

MODÈLES DE RÉGRESSION ORDINAL CUMULATIF À COVARIABLES TEMPORELLES

Simón Weinberger¹ & Jairo Cugliari² & Aurélie Le Cain³

¹ *Laboratoire ERIC, EssilorLuxottica, France, weinbes@essilor.fr*

² *Laboratoire ERIC, France, Jairo.Cugliari@univ-lyon2.fr*

³ *EssilorLuxottica, France, lecaina@essilor.fr*

Résumé. Dans ce document, on s'intéresse à la modélisation d'une réponse ordinale en utilisant des extensions du modèle ordinal cumulatif multivarié, présenté dans le livre de Agresti, A. (2010). On porte notre attention sur l'utilisation de covariables qui dépendent du temps. Pour cela, on explore deux représentations temporelles de nature différente : la représentation fonctionnelle présentée dans le monographe de Ramsay, J. O. et Silverman, B. W. (2005) et la signature du signal, présenté dans les notes Chevyrev, I. et Kormilitzin, A. (2016) ou encore la thèse de Fermanian, A. (2021). La principale contribution de cet article est l'utilisation d'une pénalité fonctionnelle dans le modèle ordinal à covariables fonctionnelles présenté par Jacques, J. et Samardzić, S. (2022). On montre la pertinence des méthodes proposées en analysant deux jeux de données publiques.

Mots-clés. Données fonctionnelles, signature, régularisation.

Abstract. We focus on modeling an ordinal response using extensions of the multivariate cumulative ordinal model, as presented in Agresti, A.'s (2010) book. Our attention is drawn to the use of covariates that depend on time. To this end, we explore two temporally based representations of different natures : the functional representation outlined in Ramsay, J. O., and Silverman, B. W.'s (2005) monograph, and the signal signature, introduced in Chevyrev, I., and Kormilitzin, A.'s (2016) notes, as well as in Fermanian, A.'s (2021) thesis. The primary contribution of this article lies in the application of a functional penalty in the ordinal model with functional covariates, as presented by Jacques, J., and Samardzić, S. (2022). We demonstrate the relevance of the proposed methods by analyzing two public datasets.

Keywords. Functional data, signatures, regularization.

Ce document est structuré de la façon suivante : dans la section 1 on définit les différentes représentations temporelles, on formule des modèles de régression appropriés et finalement, on s'intéresse à l'estimation des paramètres du modèle ordinal à covariables fonctionnelles. Dans la section 2, on applique les méthodes présentées à deux jeux de données publiques.

1 Méthode

1.1 Représentation de données temporelles

On considère une réponse discrète Y qui peut prendre $K \in \mathbb{N}$ valeurs distinctes. On suppose qu'il existe une relation d'ordre total dans cet ensemble de valeurs, on peut alors supposer, sans perdre de généralité, que Y prend des valeurs dans $\llbracket 1, K \rrbracket$. On dit alors que Y est une réponse ordinaire.

On s'intéresse alors à modéliser la variable Y en utilisant une covariable $Z(\cdot)$, qui est un processus aléatoire dépendant du temps. Par le théorème de décomposition de Wold, il existe une fonction déterministe $X(\cdot)$ et un processus aléatoire centré, $\epsilon(\cdot)$, tel que :

$$Z(t) = X(t) + \epsilon(t) \quad ; \quad t \in [0, T].$$

Dans cet article, on explore l'utilisation de $X(\cdot)$ ou de $\epsilon(\cdot)$ comme variables explicatives pour la réponse Y . Ces deux variables dépendent du temps, mais sont de nature différente : la fonction $X(\cdot)$ décrit le niveau moyen de $Z(\cdot)$ et le processus $\epsilon(\cdot)$ décrit l'écart entre ce niveau moyen et les valeurs de $Z(\cdot)$. Deux approches, déjà existantes dans la littérature, sont d'intérêt : dans l'article de Jacques, J. et Samardzić, S. (2022), un modèle ordinal à covariables fonctionnelles est décrit et dans les notes de Chevyrev, I. et Kormilitzin, A. (2016), une méthode pour classifier des processus ARMA est proposée.

Dans le domaine des données fonctionnelles, il est usuel de représenter la fonction $X(\cdot)$ en utilisant un développement fini dans une base fonctionnelle. Concrètement, on considère que la fonction $X(\cdot)$ peut s'écrire de la façon suivante :

$$X(t) = \sum_{k=1}^K a_k \psi_k(t) = \mathbf{a}^T \cdot \boldsymbol{\psi}(t) \quad ; \quad t \in [0, T],$$

où $(\psi_k)_{k=1}^K$ est une base fonctionnelle connue (base de Fourier, splines, ondelettes ...), $\mathbf{a} = (a_1, \dots, a_K)^T$ est un vecteur de coefficients et $\boldsymbol{\psi}(t) = (\psi_1(t), \dots, \psi_K(t))$ est un vecteur de fonctions.

Comme on illustre dans la section 2, dans certaines situations, il y a de l'information utile pour la prédiction de la réponse Y dans le processus $\epsilon(\cdot)$. La représentation fonctionnelle ne prend pas en compte ce processus, il est souvent considéré comme du bruit, or, il se pourrait que ce bruit soit informatif.

On présente succinctement une méthode, proposée dans Chevyrev, I. et Kormilitzin, A. (2016), pour extraire de l'information de $\epsilon(\cdot)$. Cette technique consiste à transformer le processus univarié $\epsilon(\cdot)$ dans un processus bivarié $\tilde{\epsilon}(\cdot)$ et ensuite, on utilise la "signature" de ce processus, à valeurs dans \mathbb{R}^2 , pour extraire des caractéristiques utiles pour la régression ordinaire. Définissons la signature d'un processus aléatoire, $\tilde{\epsilon}(\cdot)$, à valeurs dans \mathbb{R}^d : $\tilde{\epsilon}(\cdot) = (\tilde{\epsilon}_1(\cdot), \tilde{\epsilon}_2(\cdot), \dots, \tilde{\epsilon}_d(\cdot))^T$.

Soit ω_k un multi-index, de taille $k \in \mathbb{N}$, $\omega_k = (i_1, i_2, \dots, i_k) \in \{1, \dots, d\}^k$, l'intégrale itérée

de $\tilde{\epsilon}$ dans l'intervalle $[a, b]$ par rapport à ω_k , est la quantité suivante :

$$S_{[a,b]}^{\omega_k}(\tilde{\epsilon}) = \int_{a < s_1 < \dots < s_k < b} \dots \int d\tilde{\epsilon}_{i_1} \dots d\tilde{\epsilon}_{i_k}.$$

La signature de $\tilde{\epsilon}$ de niveau $k \in \mathbb{N}$, est l'ensemble des intégrales itérées par rapport aux multi-index de longueur plus petite ou égale à k . On note cet ensemble $S_{[a,b]}^{(k)}$. La signature de $\tilde{\epsilon}$ est l'ensemble des intégrales itérées. On note cet ensemble-là par $S_{[a,b]}(\tilde{\epsilon})$. Dans la thèse de Fermanian, A. (2021), il a été proposé d'utiliser la signature (de niveau k fixe) pour extraire des caractéristiques dans des tâches d'apprentissage automatique où la covariable est un processus multivarié. Dans le cas univarié, il est conseillé de transformer le processus dans un processus multivarié en appliquant différentes transformations.

Par exemple, pour des données univariées $\mathcal{M} = \{x_1, \dots, x_n\}$, on considère les transformations suivantes : la somme cumulée, notée $CS(\cdot)$, le lead-lag, $LeadLag(\cdot)$ et le point de base $BP(\cdot)$, sont les transformations suivantes des données \mathcal{M} :

$$\begin{aligned} CS(\mathcal{M}) &= \left\{ x_1, \sum_{i=1}^2 x_i, \sum_{i=1}^3 x_i, \sum_{i=1}^3 x_i, \dots, \sum_{i=1}^n x_i \right\}, \\ LeadLag(\mathcal{M}) &= \left\{ \begin{pmatrix} x_1 \\ x_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_3 \\ x_2 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ x_{n-1} \end{pmatrix}, \begin{pmatrix} x_n \\ x_n \end{pmatrix} \right\}, \\ BP(\mathcal{M}) &= \{0, x_1, x_2, \dots, x_n\}. \end{aligned}$$

Dans la pratique, on n'observe pas la fonction $Z(\cdot)$, mais des données de la forme $\{(t_i, Z(t_i))\}_{i=1}^T$. Pour estimer la fonction $X(\cdot)$, on fixe une base fonctionnelle $(\psi_k(\cdot))_{k=1}^K$ donnée, et il faut déterminer les coefficients, $(a_k)_{k=1}^K$. Une approche possible, qui est décrite par Ramsay, J. O. et Silverman, B. W. (2005), est de choisir les coefficients qui minimisent la quantité suivante :

$$\sum_{i=1}^T \left(Z(t_i) - \hat{X}(t_i) \right)^2 + \lambda_X \int_{[0,T]} \left(\hat{X}^{(k)}(s) \right)^2 ds,$$

où $\hat{X}(t) = \sum_{k=1}^K a_k \psi_k(t)$ est l'estimation de la fonction $X(\cdot)$ et $\lambda_X > 0$ est un réel positif connu qui contrôle l'irrégularité de $\hat{X}(t)$.

Une fois que l'on a cette représentation fonctionnelle, on doit estimer le processus $\epsilon(\cdot)$, pour cela, on propose d'utiliser les observations initiales $\{(t_i, Z(t_i))\}_{i=1}^T$ et $\{(t_i, \hat{X}(t_i))\}_{i=1}^T$. On calcule $\hat{\epsilon}_i = Z(t_i) - \hat{X}(t_i)$, pour obtenir la série temporelle $\mathcal{E} = \{\epsilon_i\}_{i=1}^T$. Ensuite, on applique une transformation sur \mathcal{E} :

$$\tilde{\mathcal{E}} = LeadLag(CS(BP(\mathcal{E}))). \quad (1)$$

Notons qu'il y a $2T + 2$ éléments dans $\tilde{\mathcal{E}}$, et chacun de ces éléments est dans \mathbb{R}^2 .

Ensuite, on utilise une interpolation linéaire des éléments de $\tilde{\mathcal{E}}$ pour obtenir une fonction à valeurs dans \mathbb{R}^2 . Finalement, on calcule la signature de cette fonction linéaire par morceaux. Il existe une implémentation de tout ce processus dans la fonction *SignatureTransformer* de la bibliothèque Python *sktime*, celle-ci a été développé par Morrill, J., Fermanian, A., Kidger, P. et Lyons, T. J. (2020).

1.2 Modèle ordinal

Le modèle ordinal cumulatif classique, présenté dans le livre d'Agresti, A. (2010), permet d'étudier le lien entre une réponse ordinaire $Y \in \llbracket 1, K \rrbracket$ et une covariable $W \in \mathbb{R}^p$. Ce modèle s'écrit de la façon suivante :

$$\text{logit}(\mathbb{P}(Y \leq j)) = \tau_j - \langle \mathbf{b}, \mathbf{w} \rangle_{\mathbb{R}^p} \quad ; \quad \llbracket 1, K-1 \rrbracket, \quad (2)$$

où $-\infty = \tau_0 < \tau_1 < \dots < \tau_{K-1} < \tau_K = \infty$. Les paramètres de ce modèle sont les seuils ordonnés $(\tau_j)_{j=1}^{K-1}$ et le vecteur de paramètres $b \in \mathbb{R}^p$. Pour simplifier les notations, on désigne Δ_{K-1} l'ensemble des seuils ordonnés :

$$\Delta_{K-1} = \{(x_1, x_2, \dots, x_{K-1}) \in \mathbb{R}^{K-1} \mid x_1 < x_2 < \dots < x_{K-1}\}.$$

Pour un échantillon \mathcal{D} , avec n observations, $\mathcal{D} = \{(W_1, Y_1), \dots, (W_n, Y_n)\}$, l'équation (2), définit une vraisemblance $\mathcal{L}_{\mathbb{R}^p}(\cdot | \mathcal{D})$ sur l'espace des paramètres du modèle : $\mathbb{R}^p \times \Delta_{K-1}$. Ces paramètres peuvent s'estimer par maximum de vraisemblance.

Jacques, J. et Samardzić, S. (2022) ont proposé une extension de ce modèle pour utiliser une donnée fonctionnelle en tant que variable explicative. Ils définissent le modèle suivant :

$$\text{logit}(\mathbb{P}(Y \leq j)) = \tau_j - \langle \beta, X \rangle_{L^2([0, T])} \quad ; \quad \llbracket 1, K-1 \rrbracket. \quad (3)$$

Notons qu'on a remplacé le produit scalaire dans \mathbb{R}^p par le produit scalaire dans $L^2([0, T])$.

Pour un échantillon $\tilde{\mathcal{D}}$, avec n observations, $\tilde{\mathcal{D}} = \{(X_1(\cdot), Y_1), \dots, (X_n(\cdot), Y_n)\}$, l'équation (3), définit une vraisemblance $\mathcal{L}_{L^2([0, T])}(\cdot | \tilde{\mathcal{D}})$ sur l'espace des paramètres du modèle : $L^2([0, T]) \times \Delta_{K-1}$. Afin d'estimer les paramètres, une hypothèse supplémentaire est faite sur la forme de la fonction $\beta(\cdot)$, on suppose qu'elle peut s'écrire de la façon suivante :

$$\beta(t) = \sum_{j=1}^P b_j \phi_j(t) = \mathbf{b}^T \cdot \boldsymbol{\phi}(t),$$

où $(\phi_k)_{k=1}^P$ est une base fonctionnelle connue, $\mathbf{a} = (b_1, \dots, b_P)^T$ est un vecteur de coefficients et $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_P(t))$ est un vecteur de fonctions. Sous cette hypothèse, on peut calculer un estimateur, $\hat{\theta}_{ML}$, des paramètres du modèle par maximum de vraisemblance. Alternative-ment, on peut calculer un estimateur, $\tilde{\theta}_{RP}$, en maximisant la vraisemblance et contrôlant la rugosité de la fonction $\hat{\beta}$ estimée. Concrètement, on définit :

$$\hat{\theta}_{ML} = \arg \max_{\theta \in L^2([0, T]) \times \Delta_{K-1}} \mathcal{L}_{L^2([0, T])}(\theta | \tilde{\mathcal{D}}), \quad (4)$$

$$\tilde{\theta}_{RP} = \arg \max_{\theta \in L^2([0, T]) \times \Delta_{K-1}} \mathcal{L}_{L^2([0, T])}(\theta | \tilde{\mathcal{D}}) + \lambda \int_{[0, T]} (\beta^{(k)}(s))^2 ds, \quad (5)$$

où $\beta^{(k)}(\cdot)$ est la dérivée d'ordre k de $\beta(\cdot)$ et $\lambda > 0$ est un hyperparamètre connu qui contrôle la pénalisation.

Dans les deux cas, on montre dans la section suivante que ces deux problèmes d'optimisation sur un espace fonctionnel, peuvent se réécrire comme des problèmes d'optimisation dans \mathbb{R}^P .

1.3 Estimation des paramètres du modèle fonctionnel

Notons que l'on peut réécrire le produit scalaire fonctionnel comme un produit vectoriel en \mathbb{R}^P :

$$\langle X, \beta \rangle_{L^2([0,T])} = \sum_{i=1}^K \sum_{j=1}^P a_i \left(\int_{[0,T]} \psi_i(s) \phi_j(s) ds \right) b_j = \mathbf{a}' \cdot \mathbf{R} \cdot \mathbf{b} = \langle \tilde{\mathbf{a}}, \mathbf{b} \rangle_{\mathbb{R}^P},$$

où $\mathbf{a} = (a_1, a_2, \dots, a_K)^T$, $\mathbf{b} = (b_1, b_2, \dots, b_P)^T$, $\mathbf{R}_{i,j} = \int_{[0,T]} \psi_i(s) \phi_j(s) ds$ et $\tilde{\mathbf{a}} = \mathbf{a}^T \cdot \mathbf{R}$.

En utilisant cette remarque, on peut calculer l'estimateur $\hat{\theta}_{ML}$, en estimant les paramètres d'un modèle ordinal classique. En effet, supposons que l'on observe un échantillon avec des données fonctionnelles : $\tilde{\mathcal{D}} = \{(X_i(\cdot), Y_i)\}_{i=1}^n$ et que chaque $X_i(\cdot)$ s'écrit $X_i(\cdot) = \mathbf{a}_i^T \cdot \boldsymbol{\psi}(\cdot)$. Considérons la vraisemblance d'un modèle ordinal à covariables dans \mathbb{R}^P sur l'échantillon suivant : $\mathcal{D} = \{\mathbf{a}_i^T \cdot \mathbf{R}, Y_i\}_{i=1}^n$. Alors, on a l'égalité suivante :

$$\hat{\theta}_{ML} = \arg \max_{\theta \in L^2([0,T]) \times \Delta_{K-1}} \mathcal{L}_{L^2([0,T])}(\theta | \tilde{\mathcal{D}}) = \arg \max_{\theta \in \mathbb{R}^P \times \Delta_{K-1}} \mathcal{L}_{\mathbb{R}^P}(\theta | \mathcal{D}).$$

Similairement, on peut estimer $\hat{\theta}_{RP}$ en estimant les paramètres d'un modèle ordinal, à covariables dans \mathbb{R}^P , avec une pénalisation ridge. On considère la matrice $P \times P$, $\mathbf{R}^{(k)}$, où $\mathbf{R}_{i,j}^{(k)} = \int_{[0,T]} \psi_i^{(k)}(s) \psi_j^{(k)}(s) ds$. Notons que :

$$\int_{[0,T]} (\beta^{(k)}(s))^2 ds = \sum_{i=1}^P \sum_{j=1}^P \mathbf{b}_i \left(\int_{[0,T]} \psi_i^{(k)}(s) \psi_j^{(k)}(s) ds \right) \mathbf{b}_j = \mathbf{b}' \mathbf{R}^{(k)} \mathbf{b}.$$

Cette matrice est symétrique positive, donc elle est diagonalisable dans une base orthonormée, on peut l'écrire de la façon suivante : $\mathbf{R}^{(k)} = \mathbf{P} \mathbf{D} \mathbf{P}'$, où \mathbf{P} est une matrice orthonormale et \mathbf{D} est une matrice diagonale avec les valeurs propres, $\lambda_1, \dots, \lambda_P$, de $\mathbf{R}^{(k)}$. Si on pose, $\mathbf{B} = \mathbf{P}^T \cdot \mathbf{b}$ et $\tilde{\mathbf{x}} = \mathbf{P}^T \mathbf{R}^T \mathbf{a}$ alors, on a :

$$\int_{[0,T]} (\beta^{(k)}(s))^2 ds = \mathbf{b}' \mathbf{R}^{(k)} \mathbf{b} = (\mathbf{P}^T \mathbf{b})^T \mathbf{D} (\mathbf{P}^T \mathbf{b}) = \mathbf{B}^T \mathbf{D} \mathbf{B} = \sum_{i=1}^P \lambda_i \mathbf{B}_i^2,$$

et

$$\langle X, \beta \rangle_{L^2([0,T])} = \mathbf{a}^T \mathbf{R} \mathbf{b} = \mathbf{a}^T \mathbf{R} \mathbf{P} \mathbf{B} = \langle \tilde{\mathbf{x}}, \mathbf{B} \rangle_{\mathbb{R}^P}.$$

Ainsi, si on considère l'échantillon $\tilde{\mathcal{D}} = \{\mathbf{P}^T \cdot \mathbf{R}^T \mathbf{a}_i, Y_i\}_{i=1}^n$, on a :

$$\mathcal{L}_{L^2([0,T])}(\theta | \tilde{\mathcal{D}}) + \lambda \int_{[0,T]} (\beta^{(k)}(s))^2 ds = \mathcal{L}_{\mathbb{R}^P}(\theta | \tilde{\mathcal{D}}) + \lambda \sum_{i=1}^P \lambda_i \mathbf{B}_i^2.$$

On reconnaît un modèle ordinal à covariables dans \mathbb{R}^P avec une pénalité Ridge, pondérée par $\lambda_1, \dots, \lambda_P$. Wurm, M. J., Rathouz, P. J., et Hanlon, B. M. (2021) ont développé des méthodes permettant d'estimer le modèle ordinal à covariables multivariés, avec une pénalisation ridge. Ainsi, on peut calculer $\hat{\theta}_{RP}$ avec ces méthodes.

2 Expériences numériques

2.1 Données Cookies

Afin d'illustrer la régularisation du modèle ordinal à covariables fonctionnelles, on utilise le jeu de données « Cookies ». Celui-ci a été utilisé dans les articles de Costanzo, G., Preda, C. et Saporta, G. (2006) et de Jacques, J. et Samardzić, S. (2022). Ce jeu de données provient du Danone Vitapole Paris Research Center et contient des mesures de résistance lors du pétrissage de 115 pâtes à cookies. Toutes les deux secondes, on mesure la résistance, cela pendant huit minutes, et on cuit chaque pâte pour obtenir des cookies. Ensuite, on va mesurer la qualité du cookie et voir s'il est « bon », « ajustable » ou « mauvais ». On s'intéresse au lien entre la résistance de la pâte lors du pétrissage et la qualité finale du cookie. On est clairement dans le cadre décrit dans cet article : la covariable est de nature fonctionnelle et la variable cible est de nature ordinale : mauvais < ajustable < bon.

On utilise une base de 22 splines cubiques pour représenter la résistance en tant que donnée fonctionnelle, on utilise une base de 12 splines d'ordre 5 pour représenter la fonction $\beta(\cdot)$ et on pénalise la troisième dérivée de ce paramètre fonctionnel. Ces valeurs sont des hyperparamètres et ils sont arbitrairement choisis dans ce document. Dans la pratique, ils pourraient être sélectionnés par validation croisée.

La figure 1 permet de visualiser la résistance lors du pétrissage, selon la qualité du cookie obtenue. On s'aperçoit déjà que les bons cookies semblent avoir une résistance plus importante que les mauvais cookies, surtout vers la fin du processus. Mais ce n'est pas évident de faire cela de façon précise, une des raisons d'utiliser un modèle de régression est de quantifier cette intuition avec un modèle statistique.

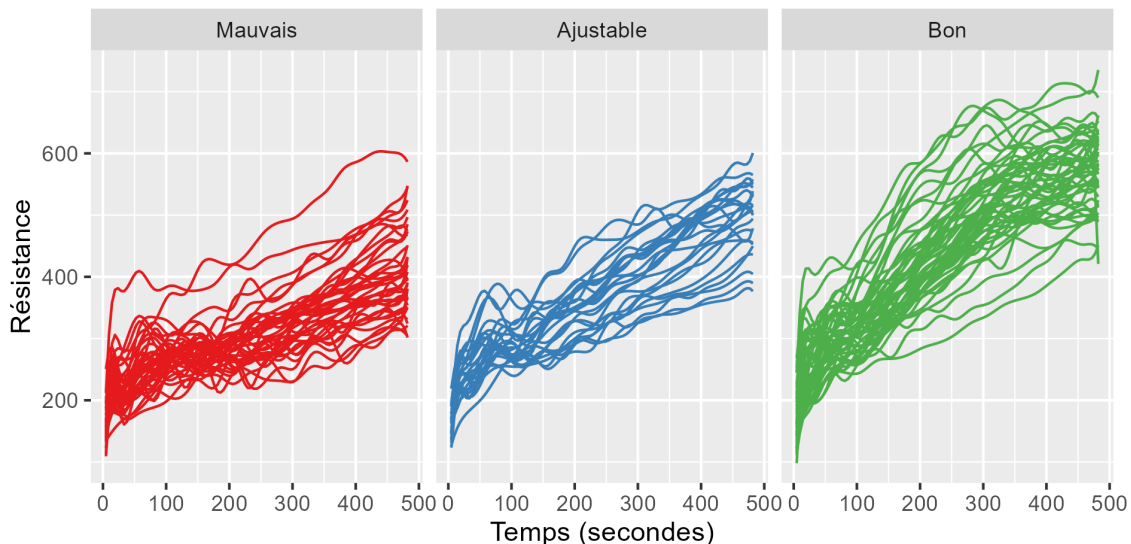


FIGURE 1 – Donnée fonctionnelle : résistance lors du pétrissage, selon qualité du cookie

Un problème pratique qui n'a pas été adressé dans la section précédente est la sélection de l'hyperparamètre λ . Une approche possible pour déterminer cet hyperparamètre est d'utiliser

une validation croisée. L'irrégularité de l'estimation de la fonction, $\beta(\cdot)$ est un signe de sur-apprentissage, on contrôle cela en choisissant le λ qui maximise la vraisemblance moyenne avec une validation croisée (voir la figure 2). Notons que la vraisemblance semble atteindre un maximum dans une zone (entre $\exp(-5)$ et $\exp(1)$), les estimations que l'on obtient si on pénalise trop ou pas assez donnent une vraisemblance inférieure. Notons aussi que c'est avec ce paramètre que l'on obtient le taux d'erreur de classification le plus faible.

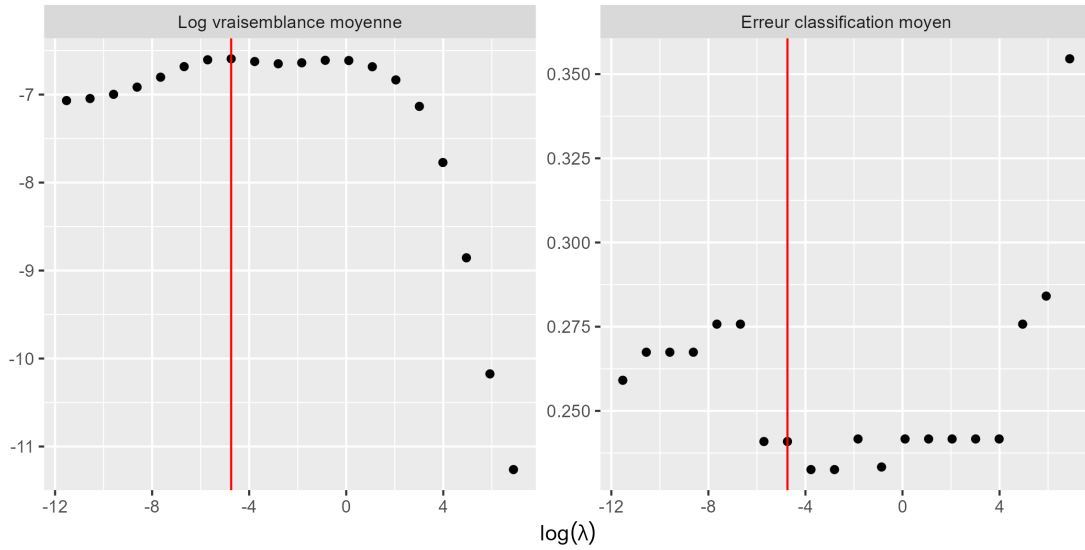


FIGURE 2 – Sélection du paramètre de l'hyperparamètre λ par validation croisée ($k=10$ blocs)

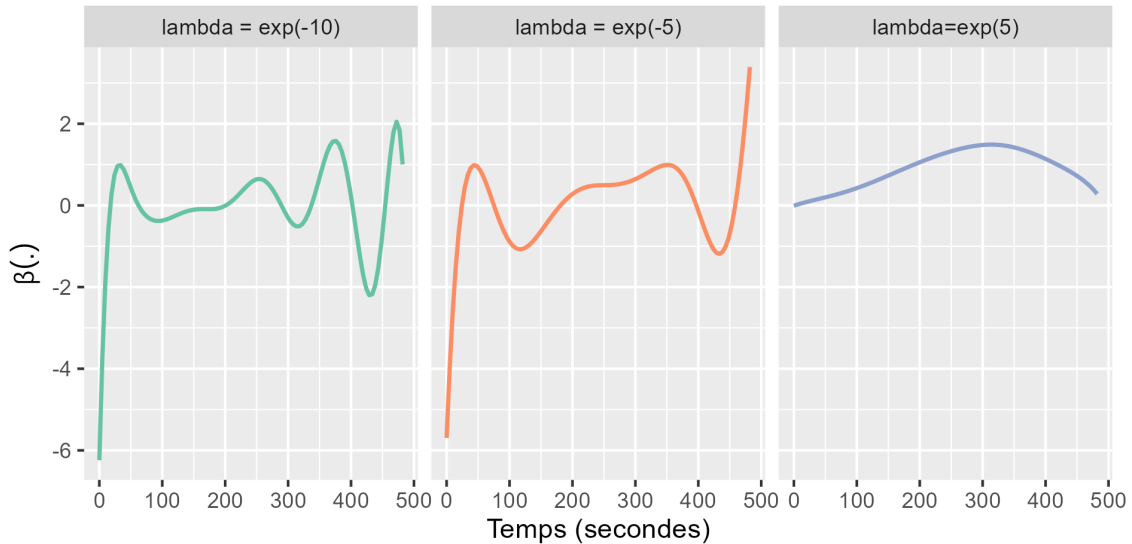


FIGURE 3 – Différentes estimations de la fonction β , selon les valeurs de λ

Finalement, notons que la forme des estimations de la fonction β n'est pas la même selon le degré de régularisation (voir la figure 3). Avec un paramètre λ trop élevé, on semble perdre de l'information temporelle, si on ne pénalise pas assez, la fonction β estimée est difficile

à interpréter dû aux fortes oscillations (surtout entre 300 et 500 secondes). En observant l'estimation obtenue avec $\lambda = \exp(-5)$, on dirait que pour avoir un cookie de bonne qualité, il faut avoir une pâte avec une faible résistance au début et une résistance élevée au milieu et surtout à la fin du pétrissage.

2.2 Données BabyECG

On s'intéresse au jeu de données « BabyECG ». Ce sont des observations de l'état de sommeil d'un bébé et de son rythme cardiaque. Une mesure est enregistrée toutes les 16 secondes et cela est fait entre 21h18 et 6h27. Cette série temporelle a été présentée par Nason, G. P., Von Sachs, R. et Kroisandt, G. (2000) comme un exemple de "locally stationary wavelet process" dont le "evolutionary wavelet spectrum" du rythme cardiaque était lié à l'état de sommeil de l'enfant. Dans la figure 4 (reproduite de l'article de Nason, G. P., Von Sachs, R. et Kroisandt, G. (2000)), on peut apercevoir ce lien.

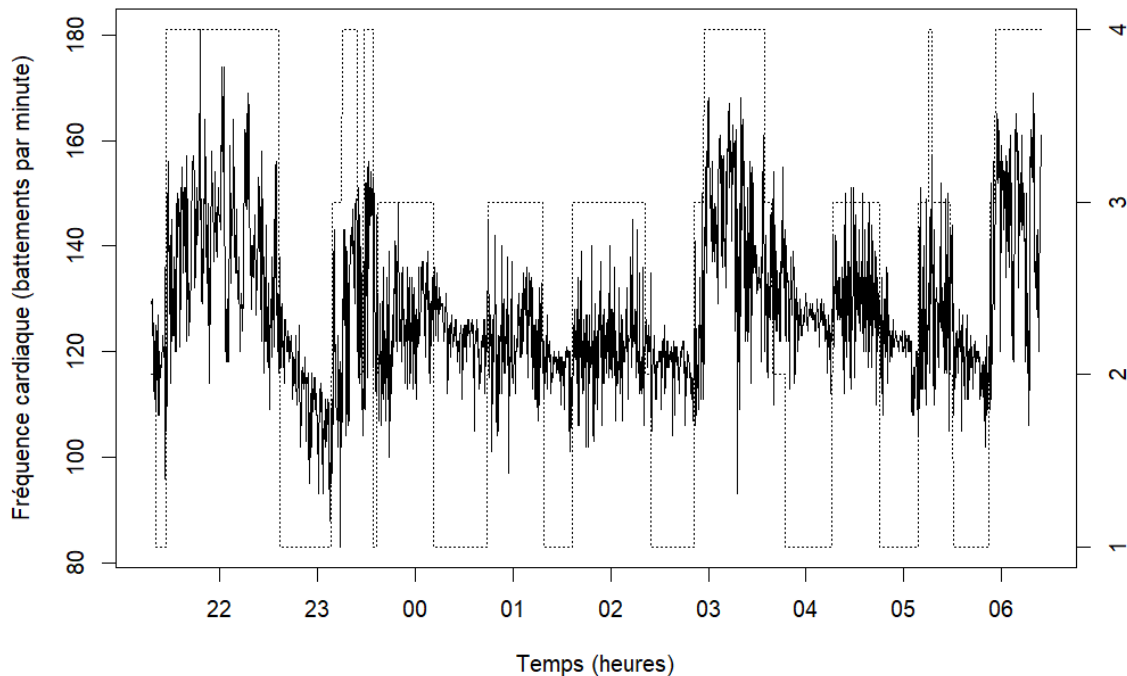


FIGURE 4 – Données BabyECG : rythme cardiaque dans le temps (ligne solide) et type de sommeil (ligne en pointillé) : les niveaux 1, 2, 3 et 4 correspondent à "réveillé", "sommeil léger", "sommeil moyen" et "sommeil profond" respectivement.

On étudie le lien entre le rythme cardiaque et le type de sommeil en utilisant la méthode de la signature présentée dans la section précédente. Pour cela, on enlève la tendance de la série entière, on divise la série obtenue dans 128 sous-séries avec 16 observations (correspondant

à 256 secondes de mesures). Pour chaque sous-série, on cherche à prédire l'état de sommeil final. On peut visualiser les sous séries obtenues dans la figure 5A.

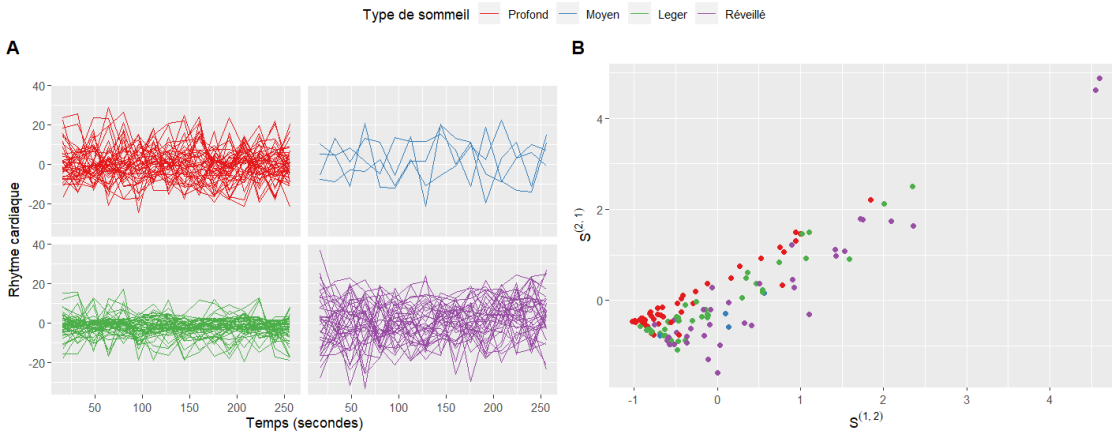


FIGURE 5 – **A** : sous-séries du rythme cardiaque obtenues, selon le type de sommeil. **B** : paire de signatures signifiantes, sélectionnées avec une pénalisation Lasso

Ensuite, pour chaque sous série de résidus, on calcule la transformation (1) et on calcule la signature de niveau deux de cette sous série bivarié. On obtient ainsi des coefficients $(S^{(1)}, S^{(2)}, S^{(1,1)}, S^{(1,2)}, S^{(2,1)}, S^{(2,2)})$, que l'on utilise comme caractéristiques pour une régression ordinaire avec une pénalité lasso.

En utilisant une validation croisée avec cinq blocs, on obtient un taux d'erreur de classification de 31,3%. Les résidus de cette série apportent bien une information sur le type de sommeil et une méthode d'utiliser cette information est d'utiliser la méthode de signature. Dans la figure 5B, on visualise deux coefficients, $S^{(1,2)}$ et $S^{(2,1)}$, sélectionnés par Lasso. On observe que les points sont, plus ou moins, séparables par des droites parallèles. Et, sauf pour le sommeil "Moyen", les types de sommeil semblent ordonnés : on passe d'une catégorie à la catégorie adjacente.

Conclusion

On a proposé des méthodes pour utiliser de l'information temporelle. Utiliser le modèle ordinal à covariables fonctionnelles avec la régularisation présentée, semble améliorer la qualité des prédictions obtenues et ça facilite l'interprétation du modèle. Utiliser la méthode basée dans des signatures permet d'extraire de l'information qui serait perdue avec un modèle à covariables fonctionnelles.

Il est de plus en plus usuel d'avoir des objets équipés de capteurs. De tels objets permettent d'obtenir des mesures successives qui peuvent être traitées comme des variables qui dépendent du temps. Les approches proposées dans cet article permettraient de se servir de telles variables.

Bibliographie

Agresti, A. (2010), *Analysis of Ordinal Categorical Data*, Wiley Series in Probability and Statistics.

Chevyrev, I. et Kormilitzin, A. (2016), A primer on the signature method in machine learning, *arXiv preprint arXiv :1603.03788*.

Costanzo, G., Preda, C. et Saporta, G. (2006), Anticipated prediction in discriminant analysis on functional data for binary response, *COMPSTAT'06, 17th Symposium on Computational Statistics*.

Fermanian, A. (2021), *Learning time-dependent data with the signature transform*, Thèse Université Sorbonne.

Jacques, J. et Samardzić, S. (2022), Analyzing cycling sensors data through ordinal logistic regression with functional covariates, *Journal of the Royal Statistical Society : Series C Applied Statistics*.

Morril, J., Fermanian, A., Kidger, P. et Lyons, T. J. (2020), A generalised signature method for time series, *arXiv preprint arXiv :2006.00873*.

Nason, G. P., Von Sachs, R. et Kroisandt, G. (2000), Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62, pp. 271–292.

Ramsay, J. O. et Silverman, B. W. (2005), *Functional data analysis*, Springer series in statistics.

Wurm, M. J., Rathouz, P. J., et Hanlon, B. M. (2021), Regularized ordinal regression and the ordinalnet r package, *Journal of Statistical Software*, 99, pp. 1-42.