

Estimation et sélection de variables dans les modèles mixtes fonctionnels

Madison Giacomci
Katholieke Universiteit Leuven

Sophie Lambert-Lacroix (TIMC - Grenoble)
Franck Picard (LBBE - Lyon)

Workshop Phon&Stat
Grenoble - 27 Novembre 2014

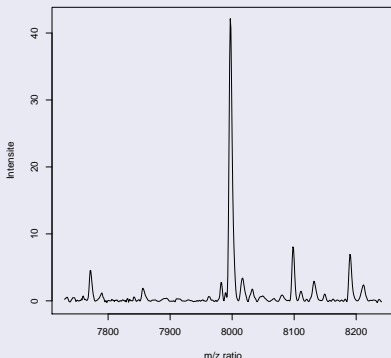
- 1 Contexte
- 2 Modélisation non paramétrique
- 3 Estimation dans les modèles mixtes fonctionnels
 - Problématiques abordées
 - Approche marginale
 - Approche jointe
- 4 Applications
 - Simulations
 - Données réelles
- 5 Perspectives

- 1 Contexte
- 2 Modélisation non paramétrique
- 3 Estimation dans les modèles mixtes fonctionnels
 - Problématiques abordées
 - Approche marginale
 - Approche jointe
- 4 Applications
 - Simulations
 - Données réelles
- 5 Perspectives

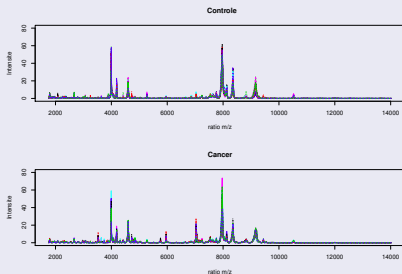
Données de spectrométrie de masse

Données liées à l'étude du protéome permettant d'obtenir une image plus complète des processus biologiques impliqués dans le développement d'une pathologie

- Les spectres individuels sont caractérisés par la présence de pics
- Chaque abscisse caractérise un unique peptide par son ratio m/z
- La hauteur d'un pic est associée à la quantité d'un peptide au sein de l'échantillon
- Les signaux produits contiennent entre 10^3 et 10^6 points de mesures



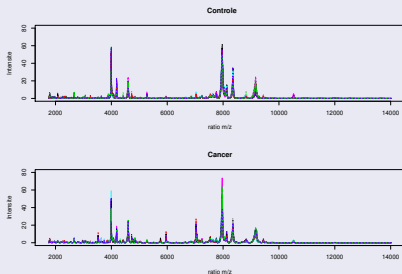
- Composé de 253 spectres sous 2 conditions
 - Individus sains (91)
 - Atteints d'un cancer (162)
- Chaque spectre est composé de la mesure de l'intensité de 15154 peptides



Parmi les enjeux statistiques

- Découverte de marqueurs moléculaires ▷ Analyse discriminante
- Découverte de sous-groupes à partir des données moléculaires ▷ Classification non supervisée
- Reconstruction des profils protéomiques moyens ▷ Estimation

- Composé de 253 spectres sous 2 conditions
 - Individus sains (91)
 - Atteints d'un cancer (162)
- Chaque spectre est composé de la mesure de l'intensité de 15154 peptides



Parmi les enjeux statistiques

- Découverte de marqueurs moléculaires ▷ Analyse discriminante
- Découverte de sous-groupes à partir des données moléculaires ▷ Classification non supervisée
- **Reconstruction des profils protéomiques moyens** ▷ Estimation

- Chaque individu est représenté par un spectre de données mesurées à haut-débit, de manière régulière

▷ Cadre de modélisation fonctionnelle

- Chaque spectre individuel est caractérisé par la présence de discontinuités qui sont informatives

▷ Utilisation des bases d'ondelettes

- Présence d'une importante variabilité biologique

▷ Contexte des modèles mixtes

- 1 Contexte
- 2 Modélisation non paramétrique
- 3 Estimation dans les modèles mixtes fonctionnels
 - Problématiques abordées
 - Approche marginale
 - Approche jointe
- 4 Applications
 - Simulations
 - Données réelles
- 5 Perspectives

Approche fonctionnelle (Ramsay et Silverman, 1997)

On parle de **données fonctionnelles** lorsque :

- ▷ les observations sont mesurées sur une grille fine et régulière
- ▷ l'unité d'observation idéale est la courbe

Modèle de régression fonctionnelle

Les données sont vues comme N courbes observées de manière bruitée sur une grille de discrétisation (t_1, \dots, t_M) telle que pour $i = 1, \dots, N$ et pour $m = 1, \dots, M$:

$$Y_i(t_m) = \mu(t_m) + E_i(t_m), \quad E_i(t_m) \sim \mathcal{N}(0, \sigma_E^2)$$

⇒ Objectif : Retrouver la fonction μ à partir de l'observation de signaux bruités.

Modèle mixte fonctionnel

Modélisation de la présence d'une variabilité spécifique due à l'individu par l'introduction d'effets aléatoires fonctionnels dans le modèle

$$Y_i(t_m) = \mu(t_m) + U_i(t_m) + E_i(t_m)$$

où $U_i \sim \mathcal{N}(0, K(t, t'))$ est un processus Gaussien centré, indépendant de E_i .

Contexte non-paramétrique

On ne spécifie pas de forme particulière pour les fonctions du modèle

▷ Le problème se place en dimension infinie

Approche usuelle

Projeter le modèle sur une base fonctionnelle bien choisie

Exemples : Bases de splines, polynomiales, de Fourier, d'ondelettes

Avantages des ondelettes

- ▷ Modélisation de courbes possédant des irrégularités
- ▷ Représentation parcimonieuse des signaux réguliers
- ▷ Propriétés décorrélantes
- ▷ Algorithme de décomposition efficace
- ▷ Contrôle fin de la régularité

Ondelettes : quelques brefs rappels

Construction des bases d'ondelettes

Construction de bases orthonormées de $L^2(\mathbb{R})$ obtenues en dilatant et translatant une fonction d'échelle ϕ et une ondelette mère ψ :

$$\{\phi_{j_0 k}(t), k = 0, \dots, 2^{j_0} - 1; \psi_{jk}(t), j \geq j_0, k = 0, \dots, 2^j - 1\}$$

avec $\psi_{jk}(t) = 2^{\frac{j}{2}} \psi(2^j t - k)$ et $\phi_{jk}(t) = 2^{\frac{j}{2}} \phi(2^j t - k)$

Projection dans la base d'ondelettes

Toute fonction $f \in L^2(\mathbb{R})$ peut être décomposée dans la base d'ondelettes :

$$f(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0 k}^* \phi_{j_0 k}(t) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} d_{jk}^* \psi_{jk}(t)$$

où $c_{j_0 k}^* = \langle f, \phi_{j_0 k} \rangle$ et $d_{jk}^* = \langle f, \psi_{jk} \rangle$ sont, respectivement, les **coefficients d'échelle et d'ondelettes théoriques**.

Transformée en ondelettes discrètes

En présence de signaux discrets $\mathbf{Y} = (Y(t_1), \dots, Y(t_M))$, un outil populaire est donné par la transformée en ondelettes discrètes (DWT) (Mallat, 1992) :

$$\underset{[M \times M]}{\mathbf{W}} \underset{[M \times 1]}{\mathbf{Y}} = \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}$$

avec \mathbf{W} matrice de filtres dépendant de la base d'ondelettes

(\mathbf{c}, \mathbf{d}) sont les coefficients d'échelle et d'ondelettes empiriques de la décomposition :

$$\begin{aligned} \mathbf{c} &\simeq \sqrt{M} \times \mathbf{c}^* \\ \mathbf{d} &\simeq \sqrt{M} \times \mathbf{d}^* \end{aligned}$$

$$\begin{aligned} W Y_i(t_m) &= W \mu(t_m) + W U_i(t_m) + W E_i(t_m) \\ \Leftrightarrow \begin{bmatrix} \mathbf{c}_i \\ \mathbf{d}_i \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\nu}_i \\ \boldsymbol{\theta}_i \end{bmatrix} + \boldsymbol{\varepsilon}_i \end{aligned}$$

$$\text{où } \begin{cases} \boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{I}) \\ \begin{bmatrix} \boldsymbol{\nu}_i \\ \boldsymbol{\theta}_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \mathbf{G}_\nu & 0 \\ 0 & \mathbf{G}_\theta \end{bmatrix}\right) \end{cases} \quad \text{et } \boldsymbol{\varepsilon}^i \perp (\boldsymbol{\nu}_i^T, \boldsymbol{\theta}_i^T)^T$$

Dans le domaine des ondelettes, le modèle mixte fonctionnel se ramène à un modèle linéaire mixte

Motivations

On souhaite une modélisation de la variabilité des effets aléatoires permettant d'obtenir :

- ▶ une structure "simple" tout en assurant une flexibilité suffisante
- ▶ des effets fixes et aléatoires partageant la même régularité (Antoniadis & Sapatinas, 2007)

Idée naturelle

Proposer un modèle pour la fonction $K(s, t)$ et en déduire des conditions sur la matrice \mathbf{G} , mais cela entraîne en général des difficultés pour :

- ▶ le contrôle du nombre de paramètres du modèle
- ▶ le contrôle de la régularité des trajectoires

On préfère donner une modélisation de la matrice \mathbf{G} dans le domaine des ondelettes suivant l'idée d'Antoniadis & Sapatinas (2007)

Hypothèses considérées

Structure de la matrice

G est supposée diagonale

▷ justifiée par la propriété décorrélante des ondelettes (Frazier et al, 1992)

Décroissance des paramètres de variances

Les termes diagonaux de **G** ont une décroissance exponentielle vis-à-vis de l'échelle j , tels que :

$$[\mathbf{G}_\theta]_{jk} = 2^{-j\eta} \gamma_{jk}^2$$

▷ Permet d'assurer que les effets fixes et aléatoires seront dans le même espace fonctionnel

Théorème - Abramovich et al (1998)

Supposons que $\mu(t) \in B_{\rho_1, \rho_2}^s$ et $\mathbb{V}(\theta_{jk}^i) = \gamma_{jk}^2 2^{-j\eta}$ alors

$$U_i(t) \in B_{\rho_1, \rho_2}^s \Leftrightarrow \begin{cases} \eta = 2s + 1 \text{ pour } 1 \leq \rho_1 \leq \infty \text{ et } \rho_2 = \infty \\ \eta > 2s + 1 \text{ sinon} \end{cases}$$

- 1 Contexte
- 2 Modélisation non paramétrique
- 3 Estimation dans les modèles mixtes fonctionnels**
 - Problématiques abordées
 - Approche marginale
 - Approche jointe
- 4 Applications
 - Simulations
 - Données réelles
- 5 Perspectives

Problématique générale

Partant de cette modélisation, on s'intéresse à la problématique de l'estimation au sein d'un groupe homogène d'individus

Estimation dans les modèles mixtes

- ▶ Estimation des effets fixes : objectif premier du statisticien car traduit le comportement moyen au sein du groupe
- ▶ Estimation des effets aléatoires : pour permettre une meilleure compréhension des sources de variabilité et une meilleure estimation des effets fixes

Estimation de l'effet fixe

- Se ramène à un problème de régression fonctionnelle basé sur les ondelettes
- La modélisation adoptée nous conduit à attendre une **représentation parcimonieuse de l'effet fixe**

Estimation des effets aléatoires

- Aussi modélisés comme des déviations individuelles fonctionnelles (partageant la même régularité que les effets fixes)
- Les résultats obtenus dans Giacomfi et al (2013) suggèrent une **représentation parcimonieuse des effets aléatoires**
 - ⇒ La sélection des effets aléatoires est réalisée par l'intermédiaire de la sélection des variances associées γ_{jk}^2

Estimation de l'effet fixe

Cadre fonctionnel

L'objectif est de proposer un estimateur $\hat{\mu}(t)$ de l'effet fixe fonctionnel $\mu(t)$

- ▷ C'est un des intérêts premiers du praticien car μ représente le profil moyen associé à un groupe homogène d'individus

Réinterprétation du modèle

$$\begin{bmatrix} \mathbf{c}_i \\ \mathbf{d}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + \tilde{\boldsymbol{\varepsilon}}_i \quad \tilde{\boldsymbol{\varepsilon}}_{i,jk} \sim \mathcal{N}(0, \gamma_{jk}^2 2^{-j\eta} + \sigma_\epsilon^2)$$

effet individuel + bruit de mesure

▷ Problème de régression non-paramétrique dans un contexte hétéroscédastique en présence de répétitions individuelles

Borne inférieure pour le risque minimax

- Pour $\mu \in B_{pq}^s(L)$ et $\tilde{\boldsymbol{\varepsilon}}_i$ à variances finies

$$\inf_{\hat{\mu}} \sup_{\mu} \mathbb{E}(\|\hat{\mu} - \mu\|_{L^2}^2) \geq \max \left\{ \mathcal{O} \left[(MN)^{\frac{-2s}{2s+1}} \right]_{(\mathbf{T}_1)}, \mathcal{O} \left[(M)^{-2s'} \right]_{(\mathbf{T}_2)} \right\}$$

Méthodes usuelles dédiées : techniques de seuillage

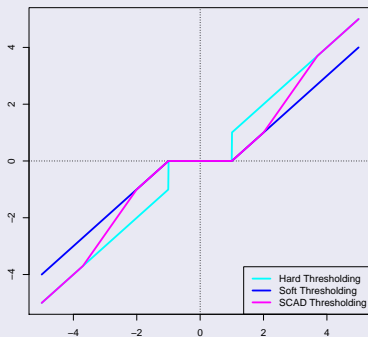
- ▷ Seuillages doux et durs (Donoho & Johnstone, 1994)
- ▷ Seuillage SCAD (Antoniadis & Fan, 2001)

- Seuillage SCAD : compromis entre les seuillages durs et doux évitant :

- le biais du seuillage doux
- l'instabilité du seuillage dur

- λ et $a > 2$ sont des paramètres de régularisation à déterminer
- Seuil universel de Donoho et Johnstone (94)

$$\lambda = \sigma \sqrt{2 \log M}$$



Idée générale

- Étendre le seuil universel à un cadre hétéroscédastique $\lambda = \sigma_{jk} \sqrt{2 \log M}$ (seuil dépendant de la position)
- Les paramètres σ_{jk} sont estimés de manière empirique grâce à la présence de répétitions individuelles
- Propriétés de reconstruction de l'estimateur fonctionnel $\hat{\mu}$?

Borne supérieure pour l'estimateur de seuillage

- $\mu \in B_{pq}^s$, $s' = s - \frac{1}{p} + \frac{1}{2} > 0$, $p \geq 1$, $q \geq 1$ et $s > 1/p$ et $\hat{\mu}$ résultant du seuillage hétéroscédastique
- σ_{jk}^2 bornées et dont on dispose d'estimateurs \sqrt{N} -consistants

$$\text{Then } \mathbb{E}(\|\hat{\mu} - \mu\|_{L^2}^2) \leq \mathcal{O} \left[\left(\frac{\log M}{MN} \right)^{\frac{2s}{2s+1}} \right]_{(T'_1)} + \mathcal{O} \left[\left(\frac{\log M}{M} \right)^{2s'} \right]_{(T'_2)}$$

Propriétés de convergence de l'estimateur

Le seuillage hétéroscédastique atteint bien le taux minimax dans le cadre hétéroscédastique avec répétitions

- Les termes T_1/T'_1 en $\left[(MN)^{\frac{-2s}{2s+1}} \right]$ correspondent au taux minimax dans le cadre homoscédastique usuel avec un bruit réduit
- La limitation vient des termes T_2/T'_2 en $\left[(M)^{-2s'} \right]$, indépendant de N (erreur approximation)

Limites - Sélection des variances des effets aléatoires

- Difficultés à proposer une sélection satisfaisante des paramètres de variances dans cette approche basée sur une procédure non itérative
⇒ Peut conduire à une mauvaise spécification des effets aléatoires

Idée générale

- Tirer parti de l'équivalence entre seuillage et régression pénalisée dans le cadre orthogonal

Sélection des effets fixes [Antoniadis et Fan (01) // Fan et Peng (04)]

- ▷ Critère de vraisemblance pénalisée pour l'estimation/sélection des coefficients d'effets fixes

$$\ell(\boldsymbol{\beta}) = -\log \mathcal{L}(\mathbf{d}, \boldsymbol{\theta}; \mathbf{G}, \boldsymbol{\beta}, \sigma_{\epsilon}^2) + \text{pen}_{\text{SCAD}}(\boldsymbol{\beta}, \lambda_1)$$

avec λ_1 paramètre de régularisation.

Idée générale

- Tirer parti de l'équivalence entre seuillage et régression pénalisée dans le cadre orthogonal

Sélection des effets fixes

$$\ell(\beta) = -\log \mathcal{L}(\mathbf{d}, \theta; \mathbf{G}, \beta, \sigma_\epsilon^2) + \text{pen}_{\text{SCAD}}(\beta, \lambda_1)$$

Double sélection des effets fixes et aléatoires

- ▷ Extension du critère pour l'estimation/sélection des coefficients d'effets fixes et des variances des effets aléatoires

$$\ell(\beta, \gamma) = -\log \mathcal{L}(\mathbf{d}, \theta; \mathbf{G}, \beta, \sigma_\epsilon^2) + \text{pen}_{\text{SCAD}}(\beta, \lambda_1) + \text{pen}_{\text{SCAD}}(\gamma, \lambda_2)$$

avec λ_1 et λ_2 paramètres de régularisation.

Notations

Soit $\Upsilon = (\beta^T, \gamma^T)^T = (\Upsilon_1^T, \Upsilon_2^T)^T$ avec Υ_1 ensemble des paramètres non nuls et Υ_2 ensemble des paramètres nuls

Propriété d'oracle

- L'optimisation du critère de vraisemblance pénalisée conduit à un estimateur $\hat{\Upsilon}$ possédant la propriété d'oracle, à savoir :
 - $\hat{\Upsilon}_2 = 0$ presque sûrement (le bon modèle est atteint presque sûrement)
 - $\hat{\Upsilon}_1$ est asymptotiquement normal
- Cadre de double asymptotique où $N \rightarrow \infty$, $M \rightarrow \infty$ et $\frac{M^5}{N} \rightarrow 0$

Limites

- Ratio $M^5/N \rightarrow 0$ peu réaliste dans un cadre fonctionnel
 - ▷ Peu de réponses dans la littérature actuelle (Kim et al., 2008 ; Fan et Li, 2012)
- Propriétés de reconstruction de l'effet fixe fonctionnel dans ce cadre ?

Optimisation du critère de vraisemblance pénalisée

- Nécessite une reparamétrisation des effets aléatoires θ_i (Chen et Dunson, 2003) pour des contraintes techniques
- Basée sur une variante ECM de l'algorithme EM : Maximisations conditionnelles au cours de l'étape M

Mise à jour de l'effet fixe

- ▷ Se ramène à un seuillage SCAD des données, corrigées des prédictions des effets aléatoires

Mise à jour des variances des effets aléatoires

- ▷ Se ramène à un seuillage SCAD des données centrées et "normalisées" par les prédictions des effets aléatoires

- 1 Contexte
- 2 Modélisation non paramétrique
- 3 Estimation dans les modèles mixtes fonctionnels
 - Problématiques abordées
 - Approche marginale
 - Approche jointe
- 4 Applications
 - Simulations
 - Données réelles
- 5 Perspectives

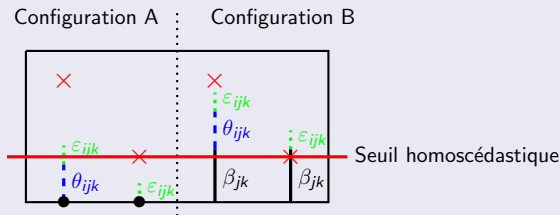
Configurations étudiées

Création de jeux de données synthétiques

- Jeux de données simulés directement dans le domaine des ondelettes
- Contrôle des niveaux de variabilité à l'aide d'un SNR et d'un τ_U fonctionnel
- Possédant une représentation parcimonieuse de l'effet fixe fonctionnel et des variances associées aux effets aléatoires

Configurations étudiées

- A - Effets aléatoires situés sur les coefficients nuls de l'effet fixe
- B - Effets aléatoires situés sur les coefficients non-nuls de l'effet fixe



Paramètres

- $N = 100$, $M = 512$, $\text{SNR} \in \{1, 3, 5, 7\}$ et $\tau_U \in \{0.1, 1/4, 1, 4\}$

Procédures

- Seuillage homoscédastique SCAD usuel [[Homoscédastique](#)]
- Seuillage hétéroscédastique avec estimation empirique des variances [[Type moment](#)]
- Estimation par vraisemblance pénalisée, avec une étape de relaxation (Meinhausen, 2007) [[Minimax.relaxe](#)]

Critères de comparaison

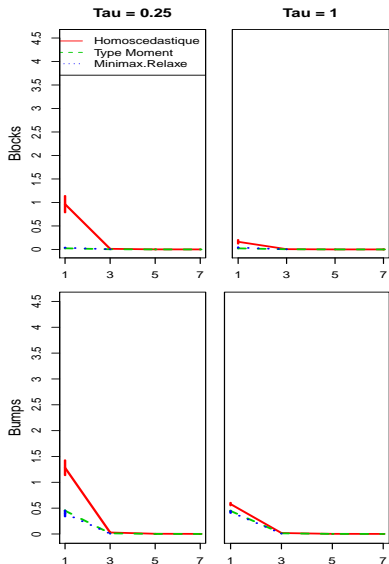
- Reconstruction : Écart Quadratique Moyen

$$\text{EQM}_\mu = \sqrt{\frac{\|\hat{\mu} - \mu_0\|_2^2}{\|\mu_0\|_2^2}} \quad \text{et} \quad \text{EQM}_\sigma = \sqrt{\frac{\|\hat{\sigma} - \sigma_0\|_2^2}{\|\sigma_0\|_2^2}},$$

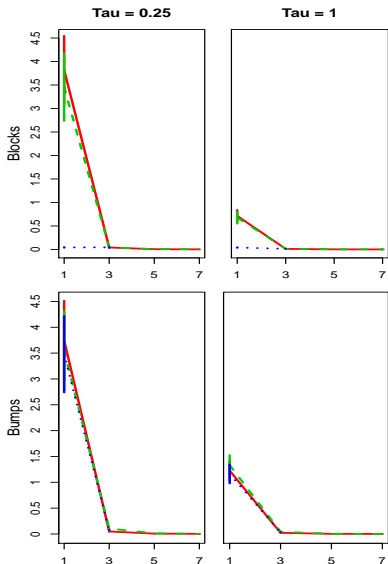
- Sélection : critères de sensibilité et spécificité

Résultats - EQM associé à l'effet fixe fonctionnel

Configuration A



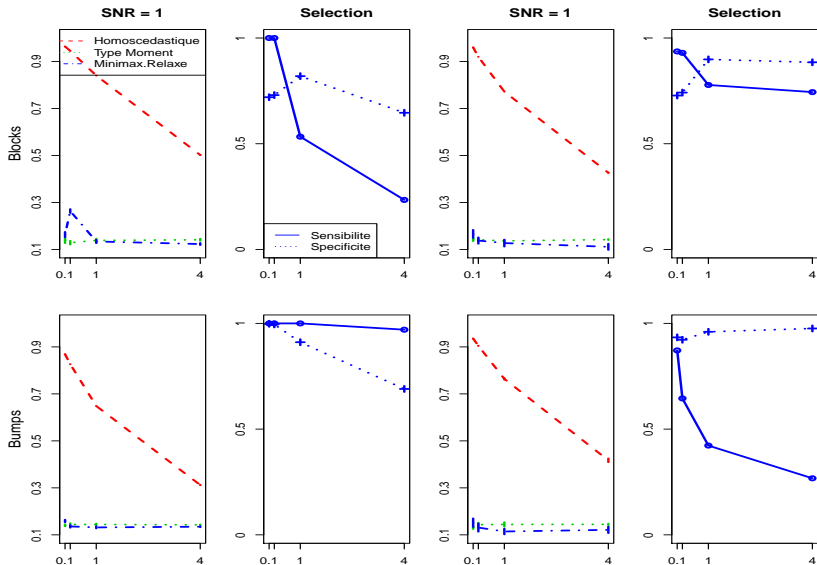
Configuration B



Résultats - Sélection des variances des effets aléatoires

Configuration A

Configuration B



Résultats - Temps d'exécution pour $\tau_U = 0.1$

			Configuration A	Configuration B
Blocks	0.1	Sélection	168.9 (9.101)	192.7 (14.64)
		Type Moment	0.159 (0.019)	0.135 (0.039)
	1	Sélection	231.9 (14.89)	349.1 (243.2)
		Type Moment	0.138 (0.027)	0.199 (0.041)
Bumps	0.1	Sélection	277.4 (25.02)	475.9 (170.1)
		Type Moment	0.157 (0.031)	0.171 (0.032)
	1	Sélection	561.2 (86.54)	476.1 (73.09)
		Type Moment	0.141 (0.021)	0.179 (0.032)

Les temps d'exécution des différentes procédures sont donnés en secondes

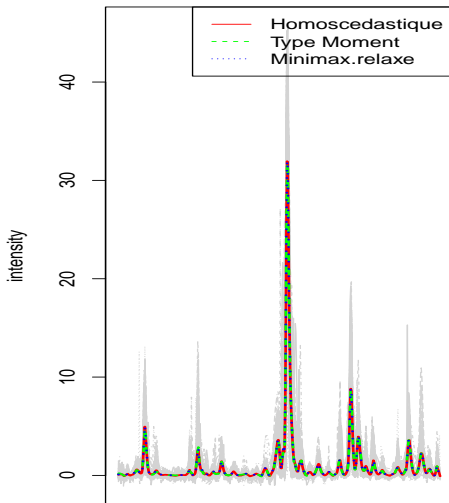
- Pré-traitement des données par groupe : artefacts de mesure, problèmes d'alignement
 - ▷ Utilisation d'une procédure de Antoniadis et al. (2007)
- Restriction à $M = 512$ points de discrétisations

Procédures comparées

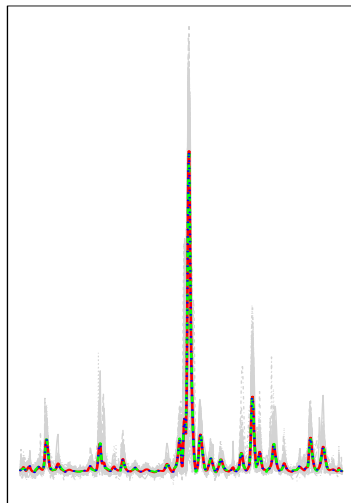
- Reconstruction de l'effet fixe avec :
 - Seuillage SCAD usuel [Homoscédastique]
 - Seuillage SCAD hétéroscédastique [Type moment]
 - Double sélection basée sur l'algorithme EM [Minimax.relaxe]
- Sélection des variances et reconstruction de la covariance des effets aléatoires pour la méthode [Minimax.relaxe]

Reconstruction des effets fixes

individus malades

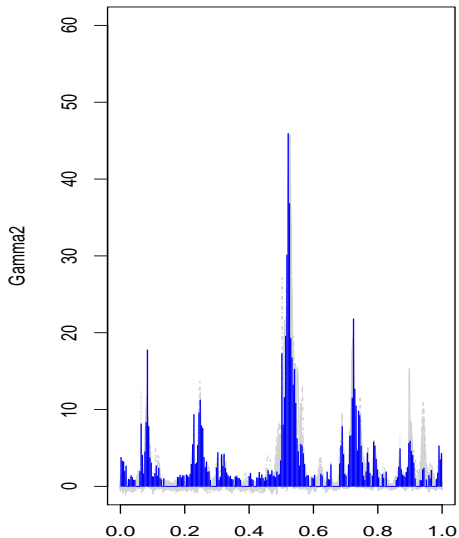


individus sains

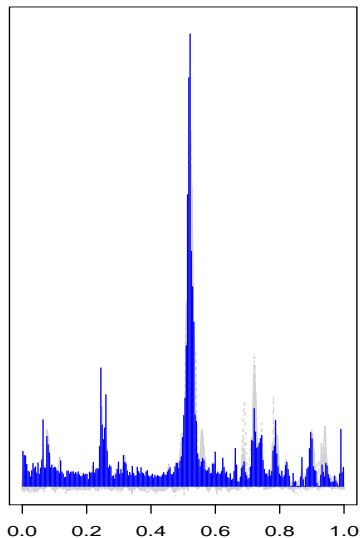


Sélection de variances

individus malades

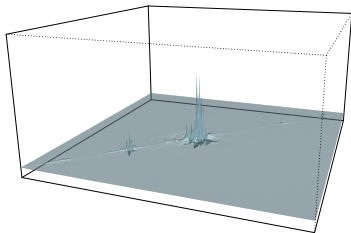


individus sains

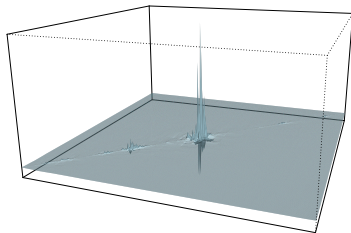


Reconstruction de la covariance des effets aléatoires

- Individus Malades



- Individus Sains



- 1 Contexte
- 2 Modélisation non paramétrique
- 3 Estimation dans les modèles mixtes fonctionnels
 - Problématiques abordées
 - Approche marginale
 - Approche jointe
- 4 Applications
 - Simulations
 - Données réelles
- 5 Perspectives

Approche Marginale








- Implémenter le seuil SURE au sein du seuillage SCAD

Approche Sélection

- Extension des résultats asymptotiques au cas $M > N$
- Propriétés de reconstruction de l'estimateur de l'effet fixe fonctionnel

Applications

- Étude de simulations avec $M \rightarrow \infty$
- Applications à des données de Microarray CGH

-  Antoniadis, A. and Fan, J. (2001), *Regularization of Wavelet Approximations*. JASA, 96 (455), p939-955.
-  Antoniadis, A. and Sapatinas, T. (2007), *Estimation and inference in functional mixed-effects models*. CSDA, 51 (10), p4793-4813.
-  Antoniadis, Bigot, Lambert-Lacroix and Letue (2007), *Non parametric Pre-processing Methods and Inference Tools for Analyzing Time-of-Flight Mass Spectrometry Data*. Current Analytical Chemistry, 3 (2), p127-147.
-  Bondell, H.D., Krishna A. and Ghosh, S.K. (2010), *Joint Variable Selection for Fixed and Random Effects in Linear Mixed-Effects Models*. Biometrics, 66 (4), p1069-1077.
-  Donoho, D.L. and Johnstone, I.M. (1994), *Ideal spatial adaptation by wavelet shrinkage*. Biometrika, 81 (3), p425-455.
-  Gervini, D. and Carter, A.C. (2014), *Warped functional analysis of variances*. Biometrics, 70, p526-535.
-  Petricoin, E.F., Ardekani, A.M. et al. (2002), *Use of proteomic patterns in serum to identify ovarian cancer*. The Lancet, 359 (Feb), p572-577.