

INFÉRENCE PAR APPROXIMATION NORMALE DE L'*a posteriori* DANS LES MODÈLES DYNAMIQUES À EFFETS MIXTES.

Mélanie Prague ¹ & Daniel Commenges ¹ & Jérémie Guedj ² & Julia Drylewicz ³ & Rodolphe Thiébaud ¹

¹ *Université de Bordeaux, ISPED INSERM U897 épidémiologie-Biostatistique, France.*

² *Université de Paris Diderot, INSERM UMR 738, Institut Claude Bernard, France*

³ *Department of Immunology, University Medical Center Utrecht, The Netherlands*

Résumé. Les modèles basés sur des équations différentielles sont des outils très utilisés pour décrire les systèmes dynamiques. L'usage de modèles à effets mixtes pour estimer les paramètres en population vient naturellement et permet de tirer avantage de la variabilité entre les individus. Depuis 2007, un algorithme basé sur une approximation normale de l'*a posteriori* permet d'estimer les paramètres dans ces problèmes, mais son utilisation difficile a limité son accessibilité. Dans un contexte Bayésien afin de réduire les problèmes d'identifiabilité, cette méthode consiste à calculer le maximum *a posteriori* (MAP) par une optimisation de la vraisemblance pénalisée. Le programme NIMROD (Normal approximation Inference in Models with Random effects based on Ordinary Differential equations) écrit en Fortran est désormais disponible. Nous décrirons la méthode et les spécificités de cet algorithme: un nouveau critère de convergence appelé la Distance Relative au Maximum (RDM) et un critère de choix de modèle basé sur l'approximation de la validation croisée (LCVa). Les illustrations et la comparaison avec des méthodes classiques MCMC se fera sur des exemples de pharmacocinétique : sur données simulées puis sur des données réelles de l'essai clinique PUZZLE pour le dosage plasmatique de la concentration d'un agent antiretroviral, l'amprenavir.

Mots-clés. Equation Différentielles; Modèles non linéaires à effets mixtes; Estimation Bayésienne; Maximum *a posteriori*; Pharmacocinétique; Programme; Optimisation.

Abstract. Models based on ordinary differential equations (ODE) are a widespread tool to describe dynamical system. The use of mixed effect models to estimate population parameters is natural to take advantage of between-subject variability. Since 2007, an algorithm based on a normal approximation of the posterior allowed estimation of parameters in such a context but lack of accessibility reduced the broadcast of the approach. In a Bayesian setting to enhance identifiability, it consists in computing the maximum of the posterior distribution (MAP) by an optimization of the penalized likelihood. The NIMROD program (Normal approximation Inference in Models with Random effects based on Ordinary Differential equations) written in Fortran is now available. We will describe the method and the specific features implemented such as a convergence criterion called the relative distance to maximum (RDM) and an approximation of the leave-one-out cross

validation for model selection (LCVa). Illustration and comparison with classical full-Bayesian methods based on MCMC will be performed on simulated Pharmacokinetics data and on real data from the PUZZLE clinical trials for the plasma concentration titration of an antiretroviral agent, the amprenavir.

Keywords. Ordinary Differential Equations; Non-Linear Mixed effects model; Bayesian estimation; Maximum *a posteriori*; Pharmacokinetics; Program; Optimization.

1 Introduction

Quand les interactions entre plusieurs variables sont dynamiques, les modèles mécanistes basés sur des équations différentielles Ordinaires (ODE) sont particulièrement adaptés. Dans ces modèles, il s'agit de prendre en compte des données longitudinales avec des variabilités inter-sujets grâce aux modèles non linéaires à effets mixtes.

Les méthodes classiques d'estimation consistent à utiliser un solveur numérique d'ODE afin de calculer la log-vraisemblance puis de la maximiser. Cependant, la présence d'effet aléatoire rend le calcul de la vraisemblance complexe avec la présence d'intégrale. Des méthodes basées sur son approximation ont été développées et implémentées: nous pouvons citer le logiciel NONMEM (Pinheiro et Bates, 1995) ou encore MONOLIX (Kuhn et Lavielle, 2005) tirant partie de l'algorithme SAEM (Stochastic Approximation Expectation Maximization). Cependant, ces algorithmes peuvent être bloqués dans des régions sous-optimales de l'espace des paramètres en particulier dû à des critères souples de convergence. Une des meilleures alternatives réside dans le calcul numérique de l'intégrale basé sur une quadrature adaptative de Gauss (Plan et al. 2012). De manière complémentaire au cadre fréquentiste, les techniques Bayésiennes ont permis d'améliorer l'estimation des paramètres en résolvant les problèmes de non identifiabilité discutés par Guedj et al (2007a). Les techniques les plus utilisées sont les méthodes de Monte Carlo par Chaînes de Markov (MCMC) implémentées notamment dans le logiciel Winbugs. Cependant, elles sont très coûteuses en temps de calcul et la convergence de ces chaînes est une fois encore difficile à évaluer. Nous proposons donc une alternative basée sur la maximisation de la vraisemblance pénalisée présentée par Guedj et al. (2007b) et le programme NIMROD associé (Prague et al. 2013).

2 Modélisation et algorithme d'estimation

2.1 Modèle mathématique, statistiques et d'observations

Nous considérons un modèle ODE pour une population n individus. Pour le sujet i nous pouvons écrire :

$$\begin{cases} \frac{dX^i(t)}{dt} = f(X^i(t), \xi^i(t)) \\ X^i(0) = h(\xi^i(0)) \end{cases}$$

où $\xi^i(t) = (\xi_1^i(t), \dots, \xi_{n_b}^i(t))$ est un vecteur de paramètres de longueur n_b qui apparaît naturellement dans le système mécaniste et a une interprétation biologique; $X^i(t) = (X_1^i(t), \dots, X_K^i(t))$ est le vecteur représentant les K compartiments. Nous supposons que la connaissance des $\xi_i(t)$ et de f permet d'obtenir la trajectoire des $X^i(t)$ au cours du temps. Nous supposons que f and h sont des fonctions possiblement non linéaires deux fois différentiable par rapport à $\xi^i(t)$.

A partir de ce modèle mathématique nous construisons un modèle statistique basé sur une transformation des paramètres $\tilde{\xi}_l^i(t) = \psi_l(\xi_l^i(t))$ (souvent le logarithme quand on considère des taux pour assurer la positivité). Pour prendre en compte les variabilités inter-patients nous appliquons un modèle à effets mixtes sur les paramètres du modèle, ainsi :

$$\tilde{\xi}_l^i(t) = \phi_l + \beta_l z_l^i(t) + u^i, \quad u^i \sim \mathcal{N}(0, \sigma_l^2).$$

où ϕ_l est la constante de base et z_l^i est un vecteur de n_e variables explicatives (possiblement dépendantes du temps) associées au l^{th} paramètre; β_l est le vecteur des coefficients de régression; u^i est le vecteur des effets aléatoire individuels.

Dans la pratique, il est rarement possible d'observer $X^i(t)$ directement, mais nous avons des observation discrètes $Y^i(t_{ij})$ de certaines fonctions $g_m(\cdot)$, $m = 1, \dots, M$ de $X^i(t)$. Ainsi, pour $i = 1, \dots, n$, $m = 1, \dots, M$ et $j = 1, \dots, n_i$ (le nombre d'observation pour le patient i), le modèle d'observations est le suivant:

$$Y_m^i(t_{ij}) = g_m(X^i(t_{ij})) + \epsilon_{ijm}, \quad \epsilon_{ijm} \sim \mathcal{N}(0, \sigma_m^2).$$

Il s'agit d'estimer le vecteur $\theta = \left((\phi_l)_{l=1 \dots n_b}, (\beta_l)_{l=1 \dots n_e}, (\sigma_l)_{l=1 \dots q}, (\sigma_m)_{m=1 \dots M} \right) \in \mathbb{R}^p$.

2.2 Vraisemblance et approximation Normale de l'*a posteriori*

La vraisemblance est calculée d'après les formules analytiques disponibles dans Guedj et al (2007). Il s'agit dans un premier temps de calculer la vraisemblance conditionnellement aux effets aléatoires individuels ($\mathcal{L}_{\mathcal{F}_i|u_i}^\theta$). La formule analytique découle du fait que les $Y_m^i(t_{ij})$ sont supposés *iid* gaussiens. Pour cela, les trajectoires $g_m(X^i(t_{ij}))$ sont calculées à l'aide d'un solveur d'équation différentielle numérique "dlsode" développé par Hindmarsh et al. (1983). Ensuite, la vraisemblance individuelle (\mathcal{L}_i^θ) est calculées par intégration numérique via une quadrature adaptative gaussienne implémentée par Genz et al. (1996). La log-vraisemblance découle de la somme des log-vraisemblances individuelles et est pénalisée par $J(\theta)$ le logarithme de l'*a priori* : $L^P(\theta) = L(\theta) + J(\theta)$.

Les *a priori* peuvent être introduits sur les paramètres ou une combinaison des paramètres pour augmenter l'identifiabilité du modèle. Traditionnellement, chaque paramètre biologique (ϕ) du modèle a pour *a priori* une loi normale de moyenne et variance issue de la littérature. Cependant, les *a priori* peuvent aussi être choisis de manière à effectuer une synthèse d'information Bayésienne : prendre pour nouvel *a priori* l'*a posteriori* issu de l'analyse d'un essai précédent. Asymptotiquement, l'approximation normale de l'*a posteriori* ne perturbe pas l'analyse bayes séquentielle.

2.3 Procédure d'optimisation

Sous les conditions du théorème de Bernstein-Von Mises (van Der Vaart, 2000), l'*a posteriori* tend vers une loi normale. Ainsi quand n est grand, la log-vraisemblance pénalisée est proche d'une quadrique en θ ce qui justifie l'utilisation d'un algorithme de type Newton-Raphson. Les gradients, $U^P(\theta^{(k)})$, sont donc calculés par des méthodes analytiques numériques analogues à celle utilisées pour la vraisemblance. Pour réduire les temps de calculs lié à la dérivée seconde de la log-vraisemblance pénalisée, nous utilisons l'algorithme Robust Variance Scoring (RVS) proposé par Commenges et al. (2006) qui approxime la Hessienne de la vraisemblance pénalisée (H_{LP}) par la variance des gradients (G). Cet algorithme permet de définir la direction du déplacement selon tous les paramètres à estimer. Nous proposons ensuite une recherche linéaire de la meilleure magnitude de déplacement. De manière itérative, nous définissons une suite $\theta^{(k)}$ tendant vers le mode de la distribution *a posteriori* des paramètres.

2.4 Critère de convergence

Les critères de convergence classique de stabilisation du déplacement ($\|\theta^{(k+1)} - \theta^{(k)}\|$) et de la valeur de la log-vraisemblance pénalisée ($L^P(\theta^{(k+1)}) - L^P(\theta^{(k)})$) sont implémentés. Cependant, ils ne suffisent pas à assurer la convergence car ces critères ne sont pas directement liés à la surface à optimiser. Nous proposons donc dans NIMROD d'utiliser la distance relative au Maximum (RDM) définie comme la norme des gradients dans la métrique des paramètres :

$$\text{RDM}(\theta^{(k)}) = \frac{U^P(\theta^{(k)})^T G^{-1}(\theta^{(k)}) U^P(\theta^{(k)})}{p}$$

Nous expliquerons pourquoi cet indicateur peut-être vu comme le ratio entre l'erreur numérique et l'erreur statistique en $\theta^{(k)}$. Pour qu'il y ait convergence il faut donc que ce ratio soit inférieur à 1 et aussi proche de 0 que possible.

2.5 Choix des variables explicatives, choix du modèle

Nous avons implémenté dans NIMROD une p-valeur Bayésienne analogue à la p-valeur fréquentiste pour le test de significativité des régresseurs d'un modèle mixte. Pour un test unilatéral " $\beta \leq 0$ ", c'est simplement $\mathbb{P}(\beta \leq 0|Y)$, une expression analogue existe pour les tests bilatéraux. Asymptotiquement, cette p-valeur Bayésienne est consistante avec la p-valeur fréquentiste obtenue pour un test de Wald.

Commenges et al. (2007) ont proposé un critère de choix de modèle basé sur l'approximation du critère de validation croisée :

$$\text{LCV}_a(\theta) = -n^{-1}[\text{L}(\theta) - \text{Tr}(H_{LP}^{-1}(\theta)H_L(\theta))],$$

Ce critère est implémenté dans NIMROD. Plus petit il est, meilleur est le modèle.

3 Illustration

Nous avons simulé des données issues d'un modèle Pharmacocinétique (PK) à deux compartiments (oral et plasma) que nous avons analysées pour évaluer les performances de NIMROD. Nous montrons de bonnes qualités d'estimation et une faible sensibilité à l'*a priori* dans le cadre de modèles bien spécifiés. De plus, nous montrons que la valeur du RDM au point de convergence est un indicateur du nombre de chiffres significatifs obtenus sur les paramètres et la vraisemblance pénalisée au maximum. Par la suite, nous avons comparé les performances de NIMROD à celles de la fonction *nlme* du package R PKTools et de Winbugs pour les méthodes MCMC en faisant varier le nombre de sujets, le nombre de données répétées et l'intensité du bruit sur ces données. Compétitif au niveau temps, nous montrons que les biais relatifs, RMSE et taux de couvertures à 95% ne sont pas significativement différents. Sur des problèmes plus complexes, Drylewicz et al. (2012) montre la supériorité de l'approche MAP par rapport à celle MCMC.

Dans un second temps, nous avons considéré des données réelles issues de l'essai clinique PUZZLE. Il s'agit de données PK de concentration plasmatique pour le dosage d'un agent antirétroviral (Amprenavir) chez des patients infectés par le VIH. Nous illustrerons comment introduire et vérifier la significativité de variables explicatives telles que le poids. Nous verrons comment se servir du LCVa comme critère de choix de modèle entre un modèle à 2 et 3 compartiments (ajout du compartiment tissulaire) pour la PK.

4 Le programme NIMROD

Le programme NIMROD est écrit en Fortran. Il est disponible en téléchargement libre au site internet <http://www.isped.u-bordeaux2.fr/NIMROD/Index.aspx>. Ce programme est diffusé sous les termes de la License générale publique GNU publiée par la Free Software Foundation. Des bibliothèques de fonction pour des modèles pharmacocinétiques et VIH (Prague et al, 2012) sont disponibles en ligne sur la page de documentation: <http://www.isped.u-bordeaux2.fr/NIMROD/documentation.aspx>.

Ce programme peut facilement être compilé et exécuté sur toutes plateformes, de préférence possédant un système Unix. De plus, en fonction de la présence sur la plateforme d'exécution de bibliothèques de calculs parallèles, la parallélisation des calculs est possible dans NIMROD. L'utilisateur peut alors paralléliser les calculs sur autant de processeurs que de sujets considérés dans son étude.

Bibliographie

[1] Pinheiro J. C. et Bates D. M. (1995), Approximations to the log-likelihood function in the nonlinear mixed-effects model *Journal of Computational and Graphical Statistics* 12-35.

- [2] Kuhn E. et Lavielle M. (2005), Maximum likelihood estimation in nonlinear mixed effects models, *Computational Statistics & Data Analysis*, 1020–1038.
- [3] Plan E.L., Maloney A., Mentré F., Karlsson M. O. et Bertrand J. (2012), Performance Comparison of Various Maximum Likelihood Nonlinear Mixed-Effects Estimation Methods for DoseResponse Models *The AAPS journal* 1-13.
- [4] Guedj J., Thiébaud R. et Commenges D. (2007), Practical identifiability of HIV dynamics models *Bulletin of Mathematical Biology* 2493-2513.
- [5] Guedj J., Thiébaud R. et Commenges D. (2007), Maximum likelihood estimation in dynamical models of HIV *Biometrics* 1198-1206.
- [6] Prague M., Commenges D., Guedj J., Drylewicz J. et Thiébaud R (2013) NIMROD: A Program for Inference via Normal Approximation of the Posterior in Models with Random effects based on Ordinary Differential Equations, *submitted in Computation Methods and Programs in Biomedecine*.
- [7] Radhakrishnan K. et Hindmarsh A.C. (1983), Description and use of LSODE, the Livermore solver for ordinary differential equations *National Aeronautics and Space Administration*.
- [8] Genz A. et Keister B.D. (1993), Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight *Journal of Computational and Applied Mathematics* 299–309.
- [9] Van der Vaart A.W. (2000), Asymptotic statistics *Cambridge Univ Pr* 138–152.
- [10] Commenges D., Jacqmin-Gadda H., Proust C. et Guedj J. (2006), A newton-like algorithm for likelihood maximization : The robust-variance scoring algorithm *Arxiv math/0610402*.
- [11] Commenges D., Joly P., Gégout-Petit A., et Liqueur B. (2007), Choice between Semi-parametric Estimators of Markov and Non-Markov Multi-state Models from Coarsened Observations *Scandinavian Journal of Statistics* 33–52.
- [12] Drylewicz J., Commenges D. et Thiébaud R. (2012), Maximum *a posteriori* Estimation in Dynamical Models of Primary HIV Infection *Statistical Communications in Infectious Diseases* in press.
- [13] Prague M., Commenges D., Drylewicz J. et Thiébaud R. (2012), Treatment monitoring of HIV infected patients based on mechanistic models *Biometrics* 902–911.