 1999.

[3] Hervé Cardot, Frédéric Ferraty, and Pascal Sarda. Spline estimators for the functional linear model. *Statistica Sinica*, 13 :571–591, 2003.

[4] Jim Ramsay and Bernard Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer Verlag, June 1997.

[5] Fabrice Rossi and Brieuc Conan-Guez. Functional multi-layer perceptron : a nonlinear tool for functional data analysis. *Neural Network*, 2004. In Press.