Nouvelles stratégies adaptatives pour l'estimation non paramétrique dans des modèles linéaires à effets mixtes.

Charlotte Dion [1,2]

charlotte.dion@imag.fr
(1)*LJK, UMR CNRS 5224, Université Joseph Fourier, 51 rue des Mathématiques, 38041 Grenoble*
(2)*MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints Pères, 75006 Paris*

Résumé. Les modèles linéaires à effets mixtes ont été souvent étudiés dans un cadre paramétrique. Dans ce travail nous présentons à l'aide de méthodes de déconvolution, deux estimateurs non paramétriques de la densité des effets aléatoires. Nous nous plaçons dans le cas où la densité du bruit est connue, mais nous ne faisons pas d'hypothèse sur sa régularité. Une première stratégie utilise toute l'observation disponible et adapte une méthode de sélection innovante due à Goldenshluger et Lepski (2011) ce qui permet d'obtenir des vitesses de convergence inattendues dans ce contexte. La seconde stratégie se base sur la dernière observation. Plus rapide et classique elle reste performante. Deux résultats théoriques principaux sont présentés.

Mots-clés. Estimation non paramétrique, déconvolution, sélection de modèles

Abstract. This study surveys new estimators of the density of a random effect in linear mixed-effects models. Data are contaminated by random noise, and we do not observe directly the random effect of interest. The density of the noise is supposed to be known, without assumption on its regularity. However it can also be estimated. We first propose an adaptive nonparametric deconvolution estimator based on a selection method set up in Goldenshluger and Lepski (2001). Then we propose an estimator based on a simpler model selection device. For both of them, non-asymptotic $\mathbb{L}_2$-risk bounds are established, implying estimation rates much better than the expected deconvolution ones.

Keywords. Nonparametric estimation, deconvolution, linear mixed models, model selection.

1 Introduction

Pour analyser des mesures répétées dans le temps pour plusieurs individus ayant le même comportement, on utilise des modèles mixtes. C'est à un modèle linéaire mixte que nous

nous intéressons ici. L'observation au temps t_j avec $j \in \{0, \dots, J\}$ pour l'individu k, $k \in \{1, \dots, N\}$, est notée $Y_{k,j}$, et vérifie:

$$Y_{k,j} = \alpha_k + \beta_k t_j + \epsilon_{k,j}, \tag{1}$$

où α_k et β_k sont les effets aléatoires de régression et dépendent du sujet k. Les temps d'observation t_j sont connus et équidistants d'un pas Δ: $t_j = j\Delta$. Les variables aléatoires $(\epsilon_{k,j})_{1\leq k\leq N, 0\leq j\leq J}$ modélisent les erreurs de mesure, centrées. Elles sont supposées indépendantes et identiquement distribuées (i.i.d.) avec pour densité f_ϵ. On suppose aussi α_k i.i.d. de densité f_α et β_k i.i.d. de densité f_β. De plus on suppose que $(\epsilon_{k,j})_{1\leq k\leq N, 0\leq j\leq J}$ et $(\alpha_k, \beta_k)_{1\leq k\leq N}$ sont des séquences indépendantes.

Les modèles mixtes sont souvent utilisés, en pharmacocinétique par exemple. Ils ont été très étudiés d'un point de vue paramétrique dans le cas d'effets aléatoires et de bruit gaussiens (voir Pinheiro et Bates (2000)). Ces hypothèses étant très fortes le but de l'article de Dion (2013) présenté ici est de donner une estimation non paramétrique de la densité des effets aléatoires à partir des observations $Y_{k,j}$ et de la suite déterministe des temps.

Nous utilisons une méthode de déconvolution. Ces méthodes ont été étudiées par Fan (1991) avec des estimateurs à noyau puis par Pensky et Vidakovic (1999), Comte et al (2006), et Butucea et Tsybakov (2007) pour les taux de convergence. L'estimation complète non paramétrique de f_β et f_α pour ce modèle a été menée dans Comte et Samson (2012) avec des méthodes de déconvolution. Les résultats présentés ici sont comparés aux leurs dans Dion (2013).

Le but de ce travail est de reconstruire la densité f_β en supprimant les hypothèses sur la loi du bruit (présentes dans Comte et Samson (2012)) et d'améliorer les vitesses de convergence des estimateurs.

Nous présentons d'abord une collection d'estimateurs utilisant toute l'observation disponible. Après une étude de leur risque $\mathbb{L}_2$, nous sélectionnons celui qui réalise le compromis biais-variance parmi la collection en adaptant une méthode de sélection due à Goldenshluger et Lepski (2011) pour des estimateurs à noyau. Enfin nous établissons un résultat de type oracle très satisfaisant. Puis nous construisons un estimateur qui utilise moins d'observations et menons la même étude mais avec une méthode de sélection différente et plus classique à l'aide d'un critère pénalisé (voir Birgé et Massart (1998)). Le cas de la densité du bruit inconnue est étudié dans Dion (2013).

2 Un estimateur utilisant toutes les observations

Notations. Pour deux fonctions f et g de $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, leur produit scalaire est défini par $< f, g >= \int_{\mathbb{R}} f(x)\overline{g(x)}dx$ et la norme associée est $\|f\|^2 = \int_{\mathbb{R}} |f(x)|^2 dx$. La transformée de Fourier de f est $f^*(x) = \int_{\mathbb{R}} e^{ixu} f(u)du$ pour tout $x \in \mathbb{R}$. Enfin le produit de convolution de f avec g pour tout $x \in \mathbb{R}$, est $f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy$.

On considère les variables aléatoires $Z_{k,m}$ telles que pour $k \in \{1,\dots,N\}$, et $m \in \{1,\dots,J\}$

$$Z_{k,m} := \frac{Y_{k,m} - Y_{k,0}}{m\Delta} = \beta_k + \frac{\epsilon_{k,m} - \epsilon_{k,0}}{m\Delta} =: \beta_k + W_{k,m}. \tag{2}$$

Pour un m donné, les variables $W_{k,m}$, $k \in \{1,\dots,N\}$ sont i.i.d. avec pour densité f_{W_m}. Grâce à l'indépendance des ϵ_k avec les β_k, les $Z_{k,m}$ sont i.i.d. et on note leur densité f_{Z_m}, et la définition (2) implique

$$f_{Z_m} = f_\beta \star f_{W_m}. \tag{3}$$

La transformée de Fourier de l'équation précédente (3) mène à $f^*_{Z_m} = f^*_\beta f^*_{W_m}$. Par indépendance des $\epsilon_{k,m}$, pour tout $u \in \mathbb{R}$, on a

$$f^*_{W_m}(u) = \mathbb{E}[e^{iu(\epsilon_{k,m}-\epsilon_{k,0})/m\Delta}] = \mathbb{E}[e^{iu\epsilon_{k,m}/m\Delta}]\mathbb{E}[e^{-iu\epsilon_{k,0}/m\Delta}] = \left|f^*_\epsilon\left(\frac{u}{m\Delta}\right)\right|^2.$$

On suppose dans la suite que les deux hypothèses suivantes sont vérifiées: **(H1)** $f_\beta \in \mathbb{L}_2(\mathbb{R})$ et $f^*_\beta \in \mathbb{L}_1(\mathbb{R})$, **(H2)** f_ϵ et connue et $f^*_\epsilon \neq 0$. Sous l'hypothèse **(H2)** on obtient pour tout $u \in \mathbb{R}$

$$f^*_\beta(u) = \frac{f^*_{Z_m}(u)}{|f^*_\epsilon(\frac{u}{m\Delta})|^2},$$

et sous l'hypothèse **(H1)**, l'inversion de Fourier nous fournit la formule fermée suivante, pour tout $x \in \mathbb{R}$

$$f_\beta(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iux}\frac{f^*_{Z_m}(u)}{|f^*_\epsilon(\frac{u}{m\Delta})|^2}du\ . \tag{4}$$

Pour estimer $f^*_{Z_m}$ on utilise son estimateur empirique. Cependant on ne remplace pas directement cet estimateur dans la formule (4), pour éviter les problèmes d'intégrabilité dûs au dénominateur. On introduit une coupure, une borne dans l'intégrale $m\Delta$ où m est l'indice du temps d'observation. On obtient:

$$\widehat{f}_{\beta,m}(x) = \frac{1}{2\pi}\int_{-m\Delta}^{m\Delta} e^{-iux}\frac{\widehat{f}^*_{Z_m}(u)}{|f^*_\epsilon(\frac{u}{m\Delta})|^2}du = \frac{1}{2\pi}\int_{-m\Delta}^{m\Delta} e^{-iux}\frac{1}{N}\sum_{k=1}^{N}\frac{e^{iuZ_{k,m}}}{|f^*_\epsilon(\frac{u}{m\Delta})|^2}du\ . \tag{5}$$

Nous allons voir comment ce choix original de coupure va produire des vitesses inattendues pour l'estimateur $\widehat{f}_{\beta,m}$.

Étude du risque de l'estimateur $\widehat{f}_{\beta,m}$

Dans la suite on définit $f_{\beta,m}$ par la relation $f_{\beta,m}(x) := \frac{1}{2\pi}\int_{-m\Delta}^{m\Delta} e^{-iux}f^*_\beta(u)du$. L'étude de la décomposition biais-variance donne la majoration du risque $\mathbb{L}_2$ suivante.

Proposition 1 *L'estimateur $\widehat{f}_{\beta,m}$ est un estimateur sans biais de $f_{\beta,m}$: $\mathbb{E}[\widehat{f}_{\beta,m}] = f_{\beta,m}$. Sous les hypothèses **(H1)-(H2)**, en notant $F_\epsilon := \int_{-1}^{1}(1/|f^*_\epsilon(v)|^4)dv$, on a*

$$\mathbb{E}[\|\widehat{f}_{\beta,m} - f_\beta\|^2] \leq \frac{1}{2\pi}\int_{|u|\geq m\Delta}|f^*_\beta(u)|^2du + F_\epsilon\frac{m\Delta}{2\pi N}\ . \tag{6}$$

Le premier terme du majorant (6) est le biais intégré $\|f_{\beta,m} - f_\beta\|^2$, il représente l'erreur commise en tronquant les bornes de l'intégrale dans $\widehat{f}_{\beta,m}$ et décroît quand m croît. Le second terme est le terme de variance. Il croît avec m et provient de l'estimation de $f_{\beta,m}$ par $\widehat{f}_{\beta,m}$. Il est d'ordre m/N: c'est l'ordre obtenu lorsque l'on a des données non bruitées. Enfin, la borne obtenue ne dépend de la densité du bruit qu'à travers la constante F_ϵ.

Sélection adaptative

Nous voulons maintenant choisir m le mieux possible parmi la collection, c'est-à-dire en minimisant le risque donc en réalisant le compromis biais-variance. Supposons que m appartienne à un ensemble fini $\mathcal{M}$, alors le choix optimal théorique pour m serait

$$m_{th} = \underset{m\in\mathcal{M}}{\operatorname{argmin}} \{\|f_{\beta,m} - f_\beta\|^2 + V(\widehat{f}_{\beta,m})\},$$

où $V(\widehat{f}_{\beta,m}) := \int \mathbb{E}[(\widehat{f}_{\beta,m}(x) - \mathbb{E}[\widehat{f}_{\beta,m}(x)])^2]dx = \mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2] \leq F_\epsilon m\Delta/(2\pi N)$. Les deux termes de cette décomposition sont inconnus et doivent être estimés. Le terme de variance est remplacé par un terme de pénalité, du même ordre:

$$\text{pen}^{(1)}(m) = \kappa^{(1)} F_\epsilon \frac{m\Delta}{N}$$

avec $m \in \mathcal{M} := \{1, \ldots, N \wedge J\}$ où $N \wedge J = \min(N, J)$ et $\kappa^{(1)}$ est une constante numérique calibrée par une étude intensive sur simulations. Notre choix de $\mathcal{M}$ implique que $m \leq N$ et la pénalité reste bornée. La dépendance en m présente à plusieurs endroits nous empêche d'écrire l'estimateur $\widehat{f}_{\beta,m}$ comme le minimiseur d'un contraste (voir Birgé et Massart (1998)). C'est pourquoi nous utilisons une méthode de sélection plus sophistiquée, due à Goldenshluger et Lepski (2011). L'indice m est choisi de manière adaptative de la façon suivante:

$$\widehat{m}^{(1)} = \underset{m\in\mathcal{M}}{\operatorname{argmin}} \{\widehat{\Gamma}_m + \text{pen}^{(1)}(m)\}$$

$$\widehat{\Gamma}_m := \max_{1\leq j\leq N\wedge J} \left(\|\widehat{f}_{\beta,m\wedge j} - \widehat{f}_{\beta,j}\|^2 - \text{pen}^{(1)}(j)\right)_+ = \max_{m\leq j\leq N\wedge J} \left(\|\widehat{f}_{\beta,m} - \widehat{f}_{\beta,j}\|^2 - \text{pen}^{(1)}(j)\right)_+$$

où $(.)_+$ est la partie positive. La quantité $\widehat{\Gamma}_m$ approxime le biais. L'estimateur final $\widehat{f}_{\beta,\widehat{m}^{(1)}}$ satisfait l'inégalité de type oracle suivante.

Théorème 2 *Sous **(H1)-(H2)**, il existe deux constantes C et C' telles que*

$$\mathbb{E}[\|\widehat{f}_{\beta,\widehat{m}^{(1)}} - f_\beta\|^2] \leq C \inf_{m\in\mathcal{M}} (\|f_{\beta,m} - f_\beta\|^2 + \text{pen}^{(1)}(m)) + \frac{C'}{N}. \tag{7}$$

La constante C est numérique dès que $\kappa^{(1)}$ est fixé.

Le théorème 2 produit un résultat non asymptotique et la borne (7) montre que $\widehat{f}_{\beta,\widehat{m}^{(1)}}$ réalise automatiquement le compromis biais-variance. De plus si f_β est *ordinary smooth* de régularité b, l'ordre du risque est $N^{-2b/(2b+1)}$ lorsque N est grand. C'est celui que l'on obtient dans un cas non bruité.

3 Un estimateur construit à l'aide de la dernière observation

Avec l'idée de proposer un estimateur et une méthode de sélection plus simple nous proposons un nouvel estimateur de f_β basé sur les variables aléatoires suivantes:

$$Z_{k,J} = \frac{Y_{k,J} - Y_{k,0}}{J\Delta} = \beta_k + \frac{\epsilon_{k,J} - \epsilon_{k,0}}{J\Delta} = \beta_k + W_{k,J},\ \forall k \in \{1, \dots, N\}$$

Avec les mêmes arguments que précédemment on peut construire l'estimateur suivant:

$$\widehat{f}^J_{\beta,m}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{k=1}^{N} \frac{e^{iuZ_{k,J}}}{|f^*_\epsilon(\frac{u}{J\Delta})|^2} du \ . \tag{8}$$

Il utilise moins d'observations que le précédent: $2N$ au lieu de NJ. Nous devons nous assurer qu'il reste malgré tout performant en étudiant son risque $\mathbb{L}_2$.

Proposition 3 *Si $m \leq N \wedge J$, et $B_\epsilon := \sup_{v\in[-1,1]} \frac{1}{|f^*_\epsilon(v)|^4}$, alors sous **(H1)-(H2)***

$$\mathbb{E}[\|\widehat{f}^J_{\beta,m} - f_\beta\|^2] \leq \frac{1}{2\pi} \int_{|u|\geq m\Delta} |f^*_\beta(u)|^2 du + B_\epsilon \frac{m\Delta}{\pi N}. \tag{9}$$

Ici encore la dépendance en la loi du bruit est contenue dans la constante B_ϵ. On obtient donc un ordre de variance en m/N, de type densité, ce qui est le mieux que l'on puisse espérer. Ici m n'est plus un temps d'observation mais seulement la coupure: cet estimateur peut être vu comme argument minimum d'un contraste de type densité. On estime la meilleur valeur théorique de m par:

$$\widehat{m}^{(2)} = \underset{m\in\mathcal{M}}{\operatorname{argmin}} \{-\|\widehat{f}^J_{\beta,m}\|^2 + \mathrm{pen}^{(2)}(m)\} \tag{10}$$

où

$$\mathrm{pen}^{(2)}(m) = \kappa^{(2)} B_\epsilon \frac{\Delta m}{N}$$

avec $\kappa^{(2)}$ est un constante calibrée pendant l'étude numérique. Nous obtenons alors pour l'estimateur final $\widehat{f}^J_{\beta,\widehat{m}^{(2)}}$ la majoration du risque $\mathbb{L}_2$ suivante .

Théorème 4 *Sous **(H1)-(H2)**, il existe deux constantes C_1 et C_2 telles que*

$$\mathbb{E}[\|\widehat{f}^J_{\beta,\widehat{m}^{(2)}} - f_\beta\|^2 \leq C_2 \inf_{m\in\mathcal{M}} (\|f_{\beta,m} - f_\beta\|^2 + \mathrm{pen}^{(2)}(m)) + \frac{C_1}{N}.$$

Le choix $\widehat{m}^{(2)}$ réalise automatiquement le compromis biais-variance. Cette méthode par pénalisation est plus simple et plus rapide que la procédure vue pour obtenir $\widehat{f}_{\beta,\widehat{m}^{(1)}}$.

4 Conclusion

Les deux estimateurs que l'on propose sont convaincants d'après les résultats théoriques et validés par une étude numérique et l'étude de vitesses dans des cas particuliers (non présentés ici). Cela permet en particulier de définir un cadre d'application pour chacun. En effet, lorsque J est petit et que toutes les observations sont disponibles on préfèrera l'estimateur $\widehat{f}_{\beta,m}$, lorsque $J\Delta$ est grand et/ou que la première et la dernière observations sont disponibles on utilisera plutôt $\widehat{f}^J_{\beta,m}$.

Bibliographie

[1]Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329–375.

[2]Butucea, C. and Tsybakov, A. (2007). Sharp optimality in density deconvolution with dominating bias. ii. *Teor. Veroyatnost. i Primenen.* **52**, 336–349.

[3]Comte, F., Genon-Catalot, V. and Samson, A. (2013). Nonparametric estimation for stochastic differential equation with random effects. *Stochastic Process. Appl.* **7**, 2522–2551.

[4]Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.* **34**, 431–452.

[5]Comte, F. and Samson, A. (2012). Nonparametric estimation of random-effects densities in linear mixed-effects model. *J. Nonparametr. Statist.* **24**, 951–975.

[6] Dion (2013) New adaptive strategies for nonparametric estimation in linear mixed models. Preprint hal-00906379

[7]Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.

[8]Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608–1632.

[9]Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342**, 877–882.

[10]Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 2033–2053.

[11]Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus.* Springer-Verlag, New York.