Modeles mixtes et pénalité fused lasso pour une comparaison de groupes

Edouard Ollier [1,2] & Adeline Samson[2] & Xavier Delavenne[3] & Vivian Viallon [4]

[1] *U.M.P.A., Ecole Normale Supérieure de Lyon, CNRS UMR 5669; INRIA, Project-team NUMED. 46 Allée d'Italie, 69364 Lyon Cedex 07, France*
[2] *Université Grenoble-Alpes, Laboratoire Jean Kuntzmann, UMR CNRS 5224*
[3] *Groupe de Recherche sur la Thrombose, EA3065, Université de Saint-Etienne, Jean Monnet, F-42023, Saint-Etienne*
[4] *Université de Lyon, F-69622, Lyon, France ; Université Lyon 1, UMRESTTE, F-69373 Lyon ; IFSTTAR, UMRESTTE, F-69675 Bron E-mail : vivian.viallon@univ-lyon1.fr*

Résumé. Nous considérons des données longitudinales possédant une structure de groupes. Par exemple, en recherche clinique, les groupes peuvent correspondre à différentes modalités de traitement. Ces données peuvent être analysées par groupe, un modèle non linéaire mixte étant alors estimé dans chacun de ces groupes. La comparaison entre les groupes est ensuite réalisée en identifiant les paramètres dont l'estimation varie significativement à travers les groupes. Dans ce travail, nous nous intéressons à l'estimation jointe de modèles non-linéaires mixtes par une méthode de vraisemblance pénalisée de type fused lasso. Cette approche permet d'identifier automatiquement les paramètres qui ne varient pas entre certains groupes. La vraisemblance d'un modèle non linéaire mixte étant non explicite, on utilise une version stochastique de l'algorithme EM. L'approche est illustrée par simulation, et utilisée pour les données d'un essai clinique étudiant l'interaction médicamenteuse entre un anticoagulant et un antibiotique.

Mots-clés. Modèles non linéaires à effets mixtes, pénalité, fused lasso, SAEM, sélection de variables.

Abstract. We consider longitudinal data with a group structure. For example in clinical research, the groups correspond to different treatment modalities. These data can be analyzed per group, a non linear mixed model being estimated for each group. The comparison between the groups is then realized identifying the parameters varying among groups. In this work, we propose to jointly estimate the data by maximizing a log-likelihood penalized with a fused lasso penalty. This approach allows to automatically identify the parameters which do not vary between groups. The likelihood of a non linear mixed model being not explicit, we use a stochastic version of the EM algorithm. This approach is illustrated by simulations and applied to real data from a clinical trial studying the drug-drug interaction between an anticoagulant and an antibiotic.

Keywords. Non linear mixed model, fused, lasso, SAEM, joint modeling, variable selection.

1 Introduction

Les Modèles Non Linéaires à Effets Mixtes (MNLEM) sont utilisés pour modéliser les données longitudinales dans de nombreux domaines comme la pharmacocinétique (PK). Dans certains cas, les données présentent une structure de groupe. Un des objectifs de l'analyse statistique consiste alors à déterminer quels paramètres varient significativement entre les différents groupes. On rencontre ce problème notamment en recherche clinique, où les groupes de patients correspondent à différentes modalités de traitement.

Dans la littérature, la détection de différences significatives entre les groupes repose généralement sur des tests statistiques. L'appartenance au groupe est alors intégrée à l'aide d'une covariable, et son influence testée par test du maximum de vraisemblance [9]. Une approche pas à pas combinée à un critère BIC est ensuite utilisée pour sélectionner le meilleur modèle. Un groupe doit donc être défini comme groupe de référence, ce qui peut rendre le problème non symétrique en présence de plus de deux groupes: seules les différences avec le groupe de référence sont étudiées. Une alternative consiste à estimer de manière conjointe les différents groupes [11, 7], en encourageant la similarité entre ceux-ci. C'est notamment le cas de la pénalité fused lasso qui pénalise les différences entre les coefficients et les encourage donc à être égaux.

Les modèles linéaires mixtes avec une pénalité l_1 sur les effets fixes ont déjà été étudiés [10, 8] ainsi qu'une pénalité sur les effets fixes et les variances [2]. La pénalisation jointe des effets fixes et des variances est plus complexe car la vraisemblance n'est pas convexe en les variances. Dans le cas des modèles non linéaire mixtes, le problème est plus difficile car la vraisemblance n'est pas explicite. Une penalité Lasso a été considérée par [1], avec un choix asymptotique de la constante de calibration. A notre connaissance, aucun travail n'a étudié l'application de *pénalité structurée* (fused lasso, group lasso) à la vraisemblance de modèles *non linéaires* à effets mixtes.

Dans ce travail, nous proposons une estimation par vraisemblance pénalisée par fused lasso. La pénalité fused lasso s'appuie sur une structure de graphe entre les groupes, structure qui a son influence dans l'estimation. Après une présentation de l'algorithme d'estimation (une version stochastique pénalisée de l'algorithme EM), nous illustrons la procédure sur données simulées avec 2, 3 ou 5 groupes, et l'influence de la structure de graphes. Enfin, l'algorithme est utilisé pour l'analyse de données réelles.

2 Estimation jointe de MNLEM

Soit $y^g_{i,j}$ l'observation au temps $t^g_{i,j}$ ($j \in \{1, ..., n_i\}$) du i-ème patient ($i \in \{1, ..., N_g\}$) du g-ème groupe ($g \in \{1, ..., G\}$). On considère des modèles de la forme :

$$\begin{aligned}
&y^g_{i,j} = f(t^g_{i,j}, \phi^g_i) + h(t^g_{i,j}, \phi^g_i)\epsilon^g_{i,j} \\
&h = af + b \\
&\epsilon^g_{i,j} \sim \mathcal{N}(0, 1) \text{ (iid)}
\end{aligned}$$

où f est une fonction non linéaire donnée. La fonction h correspond au modèle d'erreur, avec a et b deux réels à estimer. Soit ϕ_i^g le vecteur de dimension p des paramètres du i-ème sujet dans le groupe g. On suppose par ailleurs que le vecteur des paramètres individuels, ϕ_i^g, se décompose de la façon suivante :

$$\phi_i^g \;=\; \mu^g + \omega_i^g, \quad \omega_i^g \sim \mathcal{N}(0, \Omega^g)$$

avec μ^g le vecteur de dimension p des effets fixes du groupe g, ω_i^g le vecteur de dimension p des effets aléatoires du i-ième patient ($i \in \{1, ..., N_g\}$) du g-ième groupe. Enfin, Ω^g est la matrice de variance-covariance des effets aléatoires du groupe g.

La vraisemblance s'écrit:

$$LL(\theta) = \log p(y; \theta) = \log \left(\sum_{g=1}^{G} \int p(y^g, \phi^g; \theta^g) d\phi^g \right) \tag{1}$$

où $p(y^g, \phi^g; \theta^g)$ est la vraisemblance complète du groupe g:

$$\begin{aligned}\log p(y^g, \phi^g; \theta^g) = -\sum_{i,j} \log(d(t_{i,j}^g, \phi_i^g)) - \frac{1}{2} \sum_{i,j} \left(\frac{y_{ij} - f(t_{i,j}^g, \phi_i^g)}{d(t_{i,j}^g, \phi_i^g)} \right)^2 - \frac{N_g}{2} \log(|\Omega^g|) \\ -\frac{1}{2} \sum_{i} (\phi_i^g - \mu^g)^t {\Omega^g}^{-1} (\phi_i^g - \mu^g) - \frac{\sum_i n_i^g + N_g p}{2} \log(2\pi)\end{aligned}$$

avec $\theta = (\theta^1, \ldots, \theta^G)$ et $\theta^g = (\mu^g, \Omega^g, a, b)$ les paramètres à estimer.

L'optimisation de cette vraisemblance (par exemple par algorithme SAEM, une version stochastique de l'algorithme EM [6]) revient à estimer les paramètres séparément pour chacun des groupes. Il se peut que certains groupes partagent des caractéristiques communes, et donc que certains paramètres varient peu ou pas entre ces groupes. En introduisant une pénalité de type fused lasso dans le problème de maximum de vraisemblance on encourage les paramètres dont l'estimation varie faiblement entre deux groupes à être égaux.

3 Pénalité fused lasso

La pénalité fused lasso encourage les paramètres à être similaires entre les groupes. On peut être intéressé par l'étude des différences entre tous les groupes, ou entre certains groupes spécifiques. Ces comparaisons sont décrites à l'aide d'un graphe. On considère alors que deux groupes liés dans le graphe partagent des paramètres communs. Voici des exemples de graphes (voir Figure 1 pour $G = 4$ groupes):

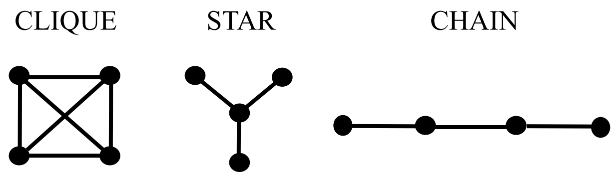

Figure 1: Examples of graphs for $G = 4$ groups

- Graphe Clique: aucune hypothèse de hierarchie sur les groupes. Toutes les différences possibles entre les groupes sont pénalisées.
- Graphe Star: un groupe de référence est choisi et seules les différences par rapport à ce groupe sont pénalisées.
- Graphe Chain: quand les groupes peuvent être ordonnés.

3.1 Paramètres fixes

Pour les paramètres fixes $(\mu^1, \ldots, \mu^G)$, la pénalité fused lasso correspond à :

$$P_F(\mu^1, \ldots, \mu^G) = \sum_{(g_1,g_2)\in\mathcal{E}} \|\mu^{g_1} - \mu^{g_2}\|_1$$

où $\|\|_1$ est la norme ℓ_1, et $\mathcal{E}$ correspond à l'ensemble des arêtes du graphe. La pénalité fused lasso encourage les groupes connectés à avoir des paramètres fixes similaires.

3.2 Paramètres de variance

Pénaliser directement la matrice de variance-covariance amène un problème d'optimisation qui n'est pas convexe. Meme si des solutions ont été proposées dans ce cadre, elles ne sont pas facilement adaptables au cas fused lasso. Nous faisons donc l'hypothèse que Ω^g est diagonale et proposons de pénaliser l'inverse de la matrice de variance-covariance:

$$P_V(\Omega^{1^{-1}}, \ldots \Omega^{G^{-1}}) = \sum_{(g_1,g_2)\in\mathcal{E}} \|\Omega^{g_1^{-1}} - \Omega^{g_2^{-1}}\|_1 = \sum_{i=1}^{p} \sum_{(g_1,g_2)\in\mathcal{E}} |\Omega_{ii}^{g_1^{-1}} - \Omega_{ii}^{g_2^{-1}}|$$

Alors le problème d'optimisation est convexe et peut être résolu rapidement. Ce n'est pas équivalent à pénaliser Ω^g et cela peut poser des problèmes quand les variances sont très hétérogènes. Une solution est alors de considérer une version adaptative, en introduisant des poids (π,ν) dans la pénalité :

$$P_F(\mu^1,\dots,\mu^G) = \sum_{(g_1,g_2)\in\mathcal{E}} \sum_{i=1}^{p} \pi_i^{g_1g_2}|\mu_i^{g_1} - \mu_i^{g_2}|$$

$$P_V(\Omega^{1^{-1}},\dots\Omega^{G^{-1}}) = \sum_{(g_1,g_2)\in\mathcal{E}} \sum_{i=1}^{p} \nu_i^{g_1g_2}|\Omega_{ii}^{g_1^{-1}} - \Omega_{ii}^{g_2^{-1}}|$$

Ces poids peuvent être des estimateurs initiaux obtenus dans chaque groupe [11].

4 Algorithme d'estimation

Nous proposons de maximiser la vraisemblance pénalisée à l'aide d'une version stochastique de l'algorithme EM, l'algorithme SAEM [6]. À l'itération k, l'étape de maximisation correspond à :

1. Mise à jour des effets fixes :

$$\left(\mu_k^1,\dots,\mu_k^G\right) \;=\; \underset{\mu^1,\dots,\mu^G}{\mathrm{ArgMax}} \begin{cases} \sum_{g=1}^{G} \tilde{Q}_k(\mu^g,\Omega_{k-1}^g,a_{k-1},b_{k-1}) \\ - \quad \lambda_F P_F(\mu^1,\dots,\mu^G) \end{cases}$$

   où λ_F est un paramètre à calibrer

2. Mise à jour des variances des effets aléatoires :

$$(\Omega_k^1,\dots,\Omega_k^G) = \underset{\Omega^1,\dots\Omega^G}{\mathrm{ArgMax}} \begin{cases} \sum_{g=1}^{G} \tilde{Q}_k(\mu_k^g,\Omega^g,a_{k-1},b_{k-1}) \\ - \quad \lambda_V P_V(\Omega^{1^{-1}},\dots,\Omega^{G^{-1}}) \end{cases}$$

   où λ_V est un paramètre à calibrer

3. Mise à jour des paramètres du modèle d'erreur : inchangé car non pénalisé.

L'etape 1 est résolue par un algorithme Alternating Direction Method of Multipliers (ADMM) [3] qui resoud ce problème convexe dans sa version duale. Pour l'étape 2, suivant [4], un algorithme ADMM peut également être utilisé, qui est explicite dans le cas de 2 groupes, et non explicite sinon.

L'estimateur proposé dépend des paramètres de sparsité λ_F, λ_V, à choisir. Nous proposons de le sélectionner à l'aide du critère BIC sur une grille prédéfinie par l'utilisateur.

5 Evaluation

L'algorithme est évalué sur des données simulées avec 2, 3 ou 5 groupes. Les différents tests montrent la bonne capacité de l'algorithme à sélectionner le bon modèle. Le choix du graphe (clique, star ou chain) est évaluée dans le cas de 5 groupes, et montre l'impact d'un mauvais choix.

Enfin, la méthode est appliquée sur un jeu de données réelles provenant d'un essai clinique étudiant l'interaction médicamenteuse entre la Clarithromycine (antibiotique) et le Dabigatran (anticoagulant oral) [5].

References

[1] Julie Bertrand and David J Balding. Multiple single nucleotide polymorphism analysis using penalized regression in nonlinear mixed-effect pharmacokinetic models. *Pharmacogenetics and genomics*, 23(3):167–174, 2013.

[2] Howard D Bondell, Arun Krishna, and Sujit K Ghosh. Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, 66(4):1069–1077, 2010.

[3] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[4] Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society, Series B*, 2013.

[5] Xavier Delavenne, Edouard Ollier, Thierry Basset, Laurent Bertoletti, Sandrine Accassat, Arnauld Garcin, Silvy Laporte, Paul Zufferey, and Patrick Mismetti. A semi-mechanistic absorption model to evaluate drug–drug interaction with dabigatran: application with clarithromycin. *British journal of clinical pharmacology*, 76(1):107–113, 2013.

[6] Estelle Kuhn and Marc Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(4):1020–1038, 2005.

[7] Edouard Ollier and Vivian Viallon. Joint estimation of k related regression models with simple l_1-norm penalties. *arXiv preprint arXiv:1411.1594*, 2014.

[8] F. Rohart, M. San Cristobal, and B. Laurent. Selection of fixed effects in high dimensional linear mixed models using a multicycle ecm algorithm. *computational Statistics and Data Analysis*, DOI: 10.1016/j.csda.2014.06.022, 2014.

[9] Adeline Samson, Marc Lavielle, and France Mentré. The saem algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Statistics in medicine*, 26(27):4860–4875, 2007.

[10] J. Schelldorfer, P. Buhlmann, and S. De Geer. Estimation for high-dimensional linear mixed-effects models using 1-penalization. *Scandinavian Journal of Statistics*, 38:197–214, 2011.

[11] Vivian Viallon, Sophie Lambert-Lacroix, Hölger Hoefling, and Franck Picard. On the robustness of the generalized fused lasso to prior specifications. *Statistics and Computing*, pages 1–17, 2014.