



Estimation dans les modèles non linéaires mixtes décrits par des systèmes d'équations différentielles stochastiques. Applications en pharmacologie de population

Julie Marcelin

► **To cite this version:**

Julie Marcelin. Estimation dans les modèles non linéaires mixtes décrits par des systèmes d'équations différentielles stochastiques. Applications en pharmacologie de population. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494845>

HAL Id: inria-00494845

<https://hal.inria.fr/inria-00494845>

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ESTIMATION DANS LES MODÈLES NON LINÉAIRES MIXTES
DÉCRIS PAR DES SYSTÈMES D'ÉQUATIONS
DIFFÉRENTIELLES STOCHASTIQUES.
APPLICATIONS EN PHARMACOLOGIE DE POPULATION

Julie Marcelin

Université Paris Sud
Département de Mathématiques
Bâtiment 425
91400 ORSAY

Abstract :

The mixed effects models are widely used in numerous domains like pharmacometry in which they allow to take into account the inter-subjects variability of physiological parameters.

In the continuous data case, the estimation of the model parameters by maximum likelihood methods has been widely studied, in particular when the regression function is governed by a system of ordinary differential equations.

Nevertheless, several modelisers in pharmacology consider that the intra-subject variability is not always well modelised by the residual errors.

They propose to add a stochastic component in the dynamic model itself, that is, the model is governed by a system of stochastic differential equations.

A method combining Kalman filtering and the SAEM algorithm has been implemented for mixed effects models governed by linear SDEs.

The convergence of the estimators towards the right parameters has been checked on simulated data.

Also, one of the challenges of our work is to define some selection and evaluation criteria on these models, in order to check, on real data, their pertinence compared to the former models.

Finally, a theoretical study (identifiability of the model, convergence of the maximum likelihood estimators, convergence of the SAEM algorithm) is underway.

Mots-clé : biostatistique, processus, pharmacocinétique, modèles non linéaires à effets mixtes, équations différentielles stochastiques, filtre de Kalman, algorithme SAEM.

On considère un modèle à effets mixtes :

$$y_{ij} = f(t_{ij}, \psi_i) + g(t_{ij}, \psi_i, \xi)\varepsilon_{ij} \quad , \quad 1 \leq i \leq N \quad . \quad 1 \leq j \leq n_i \quad (1)$$

Ici

- $y_{ij} \in \mathbb{R}$ est la j ème observation du sujet i ,
- N est le nombre de sujets,
- n_i est le nombre d'observations du sujet i ,
- les temps d'observations (t_{ij}) sont supposés connus,
- pour le sujet i , $\psi_i = (\psi_{i,\ell}; 1 \leq \ell \leq n_\psi) \in \mathbb{R}^{n_\psi}$ est un vecteur de n_ψ paramètres individuels :

$$\psi_i = h(\mu, c_i, \eta_i) \quad (2)$$

où

- $c_i = (c_{im}; 1 \leq m \leq M)$ est un vecteur de M covariables connues,
- μ est un vecteur d'effets fixes inconnus de taille n_μ ,
- η_i est un vecteur d'effets aléatoires inconnus de taille n_η :

$$\eta_i \sim_{i.i.d.} \mathcal{N}(0, \Omega)$$

- les erreurs résiduelles (ε_{ij}) sont des variables aléatoires gaussiennes *i.i.d.*; de moyenne 0 et de variance 1,

Les modèles utilisés en pharmacologie sont généralement définis par un système dynamique dont la solution n'est pas toujours connue explicitement.

Les modèles de pharmacocinétique, de glucose-insuline ou encore de dynamique virale (VIH, VHC ...) sont ainsi décrits par un système d'EDO :

$$\begin{aligned} dX_i(t) &= H(X_i(t); \psi_i) \\ y_i(t_{ij}) &= C(X_i(t_{ij})) + D(X_i(t_{ij}); \xi)\varepsilon_{ij} \end{aligned}$$

L'algorithme SAEM (Stochastic Approximation EM) couplé avec un algorithme MCMC de type Hasting-Metropolis permet d'estimer $\theta = (\mu, \Omega, \xi)$ au sens du maximum de vraisemblance. Un algorithme de Monte-Carlo par échantillonnage préférentiel permet le calcul de vraisemblance des observations et donc d'obtenir des critères d'information pour la sélection de modèle.

Des résultats de convergence de SAEM vers le maximum de vraisemblance ont été obtenus sous des hypothèses assez générales. Outre ces propriétés théoriques, l'algorithme a démontré de très bonnes propriétés pratiques, en particulier pour des modèles complexes (modèles de glucose-insuline, modèles de dynamique virale, modèle de pharmacocinétique/pharmacodynamie,...). La convergence est rapide et relativement robuste face au choix des valeurs initiales.

Cependant, différents modélisateurs en pharmacologie considèrent que la variabilité intra-sujet n'est pas toujours correctement modélisée par les erreurs résiduelles (ε_{ij}) . Ils proposent de rajouter une composante stochastique dans le modèle dynamique lui-même. En d'autres termes, le système d'EDO est remplacé par un système d'EDS :

$$\begin{aligned} dX_i(t) &= H(X_i(t); \psi_i) + \sigma(\psi_i, \beta)dW_i(t) \\ y_i(t_{ij}) &= C(X_i(t_{ij})) + D(X_i(t_{ij}); \xi)\varepsilon_{ij} \end{aligned}$$

où W_i est un mouvement brownien, β est un nouveau paramètre du modèle. C'est ici que notre travail intervient puisqu'il nous faut trouver des méthodes d'estimation du nouveau paramètre $\theta = (\mu, \Omega, \xi, \beta)$, ainsi que des méthodes d'information pour la sélection et l'évaluation de ces modèles.

Dans cette présentation, nous nous limiterons à des systèmes linéaires :

$$\begin{aligned} dX_i(t) &= H(\psi_i)X_i(t) + \sigma(\psi_i, \beta)dW_i(t) \\ y_i(t_{ij}) &= C X_i(t_{ij}) + \xi\varepsilon_{ij} \end{aligned}$$

L'équation différentielle stochastique linéaire ci-dessus admet une solution explicite :

$$X_i(t) = \exp(Ht)[X_0 + \int_0^t \exp(-Hs)\sigma dW_i(s)]$$

On peut donc exprimer $X_i(t+h)$ en fonction de $X_i(t)$

$$X_i(t+h) = \exp(Hh)X_i(t) + \exp(H(t+h)) \int_t^{t+h} \exp(-Hs)\sigma dW_i(s)$$

On peut donc obtenir, pour chaque individu, un système signal/observation du type :

$$\begin{cases} X_{i,j} = A_{ij}(\phi_i)X_{i,j-1} + B_{i,j}(\phi_i, \beta)W_{i,j} \\ y_{i,j} = C X_{i,j} + \varepsilon_{i,j} \end{cases} \quad (3)$$

où $W_{i,j}$ et $\varepsilon_{i,j}$ sont des variables gaussiennes centrées.

Le filtre de Kalman permet de calculer la vraisemblance conditionnelle $p(y_i|\phi_i; \xi, \beta)$ de y_i pour ce genre de système.

Il peut alors être combiné avec une procédure MCMC, couplée avec l'algorithme SAEM.

En effet, l'algorithme SAEM est un algorithme itératif approchant le maximum de vraisemblance, qui nécessite essentiellement le calcul de la log-vraisemblance complète $\ell(y_i, \psi_i; \theta)$. Nous avons implémenté un code permettant l'estimation de θ dans ce cadre.

Cette approche a été validée sur des données simulées. En effet, les estimateurs convergent vers les valeurs adéquates des paramètres.

L'un des enjeux de notre travail est de définir des critères d'information pour la sélection de ces modèles, afin de vérifier leur pertinence sur des données réelles, par rapport aux anciens modèles.

Enfin, une étude théorique (identifiabilité du modèle, convergence de l'estimateur du maximum de vraisemblance, convergence de l'algorithme SAEM) est en cours pour une validation théorique de la méthode.

Dans l'avenir, nous comptons étendre notre travail à des systèmes non linéaires, en utilisant des méthodes de filtrage particulière pour le calcul de la vraisemblance complète.

Bibliographie

- [1] Delyon B., Lavielle M., and Moulines E. , “Convergence of a stochastic approximation version of the EM algorithm,” The Annals of Stat., vol. 27, no. 1, pp 94-128, 1999.
- [2] Kuhn E., Lavielle M., ”Maximum likelihood estimation in nonlinear mixed effects models” Computational Statistics and Data Analysis, vol. 49, No. 4, pp 1020-1038, 2005.
- [3] Lavielle M., Mentré F. ”Estimation of population pharmacokinetic parameters of saquinavir in HIV patients and covariate analysis with the SAEM algorithm implemented in MONOLIX” Journal of Pharmacokinetics and Pharmacodynamics, vol. 34, n. 2, pp 229-249, 2007.
- [4] Mortensen, S ; Klim, S ; Dammann, B ; Kristensen, N ; Madsen, H ; Overgaard, R ”A Matlab framework for estimation of NLME models using stochastic differential equations”, Journal of Pharmacokinetics and Pharmacodynamics vol : 34, pages : 623-642, 2007.
- [5] Del Moral P. Feynman-Kac formulae. Genealogical and interacting particle approximations Springer New York, Series : Probability and Applications (2004). (Springer-Verlag, Heidelberg)
- [6] Pierre Del Moral, Arnaud Doucet and Sumeetpal S. Singh A Backward Particle Interpretation of Feynman-Kac Formulae HAL-INRIA RR-7019 [21p], [submitted] (July 2009).
- [7] Pierre Del Moral, Arnaud Doucet and Sumeetpal S. Singh Forward Smoothing Using Sequential Monte Carlo Technical Report 638. Cambridge University Engineering Department. [25p], [submitted] (2009).