

Estimation paramétrique dans des modèles définis par un système d'équations différentielles

Application aux modèles non-linéaires à effets mixtes

Sophie Donnet ¹

En collaboration avec Adeline Samson ²

¹Université Paris-Dauphine, Cérémade

²Université Paris-Descartes, MAP5

Applibugs, Juin 2008

Contexte

- Description d'un phénomène physiologique/biologique
 - *En pharmacocinétique-pharmacodynamie*
 - *En dynamique du VIH/VHC/VHB*
 - *En Imagerie par résonance magnétique*

⇒ Étude de modèles compartimentaux ou physiologiques
- Modélisation à l'aide d'un système d'équations différentielles ordinaires (EDO)
- La plupart de ces EDO n'ont pas de solution analytique

Modèles non-linéaires à effets mixtes : objectifs

- Utilisés quand on a des mesures répétées au cours du temps d'un critère biologique (concentration en médicament, en virus) sur une population de sujets.
- Même fonction de régression pour tous les sujets mais paramètres de régression propres à chaque individu.
- Permettent de différencier la variabilité inter-sujets de la variabilité résiduelle

Modèles non-linéaires à effets mixtes : écriture statistique

$$\begin{cases} y_{ij} = f(\phi_i, t_{ij}) + \varepsilon_{ij}, & i = 1 \dots l, j = 1 \dots n_i \\ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ \phi_i \sim \mathcal{N}(\mu, \Omega) \end{cases}$$

- y_{ij} observation du sujet i au temps t_{ij}
- ε_{ij} erreur au temps t_{ij} . Bruit gaussien de variance résiduelle σ^2
- ϕ_i paramètre aléatoire individuel du sujet i
- μ paramètre de moyenne
- Ω variabilité inter-sujets

A propos de la fonction de régression

- f décrivant le processus biologique \Rightarrow solution d'une ODE

$$\begin{cases} \frac{\partial f(t, \phi)}{\partial t} = F(f(t, \phi), t, \phi), & t \in [t_0, T] \\ f(t_0, \phi) = f_0(\phi) \end{cases}$$

- Non-linéaire en le paramètre ϕ ,

On cherche à estimer les paramètres $\theta = (\mu, \Omega, \sigma^2)$ par maximum de vraisemblance ou dans un cadre bayésien.

(pour faire des tests d'hypothèses, comparer des médicaments par ex...)

Exemple : modèle de cinétique enzymatique

- Modèle de Michaelis-Menten à un compartiment avec absorption du premier ordre : équations issues d'équations bilans de réactions chimiques

•

$$\frac{dC}{dt}(t, \phi) = \frac{k_a \cdot Dose}{V} e^{-k_a t} - \frac{V_m \cdot C(t, \phi)}{k_m + C(t, \phi)} \quad (1)$$

- $C(t, \phi)$ (g/L) concentration du médicament
- $Dose$ quantité connue fixée
- $\phi = (k_m, V_m, k_a, V)$
 - k_m (g/L) constante d'élimination
 - V_m (g/L/h) volume maximal d'élimination
 - k_a (h⁻¹) constante d'absorption de la molécule
 - V (L) volume total de distribution de la molécule dans le corps

Méthodes d'estimation des paramètres

- **But** : Estimer $\theta = (\mu, \Omega, \sigma^2)$
- **Cadre de travail** : modèle à données non-observées $((\phi_i)_{i=1\dots I})$
- *Par maximum de vraisemblance*
 - Algorithme Expectation-Maximization (Dempstair, Laird & Rubin, 1977) particulièrement bien adapté à ce type de modèles.
 - Ici modèle non-linéaire : utilisation d'une version stochastique de l'algorithme EM couplée avec un algorithme Monte Carlo par Chaîne de Markov (MCMC) (Kuhn et Lavielle, 2004)
- *Par approche bayésienne*
 - Loi *a priori* p_θ sur θ
 - Estimation de la distribution *a posteriori* de θ , $p_{\theta|y}$, par algorithme de Gibbs

Difficultés

- Utilisation, dans toutes les méthodes d'estimation, d'un algorithme MCMC
 - ⇒ Calcul de la quantité $f(t, \phi)$ à chaque itération de l'algorithme MCMC
 - MAIS, f est solution non explicite d'une EDO.
- Utilisation d'une méthode numérique de résolution d'EDO pour évaluer f .
- Erreurs d'estimation engendrées par l'utilisation d'une méthode d'approximation numérique de la solution de l'EDO
 - ⇒ Contrôle des erreurs d'estimation

Rappels sur l'EM et autres...

Méthodes adaptées aux modèles à données manquantes

- Maximum de vraisemblance : Expectation-Maximization
(Dempstair, Laird & Rubin, 1977)
- Cadre bayésien : Algorithme du type Gibbs ou Gibbs hybride

L'algorithme EM

- Conditions d'application
 - S'applique lorsque la vraisemblance $p_y(y; \theta)$ n'est pas maximisable de façon explicite
 - Profite de la structure des données non-observées.
 - Applicable si la **vraisemblance complète** $(p_{y,\phi}(y, \phi; \theta))$ a une expression explicite.
- Notons $Q(\theta|\theta') = \int \log p_{y,\phi}(y, \phi; \theta) p_{\phi|y}(\phi; \theta') \mu(d\phi)$.
- Principe de l'Algorithme EM : à l'itération (k)
 - Etape E : Calcul de $Q(\theta|\theta^{(k)})$
 - Etape M : Maximisation de $\theta \rightarrow Q(\theta|\theta^{(k)})$
- L'alternance des étapes E et M entraîne une croissance de la vraisemblance.

L'algorithme EM dans le cas d'un modèle exponentiel

$$p(y, \phi, \theta) = \exp \{ -\psi(\theta) + \langle S(y, \phi), \Phi(\theta) \rangle \}$$

$S(y, \phi)$ statistique exhaustive.

Algorithme EM

A l'itération k ,

- Étape E : Calcul de $s^{(k+1)} = \mathbb{E}_{\phi|y} [S(y, \phi) | \theta^{(k)}]$
- Étape M : réactualisation de $\theta^{(k)}$ par le maximum

$$\theta^{(k+1)} = \arg \max_{\theta} \left(-\Psi(\theta) + \langle s^{(k+1)}, \Phi(\theta) \rangle \right)$$

Convergence vers un maximum de la vraisemblance sous des hypothèses simples dans le cas d'un modèle exponentiel [Delyon & al (1999)] .

Algorithmes SAEM Delyon, Lavielle et Moulines, 1999

Si on ne sait pas calculer l'espérance conditionnelle :

$$\mathbb{E}_{\phi|y} [S(y, \phi) | \theta^{(k)}]$$

SAEM

- *Étape E*
 - *Étape S* : simulation de $\phi^{(k)}$ sous la loi $p(\phi|y; \theta^{(k)})$
 - *Étape SA* : approximation stochastique de s_k avec (γ_k) suite de pas décroissante vers 0

$$s^{(k+1)} = s^{(k)} + \gamma_k \{S(y, \phi^{(k)}) - s^{(k)}\}$$

- *Étape M* : réactualisation de $\theta^{(k)}$ par

$$\theta^{(k+1)} = \arg \max_{\theta} \left(-\Psi(\theta) + \langle s^{(k+1)}, \Phi(\theta) \rangle \right)$$

Algorithmes SAEM-MCMC

Kuhn et Lavielle, 2004

Si on ne sait pas simuler sous la loi a posteriori $p(\phi|y; \theta^{(k)})$

SAEM-MCMC

- *Étape S* : procédure MCMC construisant à l'étape k une chaîne de Markov de loi stationnaire $p(\phi|y; \theta^{(k)})$

Les autres étapes restent inchangées

L'algorithme Hastings-Metropolis ($H-M$)

- A chaque itération de l'algorithme SAEM-MCMC, à θ fixé
- Génère une chaîne de Markov de loi invariante $p(\phi|y; \theta)$, stationnaire sous certaines conditions

Algorithme de Hastings-Metropolis

A l'itération ℓ

1. Génération d'un candidat ϕ^c avec le noyau $q_{\theta}(\cdot|\phi^{(\ell)})$
2. Actualisation

$$\phi^{(\ell+1)} = \left\{ \begin{array}{ll} \phi^c & \text{avec probabilité } \rho(\phi^{(\ell)}, \phi^c) \\ \phi^{(\ell)} & \text{avec probabilité } 1 - \rho(\phi^{(\ell)}, \phi^c) \end{array} \right\}$$

$$\text{où } \rho(\phi^{(\ell)}, \phi^c) = \min \left\{ \frac{p(\phi^c|y; \theta) q_{\theta}(\phi^{(\ell)}|\phi^c)}{p(\phi^{(\ell)}|y; \theta) q_{\theta}(\phi^c|\phi^{(\ell)})}, 1 \right\}$$

ρ est la probabilité d'acceptation

Choix des noyaux

Théoriquement, universel en terme de noyau mais le choix du noyau conditionne les propriétés de convergence

- Noyau K_1 : la distribution a priori $q_\theta(\phi^c | \phi^{(\ell)}) = \pi(\phi^c; \theta)$
 - Algorithme de Hastings-Metropolis *indépendant*
 - Probabilité d'acceptation

$$\rho(\phi^{(\ell)}, \phi^c) = \min \left\{ \frac{p(y|\phi^c; \theta)}{p(y|\phi^{(\ell)}; \theta)}, 1 \right\}$$

- Or $p(y|\phi^c; \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left[-\frac{\sum_{i,j} (y_{ij} - f(t_{ij}, \phi_i^c))^2}{2\sigma^2} \right]$
- ⇒ Evaluation de $f(t, \phi^c)$ à chaque itération

- Noyau K_2 : marche aléatoire $\phi^c = \phi^{(\ell)} + \delta$
 - Noyau symétrique : $q_\theta(\phi' | \phi) = q_\theta(\phi | \phi')$

⇒ Ici : Evaluation de $f(t, \phi^c)$ connaissant $f(t, \phi^{(\ell)})$, avec ϕ^c dans un petit voisinage de $\phi^{(\ell)}$

Estimation bayésienne

- Loi *a priori* p_θ sur θ
- Algorithme de Gibbs/ Hastings-Metropolis avec principe de "data augmentation" : on veut échantillonner la loi $p(\theta, \phi|y)$

Algorithme de Gibbs

A l'itération k

1. Simulation de $\phi^{(k)}$ sous la loi instrumentale $q_1(\cdot|\phi^{(k-1)}, \theta^{(k-1)}, y)$ avec une proba d'acceptation ρ_1
2. Simulation de $\theta^{(k)}$ sous la loi instrumentale $q_2(\cdot|\theta^{(k-1)}, \phi^{(k)}, y)$ avec une proba d'acceptation ρ_2

- Convergence de l'algorithme vers la distribution d'intérêt $p_{\theta, \phi|y}(\theta, \phi|y)$ (la vitesse de convergence dépend de q_1 et q_2)

⇒ Evaluation de $f(t, \phi^{(k)})$ à chaque itération

Introduction d'un modèle statistique "approché"

- Les méthodes d'estimation présentées nécessitent l'évaluation de la fonction de régression f .
- Or dans notre cas, f est solution non-explicite d'une EDO.
- On peut seulement l'approcher par une méthode numérique de résolution d'EDO : soit f_h , la solution de l'EDO approchée par de f une méthode numérique

$$(\mathcal{M}) \begin{cases} y_{ij} = f(\phi_i, t_{ij}) + \varepsilon_{ij} \\ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ \phi_i \sim \mathcal{N}(\mu, \Omega) \end{cases} \Rightarrow (\mathcal{M}_h) \begin{cases} y_{ij} = f_h(\phi_i, t_{ij}) + \varepsilon_{ij} \\ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ \phi_i \sim \mathcal{N}(\mu, \Omega) \end{cases}$$

On travaille sur ce nouveau modèle \mathcal{M}_h (application des algorithmes précédents (SAEM-MCMC et Gibbs)) et on indice par h toutes les distributions se rattachant à ce modèle

Méthodes de résolution classiques d'EDO

- Une méthode numérique est caractérisée par son *pas* h et son *ordre* p
 - $[t_0, T]$ divisé en intervalles $[t_n, t_{n+1}[$ de longueur h
 - $\sup_{t, \phi} |f(t, \phi) - f_h(t, \phi)| = O(h^p)$
- Méthodes existantes :
 - Méthodes à pas séparé (Euler, Runge-Kutta) / à pas multiples (Adams-Bashforth, Moulton) OU PECE (prédiction-Evaluation-Correction-Evaluation)
 - Méthodes explicite / implicite (systèmes Stiffs)
 - Fonctions Matlab
 - ode45 : Runge-Kutta d'ordre 4
 - ode23 : Runge-Kutta d'ordre 3
 - ode113 : PECE (Adams-Bashforth et Adams-Moulton)
 - ode23s : Runge-Kutta implicite pour systèmes "Stiff"

On cherche un compromis entre
précision, stabilité et temps de calculs

Schéma de Linéarisation Locale (LL)

- Principe :
 - Linéarisation locale par rapport au temps de l'EDO
 - Intégration exacte de l'EDO linéaire obtenue
- Résultat de convergence (Ramos et Garcia-López, 1997)
 - Schéma d'ordre 2
- Inconvénients
 - Nécessite le calcul d'exponentielles de matrices
 - Nouvelles méthodes de Padé, Schur, ...
- Avantages
 - Grande stabilité dans les systèmes de grande dimension et/ou "Stiff"
 - Généralisation aux équations différentielles stochastiques avec bruit additif

Avantage : une généralisation du schéma LL

- Contexte

- Algorithme $H-M$ par marche aléatoire
"Evaluation de $f(t, \phi^c)$ connaissant $f(t, \phi^{(\ell)})$ avec ϕ^c dans un petit voisinage de $\phi^{(\ell)}$ "

- Principe

- Développement de Taylor en (t, ϕ) de l'EDO
- Pas de nouveau calcul d'exponentielle de matrice

- Résultat de convergence

Il existe C_1 et C_2 telles que

$$\sup_{t \in [t_0, T]} |f(t, \phi^c) - f_{h, \phi^{(\ell)}}(t, \phi^c)| \leq \max(C_1 h^2, C_2 \left\| \phi^c - \phi^{(\ell)} \right\|_{\mathbb{R}^k}^2)$$

Un algorithme Hastings-Metropolis modifié

Modification de l'algorithme *H-M* en utilisant LL2

Hastings-Metropolis modifié

A l'itération ℓ

1. Génération d'un candidat ϕ^c par $\phi^c = \phi^{(\ell)} + \delta$
2. Actualisation

$$\phi^{(\ell+1)} = \begin{cases} \phi^c & \text{avec probabilité } \rho^{(2)}(\phi^{(\ell)}, \phi^c) \\ \phi^{(\ell)} & \text{avec probabilité } 1 - \rho^{(2)}(\phi^{(\ell)}, \phi^c) \end{cases}$$

$$\text{où } \rho^{(2)}(\phi^{(\ell)}, \phi^c) = \min \left\{ \frac{p^{LL2}(y|\phi^c; \theta)\pi(\phi^c; \beta)}{p^{LL}(y|\phi^{(\ell)}; \theta)\pi(\phi^{(\ell)}; \beta)}, 1 \right\}$$

3. Si acceptation du candidat, résolution de l'EDO avec le schéma LL

⇒ Si on rejette le candidat, on ne calcule pas d'exponentielle de matrice

Propriété de l'algorithme SAEM-MCMC

Résultat 1

Soit (γ_k) tels que $\sum_k \gamma_k = \infty$ et $\sum_k \gamma_k^2 < \infty$.

En considérant une méthode numérique de pas h et d'ordre p , sous les hypothèses \mathcal{H}_1 et \mathcal{H}_2

1. Supposons que $(s_k)_{k \geq 0}$ appartienne à un compact, alors la suite (θ_k) d'estimateurs fournie par l'algorithme SAEM sur le modèle \mathcal{M}_h converge presque sûrement vers un point θ_∞ , maximum (local) de la fonction de vraisemblance $p_{h,y}$
2. $\sup_{\theta \in \Theta} |p_y(y; \theta) - p_{h,y}(y; \theta)| = O(h^p)$

Propriétés de l'algorithme de Gibbs (bayésien)

Résultat 2

Considérons une méthode numérique de pas h et d'ordre p Soit p_θ une loi a priori sur θ . Sous les hypothèses \mathcal{H}_1 et \mathcal{H}_2 , il existe une constante C_y telle que

$$D(p_{\theta|y}(\cdot), p_{h,\theta|y}(\cdot)) \leq C_y h^p$$

Applications sur données simulées

Modèle de Michaelis-Menten :

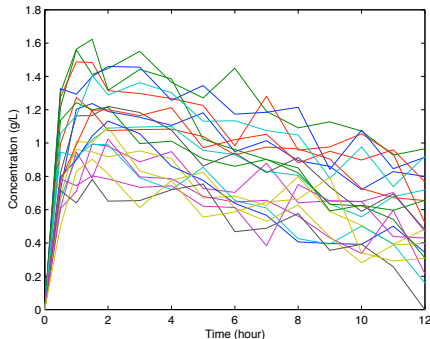
$$\frac{dC}{dt}(t, \phi) = \frac{k_a \cdot Dose}{V} e^{-k_a t} - \frac{V_m \cdot C(t, \phi)}{k_m + C(t, \phi)} \quad (2)$$

- $C(t, \phi)$ (g/L) concentration du médicament
- $Dose$ quantité connue fixée
- $\phi = (k_m, V_m, k_a, V)$

Jeu de données simulées

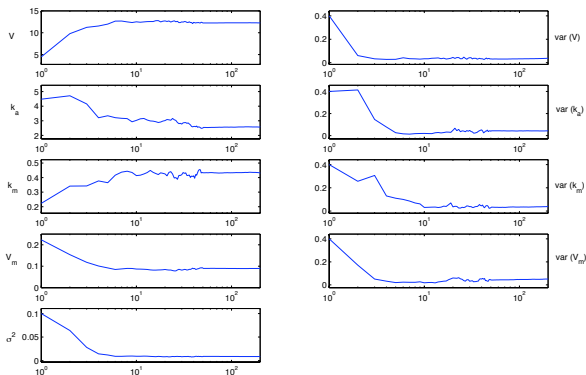
- $N=20$ patients indépendants. Observations à $t = 0, 0.5, 1, 1.5, 2$ heures puis toutes les 2 heures
- $\phi = [V, k_a, k_m, V_m] \in \mathbb{R}^4$. $\phi_i \sim \mathcal{N}(\mu, \Omega)$

$$\mu = (12.2, 2.72, 0.37, 0.082) \quad \Omega = 0.4I_4$$
- $D = 13.8$ mmol. $\sigma^2 = 0.01$



Estimation par SAEM-MCMC

- 200 itérations SAEM, 20 itérations de l'algo H-M pour chaque noyau.
- Résolution de l'EDO par LL et LL2.
- Temps de calcul env. 10 min
- Exemple de convergence :



Estimation Bayésienne

(en collaboration avec Jean-Louis Foulley)

- Loi a priori sur Ω :
 - Inverse Gamma : résultats de convergence inacceptables
 - Utilisation d'une Inverse Wishart
- Utilisation de WBDiff (WinBUGS with Differential Interface) :
100000 itérations, 1 chaîne
- Temps de calculs : 1h

Résultats d'estimation

- NONMEM : logiciel d'estimation des modèles non linéaires mixtes en PK/PD (méthode d'estimation par linéarisation de la fonction de régression)
- Comparaison de SAEM et Bayésien avec NONMEM
 - NONMEM ne converge pas
 - SAEM et Bayésien estiment correctement les paramètres

	V	k_a	k_m	V_m	var V	var k_a	var k_m	var V_m	σ^2
initial values	5.0	5.00	0.50	0.100	0.400	0.400	0.400	0.400	0.1000
simulation values	12.2	2.72	0.37	0.082	0.040	0.040	0.040	0.040	0.0100
SAEM									
estimates	12.3	2.79	0.40	0.085	0.038	0.049	0.039	0.038	0.0081
SE	0.5	0.21	0.01	0.004	0.012	0.019	0.013	0.013	6.10^{-4}
NONMEM									
estimates	12.3	2.57	0.60	0.100	0.036	0.068	10^{-8}	0.062	0.0088
SE	-	-	-	-	-	-	-	-	-
Bayesian Gibbs									
mode estimates	12.4	2.64	0.39	0.085	0.040	0.040	0.048	0.052	0.0082
SE	0.6	0.23	0.62	0.044	0.014	0.046	0.028	0.029	7.58^{-4}

Perturbation de l'ODE

$$\begin{cases} y_{ij} = Z(\phi_i, t_{ij}) + \varepsilon_{ij} \\ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \\ \phi_i \sim \mathcal{N}(\mu, \Omega) \end{cases}$$

- La fonction de régression Z du modèle est solution de

$$\begin{aligned} dZ(t, \phi) &= F(Z(t, \phi), t, \phi)dt + \gamma G(Z(t, \phi), t, \phi)dW_t \\ Z(t_0, \phi) &= Z_0(\phi) \end{aligned}$$

où W_t est un mouvement brownien

- Estimation de $\theta \in \Theta = \{(\mu, \Omega, \sigma^2, \gamma^2) \in \mathbb{R}^{p+2}\}$

Estimation par maximum de vraisemblance

- Modèle à données non-observées : $(\phi_i, (z(t_{ij})))$
- Vraisemblance complète du modèle

$$p_{y,\phi,z}(y, z, \phi; \theta) = p_{y|z}(y|z; \sigma^2) \prod_i \pi(\phi_i; \mu, \Omega) \prod_{i,j} p_{z|\phi}(z(t_{ij})|z(t_{i,j-1}), \phi_i; \gamma^2)$$

- N'a pas d'expression explicite en général. Euler-Maruyama en donne une **approximation gaussienne**.

Estimation par maximum de vraisemblance

- Difficultés :
 - Comme précédemment, approximation de Z par un schéma numérique (Euler-Maruyama)
 - Introduction de temps intermédiaires \Rightarrow augmentation du volume de données manquantes
 - Difficultés d'implémentation d'un algorithme de Gibbs : utilisation de ponts browniens
- Résultats :
 - Convergence des algorithmes
 - Contrôle de l'erreur engendrée par l'utilisation du schéma d'Euler-Maruyama
 - Mise en évidence de la pertinence de l'approche par SDE versus ODE sur données réelles

Estimation bayésienne

Travail en cours avec Adeline Samson, Jean-Louis Foulley et Florence Jaffrézic

- Modèle de croissance de Gompertz.
- Choix du type de bruit ? Multiplicatif ou additif
- Existence de solution exacte de la SDE dans certains cas \Rightarrow écriture facile d'un algorithme de Gibbs
- Utilisation de Winbugs : une des variances s'exprime comme l'intégrale d'une fonction sans primitive évidente \Rightarrow question par la séance suivante

Perspectives

- Certaines modèles peuvent-être vu comme des ARMA \Rightarrow utilisation de la littérature dans ce domaine
- Certaines modèles peuvent-être vu comme chaînes de Markov cachées \Rightarrow idem
- EDS sur des systèmes en grande dimension ?

Références

- DEMPSTER, LAIRD, RUBIN (1977). Maximum likelihood from incomplete data via the EM algorithm. *JRSS B*, 39, 1-38
- DELYON, LAVIELLE, MOULINES (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27, 94-128.
- KUHN and LAVIELLE (2004). Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM PS*, 8, 115-131.
- DONNET, SAMSON (2007). Estimation of parameters in missing data models defined by differential equations. *J. Statist. Plann. Inference*, 137, 2815-2831.
- DONNET, SAMSON (2008). Parametric inference for mixed models defined by stochastic differential equations, *ESAIM P&S*, 27 :196-218