

ESTIMATION ADAPTATIVE OPTIMALE DE FONCTIONS MOYENNES ET COVARIANCES IRRÉGULIÈRES

Steven Golovkine¹ & Nicolas Klutchnikoff² & Valentin Patilea³

¹*University of Limerick, steven.golovkine@ul.ie*

²*Université Rennes 2, nicolas.klutchnikoff@univ-rennes2.fr*

³*ENSAI, valentin.patilea@ensai.fr*

Résumé. Nous proposons des estimateurs non-paramétriques pour les fonctions moyenne et covariance de données fonctionnelles. Les courbes aléatoires ne sont pas nécessairement différentiables, de régularité inconnue et mesurées avec erreur sur un ensemble de points discret tirés aléatoirement. La définition de nos estimateurs non-paramétriques dépend de la régularité locale du processus stochastique générant les données. D’abord, nous proposons un estimateur simple de cette régularité locale utilisant les informations intra- et inter-courbes. Ensuite, l’approche “smoothing first, then estimate” est utilisée pour l’estimation des fonctions moyenne et covariance. Ces nouveaux estimateurs non-paramétriques atteignent des vitesses de convergence optimales.

Mots-clés. Données fonctionnelles, Exposant de Hölder, Polynômes locaux

Abstract. We propose nonparametric estimators for the mean and the covariance functions of functional data. The random trajectories are, not necessarily differentiable, have unknown regularity, and are measured with error at randomly drawn discrete design points. The definition of our nonparametric estimators depends on the local regularity of the stochastic process generating the functional data. We first propose a simple estimator of this local regularity which takes strength from the replication and regularization features of functional data. Next, we use the “smoothing first, then estimate” approach for the estimation of the mean and the covariance functions. The new nonparametric estimators achieve optimal rates of convergence.

Keywords. Functional data, Hölder exponent, Local polynomials

1 Introduction

Un intérêt croissant est porté sur la modélisation de données enregistrées sous forme de séquence de mesures discrète à travers le temps. L’analyse de données fonctionnelles (ADF) considère ces données comme des réalisations d’un processus stochastique, enregistré avec erreur à des instants aléatoires. Les fonctions moyenne et covariance jouent un rôle crucial en ADF.

Les données consistent en des réalisations d'un processus stochastique de second ordre $X = (X_t : t \in [0, 1])$ ayant des trajectoires continues. Les fonctions moyenne et covariance sont définies, respectivement, par

$$\mu(t) = \mathbb{E}(X_t) \quad \text{et} \quad \Gamma(s, t) = \mathbb{E} \{ [X_s - \mu(s)][X_t - \mu(t)] \}, \quad s, t \in [0, 1].$$

Si des réalisations indépendantes $X^{(1)}, \dots, X^{(N)}$ de X étaient observées sans bruit et de façon continues, les estimateurs idéaux seraient donnés par

$$\tilde{\mu}(t) = \frac{1}{N} \sum_{i=1}^N X_t^{(i)} \quad \text{et} \quad \tilde{\Gamma}(s, t) = \frac{1}{N-1} \sum_{i=1}^N \{X_s^{(i)} - \tilde{\mu}(s)\} \{X_t^{(i)} - \tilde{\mu}(t)\}, \quad s, t \in [0, 1].$$

Dans le cas d'applications réelles, les courbes sont rarement observées sans erreur et jamais pour chaque $t \in [0, 1]$. Pour chaque $1 \leq i \leq N$, soit M_i un entier positif et soit $T_m^{(i)} \in [0, 1]$, $1 \leq m \leq M_i$, les temps d'échantillonnage aléatoires de la courbe $X^{(i)}$. Ainsi, nous observons les paires $(Y_m^{(i)}, T_m^{(i)}) \in \mathbb{R} \times [0, 1]$ avec $Y_m^{(i)}$ défini comme

$$Y_m^{(i)} = X^{(i)}(T_m^{(i)}) + \epsilon_m^{(i)}, \quad 1 \leq m \leq M_i, \quad 1 \leq i \leq N, \quad (1)$$

et les $\epsilon_m^{(i)}$ sont des réalisations indépendantes de ϵ , une variable aléatoire de moyenne nulle et de variance σ^2 .

L'idée est de calculer $\tilde{\mu}(\cdot)$ et $\tilde{\Gamma}(\cdot, \cdot)$ en utilisant des estimateurs non-paramétriques de $X_t^{(i)}$ et $X_s^{(i)} X_t^{(i)}$, comme celui par polynômes locaux. Cette approche est appelée "smoothing first, then estimate" [Hall *et al.* (2006), Zhang and Chen (2007)]. En général, une hypothèse de régularité est faite sur les trajectoires de X . Celles-ci doivent admettre des dérivées d'ordre 2 ou supérieure. Cependant, dans certaines applications, par exemple dans le domaine de l'énergie ou dans les applications médicales, la régularité semble bien plus faible et cette hypothèse peut sembler irréaliste. Nous proposons donc des estimateurs non-paramétriques des fonctions moyenne et covariance d'un processus de régularité inconnue atteignant des vitesses de convergence optimale pour le risque minimax, basé sur l'estimation de la régularité locale du processus.

2 Estimation locale de la régularité

Soit t_0 un point de $[0, 1]$ et soit \mathcal{O}_* un voisinage de t_0 . Pour $H_{t_0} \in (0, 1]$, nous faisons l'hypothèse que le processus X vérifie la condition suivante :

$$\mathbb{E}((X_u - X_v)^2) \asymp L_{t_0}^2 |v - u|^{2H_{t_0}}, \quad \text{avec } u \text{ et } v \in \mathcal{O}_*. \quad (2)$$

La quantité H_{t_0} correspond à ce qu'on appellera la *régularité locale* du processus X sur \mathcal{O}_* . Lorsque la régularité locale est constante ($H_{t_0} \equiv H$), elle est associée à une hypothèse

habituelle sur la vitesse de décroissance des valeurs propres ordonnées de l'opérateur de covariance du processus. En effet, pour une large classe des processus, en notant $\lambda_j, j \geq 1$, ces valeurs propres ordonnées, alors pour $\nu > 1$, $\lambda_j \sim j^{-\nu}$, nous avons $2H = \nu - 1$.

Nous présentons ensuite les principales idées conduisant à la construction de l'estimateur local de la régularité. Pour $u, v \in \mathcal{O}_*$, notons

$$\theta(u, v) = \mathbb{E} [(X_u - X_v)^2].$$

Considérons t_1 et t_2 tels que $[t_1, t_2] \subset \mathcal{O}_*$ et t_0 est le point central de $[t_1, t_2]$. On remarque facilement que

$$H_{t_0} \approx \frac{\log(\theta(t_1, t_2)) - \log(\theta(t_1, t_0))}{2 \log(2)} \quad \text{si } t_2 - t_1 \text{ est petit.}$$

Considérons un échantillon d'apprentissage de N courbes, générées suivant (1). Soit \tilde{X}_u , un estimateur non-paramétrique de X_u , pour $u \in \mathcal{O}_*$. Nous définissons un estimateur naturel de H_{t_0} par

$$\hat{H}_{t_0} = \frac{\log(\hat{\theta}(t_1, t_2)) - \log(\hat{\theta}(t_1, t_0))}{2 \log(2)}, \quad \text{où } \hat{\theta}(u, v) = \frac{1}{N} \sum_{i=1}^N (\tilde{X}_u^{(i)} - \tilde{X}_v^{(i)})^2.$$

En particulier, sous certaines conditions sur le processus X , sur son estimateur non-paramétrique \tilde{X} , et sur les vitesses de croissance de N et $\mathbf{m} := \mathbb{E}(M_i)$, nous montrons que, pour \mathbf{m} suffisamment grand,

$$\mathbb{P} \left(|\hat{H}_{t_0} - H_{t_0}| > \log^{-2}(\mathbf{m}) \right) \leq \exp(-\mathbf{m}).$$

Notre résultat non-asymptotique sur la régularité locale de X permet de construire les fenêtres optimales pour le lissage des fonctions moyenne et covariance qui permettent d'atteindre les vitesses minimax pour l'estimation de ces fonctions. Pour l'implémentation de la méthode, nous avons également besoin d'un estimateur de L_{t_0} qui peut se construire et étudier d'une manière similaire avec l'estimateur de H_{t_0} .

3 Estimation des fonctions moyenne et covariance

Nous expliquons maintenant comment sélectionner une fenêtre ayant une vitesse de convergence optimale pour le lissage par polynômes locaux des trajectoires, et ensuite construire les estimateurs des fonctions moyenne et covariance.

Pour chaque $1 \leq i \leq N$, nous utilisons une approche par polynômes locaux pour construire des estimateurs non-paramétrique $\hat{X}^{(i)}$ de $X^{(i)}$. Pour $t \in [0, 1]$, si la régularité

locale H_t est connue, alors en utilisant les mesures $(Y_m^{(i)}, T_m^{(i)})$, $1 \leq m \leq M_i$, d'une trajectoire $X^{(i)}$, nous considérons l'estimateur défini par

$$\widehat{X}_t^{(i)} = \widehat{X}_t^{(i)}(h_t) = \sum_{m=1}^{M_i} Y_m^{(i)} W_m^{(i)}(t), \quad 1 \leq i \leq N,$$

avec h_t une fenêtre dépendante de H_t , et $W_m^{(i)}(t)$ des poids dépendants de h_t .

Soit k_0 un entier. Notons $\mathbf{1}\{\cdot\}$ la fonction indicatrice. Pour $t \in [0, 1]$, définissons $\mathcal{W}(t; h) = \sum_{i=1}^N w_i(t; h)$ et $\mathcal{W}(s, t; h) = \sum_{i=1}^N w_i(s; h) w_i(t; h)$ tels que

$$w_i(t; h) = 1 \quad \text{si} \quad \sum_{m=1}^{M_i} \mathbf{1}\{|T_m^{(i)} - t| \leq h\} \geq k_0, \quad \text{et} \quad w_i(t; h) = 0 \quad \text{sinon.}$$

L'estimateur adaptatif de la fonction moyenne est donné par $\widehat{\mu}^*(t) = \widehat{\mu}(t; h_\mu)$ avec

$$\widehat{\mu}(t; h) = \mathcal{W}(t; h)^{-1} \sum_{i=1}^N w_i(t; h) X_t^{(i)}, \quad t \in [0, 1]. \quad (3)$$

L'estimateur $\widehat{\mu}(t; h)$ est un estimateur de la moyenne prenant en compte le fait qu'il puisse y avoir moins de k_0 observations entre $t - h_\mu$ et $t + h_\mu$ pour certaines trajectoires. Nous définissons la fenêtre optimale pour calculer $\widehat{\mu}^*(t)$ telle que la différence au carré entre $\widehat{\mu}(t; h)$ et $\widetilde{\mu}(t)$ soit minimisée. La fenêtre optimale est donc définie par

$$h_\mu(t) = \arg \min_{h>0} \mathcal{R}_\mu(t; h), \quad \text{avec} \quad \mathcal{R}_\mu(t; h) = q_1^2 h^{2\hat{H}_t} + \frac{q_2^2}{\mathcal{N}_\mu(t; h)} + q_3^2 \left[\frac{1}{\mathcal{W}(t; h)} - \frac{1}{N} \right],$$

et q_1 , q_2 et q_3 des constantes dépendantes de la variance du processus au point t , de la variance de ϵ , de la constante L_{t_0} de l'équation (2), du noyau utilisé et tel que

$$\mathcal{N}_\mu(t; h) = \left[\frac{1}{\mathcal{W}^2(t; h)} \sum_{i=1}^N \frac{w_i(t; h)}{\mathcal{N}_i(t; h)} \sum_{m=1}^{M_i} |W_m^{(i)}(t; h)| \right]^{-1} \quad \text{avec} \quad \mathcal{N}_i(t; h) = \frac{w_i(t; h)}{\max_{1 \leq m \leq M_i} |W_m^{(i)}(t; h)|}.$$

Nous montrons que $\mathcal{R}_\mu(t; h)$ est une borne de $\mathbb{E} [\{\widehat{\mu}(t; h) - \widetilde{\mu}(t)\}^2]$. La minimisation de $\mathcal{R}_\mu(t; h)$ peut se faire sur une grille de valeurs pour h . En particulier, sous certaines conditions sur la densité des points d'observations $T_m^{(i)}$, sur la vitesse de convergence de H_t et sur les vitesses de croissance de N et \mathbf{m} , nous montrons que, pour $t \in [0, 1]$,

$$\widehat{\mu}^*(t) - \widetilde{\mu}(t) = O_{\mathbb{P}} \left((N\mathbf{m})^{-\frac{H_t}{1+2H_t}} \right) \quad \text{et} \quad \widehat{\mu}^*(t) - \mu(t) = O_{\mathbb{P}} \left((N\mathbf{m})^{-\frac{H_t}{1+2H_t}} + N^{-1/2} \right).$$

En ce qui concerne la fonction de covariance, la fonction variance $\Gamma(s, s)$ induit une singularité lors de l'estimation de la covariance $\Gamma(\cdot, \cdot)$, nous distinguons donc les points sur

la diagonale de ceux en dehors [Zhang and Wang (2016)]. Pour les points hors-diagonale, de façon similaire à l'estimation de la moyenne, l'estimateur adaptatif de la fonction covariance est donné par

$$\widehat{\Gamma}^*(s, t) = \widehat{\Gamma}(s, t; h_\Gamma) \quad \text{avec} \quad \widehat{\Gamma}(s, t; h) = \widehat{\gamma}(s, t; h) - \widehat{\mu}^*(s)\widehat{\mu}^*(t), \quad (4)$$

où $\widehat{\mu}^*(s)$ et $\widehat{\mu}^*(t)$ sont définis suivant (3) avec les fenêtres correspondantes, et

$$\widehat{\gamma}(s, t; h) = \mathcal{W}(s, t; h)^{-1} \sum_{i=1}^N w_i(s; h)w_i(t; h)\widehat{X}_s^{(i)}\widehat{X}_t^{(i)}, \quad s, t \in [0, 1], s \neq t. \quad (5)$$

Nous définissons la fenêtre optimale pour calculer $\widehat{\gamma}(s, t; h)$ telle que la différence au carré entre $\widehat{\gamma}(s, t; h)$ et $\widetilde{\gamma}(s, t)$, l'estimateur de $\mathbb{E}(X_s^{(i)}X_t^{(i)})$, soit minimisée. La fenêtre optimale est donc définie par

$$h_\Gamma = h_\Gamma(s, t) = \arg \min_{h>0} \{\mathcal{R}_\Gamma(s|t; h) + \mathcal{R}_\Gamma(t|s; h)\}, \quad (6)$$

avec

$$\mathcal{R}_\Gamma(t|s; h) = q_1^2 h^{2\widehat{H}_t} + \frac{q_2^2}{\mathcal{N}_\Gamma(t|s; h)} + q_3^2 \left[\frac{1}{\mathcal{W}(s, t; h)} - \frac{1}{N} \right]. \quad (7)$$

et q_1 , q_2 et q_3 des constantes dépendantes de différents moments du processus, de la variance de ϵ , de la contante L_{t_0} de l'équation (2), du noyau utilisé et tel que

$$\mathcal{N}_\Gamma(t|s; h) = \left[\frac{1}{\mathcal{W}^2(s, t; h)} \sum_{i=1}^N \frac{w_i(s; h)w_i(t; h)}{\mathcal{N}_i(t|s; h)} \sum_{m=1}^{M_i} |W_m^{(i)}(t; h)| \right]^{-1}$$

avec $\mathcal{N}_i(t|s; h) = w_i(s; h)w_i(t; h) \min_{1 \leq m \leq M_i} |W_m^{(i)}(t; h)|^{-1}$. La définition de $\mathcal{R}_\Gamma(s|t; h)$ est la même que dans l'équation (7) avec s et t échangés. Nous montrons la fonction de h minimisée en (6) est une borne de $\mathbb{E} [\{\widehat{\gamma}(s, t; h) - \widetilde{\gamma}(s, t)\}^2]$. La minimisation de $\mathcal{R}_\Gamma(s|t; h) + \mathcal{R}_\Gamma(t|s; h)$ peut se faire sur une grille de valeurs pour h . En particulier, en notant $H(s, t) = \min\{H_s, H_t\}$ et $H(s, t) < 1$, sous certaines conditions sur la densité des points d'observations $T_m^{(i)}$, sur la vitesse de convergence de H_t , sur les vitesses de croissance de N et \mathbf{m} et sur les moments du processus X , nous montrons que, pour $s, t \in [0, 1], s \neq t$,

$$\widehat{\Gamma}^*(s, t) - \widetilde{\Gamma}(s, t) = O_{\mathbb{P}} \left((N\mathbf{m})^{-\frac{H(s,t)}{1+2H(s,t)}} \right)$$

et

$$\widehat{\Gamma}^*(s, t) - \Gamma(s, t) = O_{\mathbb{P}} \left((N\mathbf{m})^{-\frac{H(s,t)}{1+2H(s,t)}} + N^{-1/2} \right).$$

Nous proposons d'utiliser cet estimateur (5) seulement en dehors de la bande diagonale définie par $\{(s, t) : |s - t| \leq \mathfrak{d}\}$, pour $\mathfrak{d} > 0$.

Il nous reste à donner un estimateur de cette bande diagonale pour la fonction de covariance. Pour cela, nous proposons une règle basée sur les données pour le choix de \mathfrak{d} et ensuite donner un estimateur de $\mathbb{E}(X_s X_t)$ quand s et t sont sur la diagonale. Notons H la régularité locale au point $u = (s + t)/2$. Sous certaines hypothèses sur les moments de X_t et de $X_t - X_s$, nous montrons que

$$\int \int_{t-\mathfrak{d} \leq s \leq t} \left\{ \tilde{\Gamma}(u - \mathfrak{d}/2, u + \mathfrak{d}/2) - \tilde{\Gamma}(s, t) \right\}^2 ds dt = O_{\mathbb{P}}(\mathfrak{d}^{2H+1}).$$

Pour (s, t) dans la bande diagonale, nous pouvons prendre $\hat{\Gamma}(u - \mathfrak{d}/2, u + \mathfrak{d}/2)$ définie par l'équation (4) quand $s \leq t$ et utiliser la symétrie de la covariance pour $s > t$. Si \hat{H} est un estimateur de H , nous pouvons, par exemple, prendre

$$\mathfrak{d} = \left\{ N^{-2} \sum_{i=1}^N (1/M_i) \right\}^c \quad \text{avec} \quad c = \frac{2\hat{H} + 1/2}{(2\hat{H} + 1)^2}.$$

Nous avons étudié la qualité de ces estimateurs de la fonction moyenne et covariance par rapport à deux estimateurs classiques, des estimateurs de lissage par splines sur les fonctions moyenne et covariance empiriques [Cai et Yuan (2010, 2011)] et des estimateurs par polynômes locaux pondérés nécessitant l'ensemble des courbes [Zhang and Wang (2016)]. Nous obtenons de bons résultats au regard de l'erreur quadratique intégrée avec des temps de calculs raisonnables.

La méthode peut s'étendre aux cas où la régularité est supérieur à 1 en estimant la régularité des dérivées successives. Enfin, le cadre peut aussi s'élargir aux cas où la variance du bruit n'est plus supposée constante mais d'esperance contionnnelle constante, ce qui permet de considérer un bruit hétéroscédastique. Une version complète de l'article est disponible à l'adresse suivante : arxiv:2108.06507.

Bibliographie

- Cai, T. and Yuan, M. (2010) *Nonparametric covariance function estimation for functional and longitudinal data*. University of Pennsylvania and Georgia Institute of Technology.
- Cai, T. T. and Yuan, M. (2011) *Optimal estimation of the mean function based on discretely sampled functional data: Phase transition*. *Ann. Statist.*, 39, 2330–2355.
- Hall, P., Müller, H.-G. and Wang, J.-L. (2006) *Properties of principal component methods for functional and longitudinal data analysis*. *Ann. Statist.*, 34, 1493–1517.
- Zhang, J.-T. and Chen, J. (2007) *Statistical inferences for functional data*. *Ann. Statist.*, 35, 1052–1079.
- Zhang, X. and Wang, J.-L. (2016) *From sparse to dense functional data and beyond*. *Ann. Statist.*, 44, 2281–2321.