Estimation consistante d'un modèle paramétrique fonctionnel en présence de discrétisation aléatoire[0]

Fabrice ROSSI*[1] — Brieuc CONAN-GUEZ**

**LISE/CEREMADE, UMR CNRS 7534,*
Université Paris-IX Dauphine,
Place du Maréchal de Lattre de Tassigny,
75016 Paris, France

Fabrice.Rossi@dauphine.fr

***INRIA,*
Domaine de Voluceau, Rocquencourt, B.P. 105
78153 Le Chesnay Cedex, France

Brieuc.Conan-Guez@inria.fr

RÉSUMÉ. *Lors de travaux précédents, on s'est intéressé aux propriétés asymptotiques d'un modèle fonctionnel paramétrique. L'apport de ces travaux résidait dans l'utilisation d'une approche probabiliste afin de modéliser la connaissance discrète des fonctions traitées. Le travail présenté ici est une amélioration de ces précédents résultats : l'hypothèse de compacité de l'ensemble des fonctions observées n'est plus nécessaire.*

ABSTRACT. *In our previous works, we studied the asymptotical properties of a parametric functional model. Our contribution lay in the use of a probalistic approach to model the sampling of observed functions. This present work is an improvement over these previous results: the compacity assumption of the function set is no more necessary.*

MOTS-CLÉS : *analyse de données fonctionnelles, consistance, apprentissage supervisé.*

KEYWORDS: *functional data analysis, consistency, supervised classification.*

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 [RAM 97] est d'adapter les techniques classiques de l'analyse de données traditionnelles à ce type de description.

Dans ce travail, on considère une classe particulière de régresseurs fonctionnels paramétriques. Ces régresseurs permettent une modélisation non-linéaire du lien fonctionnel existant entre la variable explicative (v.a. fonctionnelle) et la variable à expliquer (v.a. réelle). En Analyse de Données Fonctionnelles, il est habituel de supposer que les fonctions observées ne sont connues que de manière discrète (par le biais d'une échantillonnage). Une manière naturelle de prendre en compte cette caractéristique consiste à modéliser les points d'observation des fonctions traitées de manière pro-

0. Publié dans les actes des XXXVèmes journées de Statistique de la SFdS.
Disponible à `http://apiacoa.org/publications/2003/sfds2003.pdf`

1. Les coordonnées actuelles de Fabrice Rossi sont disponibles à l'URL `http://apiacoa.org/`

babiliste. Le résultat énoncé à la fin de cet article s'appuie sur cette modélisation, afin de montrer que l'estimation des paramètres optimaux du modèle est consistante.

2. Modèle fonctionnel paramétrique

Dans la suite de cet article, on suppose que toutes les variables aléatoires sont définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$. On considère l'espace mesuré $(Z, \mathcal{B}, \mu)$, où μ est une mesure de Borel. Le modèle fonctionnel auquel on s'intéresse s'écrit sous la forme suivante :

$$H(w,g) = U\left(w_o, \int F_1(w_1,.)gd\mu, \ldots, \int F_k(w_k,.)gd\mu\right),$$

où la variable explicative g est une fonction définie sur Z et à valeurs réelles, et où w est le vecteur paramètre du modèle. Chaque F_l est un modèle paramétrique (par exemple un modèle linéaire généralisé, ou un perceptron multi-couches) de vecteur paramètre w_l. On note $w = (w_0, w_1, \ldots, w_k)$ un élément de $W = \Pi_{l=0}^{k} W_l$. Enfin, U est une fonction régulière. On peut par exemple choisir U de sorte à obtenir un perceptron multi-couches fonctionnel [ROS 02]. Ce cas particulier est un approximateur universel, ce qui motive l'utilisation du modèle en pratique. On peut noter que le modèle présenté est en fait très similaire à celui proposé dans le contexte de la Régression Inverse par Tranche Fonctionnelle [FER 03].

On considère les hypothèses suivantes (hypothèses H_a) :

1. pour $0 \leq l \leq k$, W_l est une partie compacte de $\mathbb{R}^{v_l}$
2. pour $1 \leq l \leq k$, F_l est une fonction définie de $W_l \times Z$ vers $\mathbb{R}$ telle que :

(a) pour chaque $x \in Z$, $F_l(., x)$ est continue

(b) pour chaque $w_l \in W_l$, $F_l(w_l, .)$ est mesurable

(c) il existe $d_l \in L^q(\mu) / \forall\, (w_l, x) \in W_l \times Z$, $|F_l(w_l, x)| \leq d_l(x)$.

3. U est une fonction bornée uniformément continue de $W_0 \times \mathbb{R}^k$ vers $\mathbb{R}^o$

Si l'on considère le cas particulier du perceptron multi-couches fonctionnel, ces différentes hypothèses sont naturellement vérifiées.

3. Connaissance parfaite des fonctions observées

Comme indiqué dans l'introduction, notre but est d'expliquer la variable $t \in \mathbb{R}^o$ grâce à l'observation g. Pour cela, on veut trouver un vecteur w optimal tel que $H(w, g)$ prédise au mieux la valeur t. On se donne pour cela une mesure de similarité c (par exemple la distance quadratique), et on cherche à minimiser la quantité $c(t, H(w, g))$ "en moyenne". Plus formellement, on considère les hypothèses suivantes (hypothèses H_b) :

1. Z est une partie compacte de $\mathbb{R}^u$
2. (G^i, T^i) est une suite i.i.d de variables aléatoires à valeurs dans $C(Z, \mathbb{R}) \times \mathbb{R}^o$ (i.e., chaque G^i est une fonction mesurable de $(\Omega, \mathcal{A}, P)$ dans $C(Z, \mathbb{R})$ muni de sa tribu borélienne et chaque T^i est une variable aléatoire à valeurs dans $\mathbb{R}^o$).

On définit alors l'erreur théorique réalisée par le modèle paramétrique H pour le paramètre w : $\lambda(w) = E(c(T^1, H(w, G^1)))$. On note W^* l'ensemble des minimiseurs de $\lambda(w)$. Dans la pratique cependant, la quantité $\lambda(w)$ ne peut malheureusement pas être évaluée, car seul un nombre fini d'observations est disponible (les couples (g^i, t^i)). On est donc naturellement amené à considérer l'erreur empirique suivante : $\widehat{\lambda}_n(w) = \frac{1}{n} \sum_{i=1}^{n} c(T^i, H(w, G^i))$, ainsi qu'un minimiseur de cette erreur empirique, noté $\widehat{w}_n$.

On démontre que l'estimateur $\widehat{w}_n$ converge presque sûrement vers W^* (la preuve est basée sur [AND 87]). Ce résultat nécessite l'hypothèse de domination suivante (hypothèses H_c) :

1. $\forall w \in W, g \in C(Z, \mathbb{R}), t \in \mathbb{R}^o, c(t, H(w, g)) \leq c_{max}(t)$
2. $E(c_{max}(T^1)) < \infty$

Ce premier résultat de consistance impose que les fonctions d'entrée soient connues de manière parfaite. Dans la pratique cependant, les fonctions à traiter sont connues sous la forme d'un échantillonnage (nombre fini de couples $(x_j^i, g(x_j^i))$). Dans la section suivante, on démontre un second résultat de consistance prenant en compte l'aspect discret des fonctions.

4. Connaissance discrète des fonctions observées

Afin de modéliser le fait que chaque fonction d'entrée n'est connue qu'en un nombre fini de points d'observation, on utilise le modèle probabiliste suivant (hypothèses H_d) :

1. $(X_j^i)_{i \in \mathbb{N}, j \in \mathbb{N}}$ est une suite de variables aléatoires définies de Ω vers Z.
2. les X_j^i sont identiquement distribuées, la probabilité induite sur Z est $\mu = P_X$.
3. $(\mathcal{E}_j^i)_{i \in \mathbb{N}, j \in \mathbb{N}}$ est une suite de variables aléatoires identiquement distribuées, définies de Ω vers $\mathbb{R}$.
4. $E(\mathcal{E}_j^i) = 0$ et $E(|\mathcal{E}_j^i|^p) < \infty$ avec p et q exposants conjugués (hypothèses H_a).
5. Pour i fixé, la suite de variables aléatoires $(X_0^i, \mathcal{E}_0^i, X_1^i, \mathcal{E}_1^i, \ldots)$ est indépendante.

Pour chaque i, la suite $(X_j^i)_{j \in \mathbb{N}}$ correspond aux points d'observation de la fonction G^i, la suite $(\mathcal{E}_j^i)_{j \in \mathbb{N}}$ correspond aux erreurs commises lors de l'évaluation de la fonction en ces points. Si g^i, x_j^i et ε_j^i sont respectivement des réalisations de G^i, X_j^i et $\mathcal{E}_j^i$, on a donc $y_j^i = g^i(x_j^i) + \varepsilon_j^i$.

Les fonctions n'étant connues que de façon discrète, les intégrales ne peuvent plus être calculées. On remplace donc chaque intégrale $\int F_l(w_l, .) g^i d\mu$ par la moyenne empirique suivante : $\frac{1}{m^i} \sum_{j=1}^{m^i} F_l(w_l, x_j^i) y_j^i$. En conséquence, l'erreur empirique $\widehat{\lambda}_n(w)$ est approchée par l'erreur empirique suivante :

$$\lambda_n^m(w) = \frac{1}{n} \sum_{i=1}^{n} c \left(t^i, U \left(w_0, \frac{1}{m^i} \sum_{j=1}^{m^i} F_1(w_1, x_j^i) y_j^i, \ldots, \frac{1}{m^i} \sum_{j=1}^{m^i} F_k(w_k, x_j^i) y_j^i \right) \right)$$

où $m = inf_{1 \leq i \leq n} m^i$. On note w_n^m un minimiseur de λ_n^m, on a alors le résultat suivant :

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

$$\lim_{n \to \infty} \lim_{m \to \infty} d(w_n^m, W^*) = 0$$

Il est important de noter que le résultat de convergence ci-dessus est séquentiel. En effet, le nombre de points d'observation (m) nécessaire à l'obtention d'une précision donnée dépend du nombre de fonctions (n).

5. Conclusion

On a pu voir au cours de cet article que la modélisation probabiliste des points d'observation était naturelle, et permettait d'obtenir un résultat de consistance satisfaisant. Les hypothèses liées à ce dernier résultat ont, par rapport à nos travaux précédents, été affaiblies : la compacité de l'ensemble des fonctions observées n'est plus nécessaire. On se rapproche ainsi des hypothèses classiquement utilisées en Analyse de Données Fonctionnelles (fonctions hölderiennes, cf [CAR 03]), sans pour autant demander une discrétisation déterministe.

6. Bibliographie

[AND 87] ANDREWS D. W. K., Consistency in Nonlinear Econometric Models : A Generic Uniform Law of Large Numbers, *Econometrica*, vol. 55, n^o 6, 1987, p. 1465–1471.

[CAR 03] CARDOT H., FERRATY F., SARDA P., Spline Estimators for the Functional Linear Model, *Statistica Sinica*, vol. 13, 2003, p. 571–591.

[FER 03] FERRÉ L., YAO A.-F., Functional sliced inverse regression analysis, *Statistics*, vol. 37, n^o 6, 2003, p. 475–488.

[RAM 97] RAMSAY J., SILVERMAN B., *Functional Data Analysis*, Springer Series in Statistics, Springer Verlag, June 1997.

[ROS 02] ROSSI F., CONAN-GUEZ B., Modélisation supervisée de données fonctionnelles par perceptron multi-couches, *Actes des neuvièmes journées de la SFC (conférence invitée)*, Toulouse, France, Septembre 2002, p. 93–100.