Parametric estimation of complex mixed models based on meta-model approach

Pierre BARBILLON[1], Célia BARTHÉLÉMY[2], Adeline LECLERCQ-SAMSON[3]

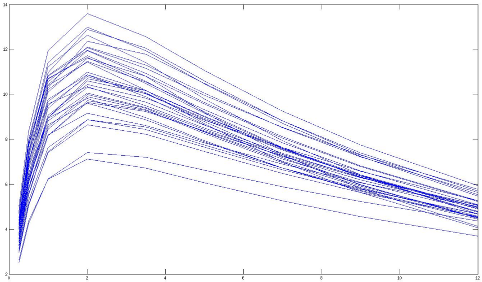
[1]AgroParisTech / INRA MIA UMR 518
pierre.barbillon@agroparistech.fr

[2]INRIA, POPIX team / Université Paris-Sud

[3]Laboratoire Jean Kuntzmann, Université Joseph Fourier, Grenoble

Modélisation en biologie et statistique de données biomédicales
Grenoble 02/11/2015

Context: Pharmacokinetics



Models:

- compartment models for diffusion of a drug, ...
- modeled by ODE, PDE

Population approach:

- Parameters in this differential equation depend on individuals.
- distribution has to be estimated on the basis of longitudinal data.

Outline

Outline

Outline

Model

$\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^t$ where $y_{ij} \in \mathbb{R}^p$ is the response for individual i at time t_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$.

For $i = 1, \ldots, N$, $j = 1, \ldots, n_i$:

$$\begin{aligned} y_{ij} &= f(t_{ij}, \psi_i) + \sigma_\varepsilon \, \varepsilon_{ij}, \quad \varepsilon_{ij} \sim_{iid} \mathcal{N}(0, 1) \\ \psi_i &\sim_{iid} \mathcal{N}(\mu, \Omega), \end{aligned}$$

where

- $f(\cdot, \cdot) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^p$ is the regression function,
- ψ_i vector of individual parameters,
- $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^t$ represents the Gaussian centered residual error, independent of ψ_i.

Goal: estimate from observation of **y**, the population parameters $\theta = (\mu, \Omega, \sigma_\varepsilon^2)$.

Difficulty: f may be solution of complex ODE, PDE without analytical expression, achievable by numerical solver, expensive to compute.

Likelihood

Estimate θ by maximum likelihood procedure.

The likelihood of the mixed model is the following

$$p(\mathbf{y},\theta) = \int p(\mathbf{y},\boldsymbol{\psi}\,;\theta)\,d\boldsymbol{\psi} = \prod_{i=1}^{N}\int p(\mathbf{y}_i|\psi_i;\theta)p(\psi_i;\theta)d\psi_i$$

$$= \prod_{i=1}^{N}\int \frac{1}{(2\pi\sigma_\varepsilon^2)^{n_i/2}} \exp\left(-\frac{1}{2}\,{}^t(\mathbf{y}_i - f(\mathbf{t}_i,\psi_i))(\sigma_\varepsilon^2 I_{n_i})^{-1}(\mathbf{y}_i - f(\mathbf{t}_i,\psi_i))\right)$$
$$\times \frac{1}{\sqrt{(2\pi)^N|\Omega|}} \exp\left(-\frac{1}{2}\,{}^t(\psi_i-\mu)\Omega^{-1}(\psi_i-\mu)\right)d\psi_i$$

Outline

EM principle

[Dempster and al., 1977]

- Observed log-likelihood: $L(\mathbf{y}, \theta) = \log p(\mathbf{y}, \theta)$,
- Complete log-likelihood: $L(\mathbf{y}, \psi\,; \theta) = \log p(\mathbf{y}, \psi\,; \theta)$

EM decomposition: $L(\mathbf{y}; \theta) = \underbrace{\mathbb{E}(L(\mathbf{y}, \psi; \theta)|\mathbf{y}; \theta^{(c)})}_{Q(\theta, \theta^{(c)})} - \mathbb{E}(\log(p(\psi|\mathbf{y}; \theta))|\mathbf{y}; \theta^{(c)})$.

For fixed $\theta^{(c)}$, if $Q(\theta, \theta^{(c)})$ increases, then $L(\mathbf{y}; \theta)$ increases.

EM

Random initialisation: $\theta^{(0)}$.
For iteration k,

1. **Expectation** Compute $Q(\theta, \theta^{(k)})$,
2. **Maximisation** Update $\theta^{(k+1)} = \text{argmax}_\theta Q(\theta, \theta^{(k)})$

Stochastic version of EM

If Expectation step not possible simulate ψ according to $p(\cdot|\mathbf{y};\theta^{(c)})$:

- SEM [Celeux and Diebolt, 1985],
- MCEM [Wei and Tanner, 1990],
- SAEM [Delyon and al.(1999)].

Stochastic Approximation of EM: SAEM

SAEM algorithm

Random initialisation: $\theta^{(0)}$.
For iteration k,

1. **Simulation step:** Simulate $\psi^{(k)}$ according to $p(\cdot|\mathbf{y};\theta^{(k-1)})$,
2. **Stochastic Approximation step:** update the sufficient statistics S_D

$$\begin{aligned} s_{k,1} &= s_{k-1,1} + \gamma_k \left(\sum_{i=1}^{N} \psi_i^{(k)} - s_{k-1,1} \right) \\ s_{k,2} &= s_{k-1,2} + \gamma_k \left(\sum_{i=1}^{N} \psi_i^{(k)}\, {}^t\psi_i{}^{(k)} - s_{k-1,2} \right) \\ s_{k,3} &= s_{k-1,3} + \gamma_k \left(\sum_{i=1}^{N} \sum_{j=1}^{n_i} (y_{ij} - f(t_{ij}, \psi_i^{(k)}))^2 - s_{k-1,3} \right) \end{aligned}$$

3. **Maximisation step:** update the parameters

$$\begin{aligned} \widehat{\mu}^{(k)} &= \frac{s_{k,1}}{N}, \quad \widehat{\Omega}^{(k)} = \frac{s_{k,2}}{N} - \frac{s_{k,1}\, {}^t s_{k,1}}{N^2} \\ \widehat{\sigma_\varepsilon}^{2\,(k)} &= \frac{s_{k,3}}{n_{tot}} \end{aligned}$$

MH algorithm:

Coupling the SAEM algorithm with a MCMC procedure for step S: [Kuhn and Lavielle, 2004]

Simulation step: For each individual i separately and successively, update $\psi_i^{(k)}$ with m iterations of a MH algorithm with $p(\psi_i|\mathbf{y}_i;\theta^{(k-1)})$ as stationary distribution.

Drawback: Each computation of the acceptation rate needs a resolution of the ODE /PDE to compute f.

Outline

Outline

Gaussian Process Emulator

Sacks et al. (1989).

- Assumption: f realization of a Gaussian process F:
  $\forall \mathbf{x} = (t, \psi) \in E$,

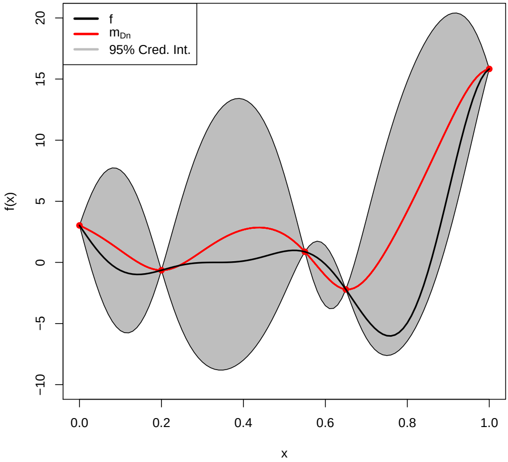
$$F(\mathbf{x}) = \sum_{k=1}^{Q} \beta_k h_k(\mathbf{x}) + \zeta(\mathbf{x}) = H(\mathbf{x})^T \boldsymbol{\beta} + \zeta(\mathbf{x}) .$$

- Pre-computation step: $y_1 = f(\mathbf{x}_1), \ldots, y_n = f(\mathbf{x}_n)$ evaluations of f on a design D.

- Process F^D: Conditioning F to $F(\mathbf{x}_1) = y_1, \ldots, F(\mathbf{x}_n) = y_n$.
  Gaussian Process with mean $m_D(\mathbf{x})$ and covariance $C_D(\mathbf{x}, \mathbf{x}')$ $\forall \mathbf{x}, \mathbf{x}'$.

  For all $\mathbf{x} \in E$,
  - $m_D(\mathbf{x})$ approximates $f(\mathbf{x})$,
  - $C_D(\mathbf{x}, \mathbf{x})$ uncertainty on this approximation.

Gaussian process emulator: illustration



Outline

Mixed meta-model

f is replaced with F^D:

$$\begin{array}{rcl} y_{ij} & = & \underbrace{F^D(t_{ij},\psi_i)}_{m_D(t_{ij},\psi_i)+r(t_{ij},\psi_i)} + \sigma_\varepsilon\, \varepsilon_{ij}, \quad \varepsilon_{ij} \sim_{iid} \mathcal{N}(0,1) \\ \psi_i & \sim_{iid} & \mathcal{N}(\mu,\Omega) \\ r(t,\psi) & = & F^D(t,\psi) - m_D(t,\psi) \sim \mathcal{GP}(0, C_D(t,\psi;t,\psi)), \end{array}$$

$\Rightarrow r()$ takes into account the approximation error but makes the y_{ij} not independent.

3 situations:

1. Complete mixed meta-model, keeping r as it is,
2. Simple mixed meta-model, neglecting r (replacing f with m_D),
3. Intermediate mixed meta-model, replacing r with $\bar{r}$ where independence is forced by setting to 0 the correlations in the $\mathcal{GP}$.

Likelihood

$$
\begin{aligned}
p_D(\mathbf{y};\boldsymbol{\theta}) &= \int p(\psi;\boldsymbol{\theta}) p_D(\mathbf{y}|\psi;\boldsymbol{\theta}) d\psi\,, \\
&= \int p(\psi;\boldsymbol{\theta}) \, \frac{1}{(2\pi)^{1/2} |\sigma_\varepsilon^2 I_{n_{tot}} + \mathbf{C}_D(\mathbf{t},\psi)|^{1/2}} \\
& \exp\left(-\frac{1}{2} {}^t(\mathbf{y} - \mathbf{m}_D(\mathbf{t},\psi))(\sigma_\varepsilon^2 I_{n_{tot}} + \mathbf{C}_D(\mathbf{t},\psi))^{-1}(\mathbf{y} - \mathbf{m}_D(\mathbf{t},\psi)) \right) d\psi.
\end{aligned}
$$

- Likelihood not explicit, because $m_D(t_{ij},\psi_i)$ not linear in ψ_i,
- Likelihood cannot be simplified as a product of individual likelihoods because $\mathbf{y}_i$ not independent (matrix $\mathbf{C}_D(\mathbf{t},\psi)$ is a full matrix),
- computational burden to invert $\mathbf{C}_D(\mathbf{t},\psi)$ at each iteration of the MCMC algorithm.

Estimation issues

1. Complete mixed meta-model
   - takes into account uncertainty due to meta-modeling,
   - computational burden due to inversion of the covariance matrix in the MH algorithm,
   - dependence between individuals $\Rightarrow$ bad mixing properties of the MH algorithm.
2. Simple mixed meta-model
   - does not take into account uncertainty due to meta-modeling,
   - computational efficient since the likelihood is decomposable as a product of individual likelihoods (as the exact mixed model).
3. Intermediate mixed-model
   - takes into account uncertainty due to meta-modeling,
   - neglects dependence between GP emulator approximation errors may biased variance estimates,
   - computational efficient since the likelihood is decomposable as a product of individual likelihoods.

Outline

Convergence to MLE

Proposition

Under general condition as in [Kuhn and Lavielle, 2004]:
for the complete, intermediate or simple mixed meta-model, if the sequence (s_k) stays in a compact set, the SAEM algorithm produces a sequence $(\hat{\theta}^{(k)})_{k\geq 1}$ which converges to the (local) maximum of the corresponding approximated likelihood.

Distance between likelihoods

Proposition

- $p(\mathbf{y};\theta)$ likelihood of the exact mixed model, $\tilde{p}_D(\mathbf{y};\theta)$ likelihood of a mixed meta-model where D is a minimax design.
- The support of the distribution of ψ is compact.
- The functions f and m_D are uniformly bounded on the support of the distribution of ψ.

Then, there exists a constant $\tilde{C}_y$ which depends only on $\mathbf{y}$ such that

$$|p(\mathbf{y};\theta) - \tilde{p}_D(\mathbf{y};\theta)| \leq \tilde{C}_y \frac{n_{tot}}{\sigma_\varepsilon^{n_{tot}+2}} G_K(a_D)$$

where the function $G_K(a)$ tends to 0 when $a \to 0$ and the constant a_D is the covering distance of the design of experiments D.

With regularity hypotheses, results similar to [Donnet and Samson(2007)]: distance between $p(\mathbf{y};\theta)$ and $\tilde{p}_D(\mathbf{y};\theta)$ can be as small as we want for D rich enough.

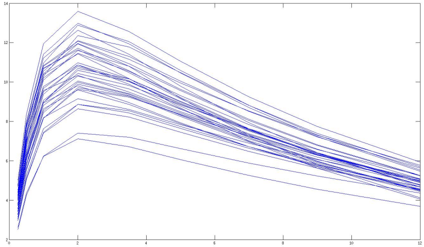
Outline

Simulation model

- one-compartment pharmacokinetic model, first order absorption and elimination.
- at time 0, a dose D of a drug is given to patient
- drug concentration described by equation:

$$\frac{dC}{dt} = D\frac{k_a k_e}{C_l} exp(-k_a t) - k_e C, \quad C(t_0) = 0$$

where k_a and k_e are the absorption and elimination constants, C_l is the clearance.



Simulation model

Parameters

- pharmacokinetic parameters for the theophyllin: $\log k_e = -2.52$, $\log k_a = 0.4$, $\log C_k = -3.22$.
- dataset of 36 patients is simulated with a dose D=6 *mmol* and measurements at time $t = 0.25, 0.5, 1, 2, 3.5, 5, 7, 9, 12$ hours.
- random effects were simulated assuming a diagonal variance-covariance matrix Ω with the following diagonal elements: $\omega_{k_e} = \omega_{k_a} = \omega_{C_l} = 0.1$.
- Then a homoscedastic additive error model is simulated with a standard error $\sigma_\varepsilon = 0.1$.

SAEM settings

- 3 SAEM algorithms: exact, intermediate, simple.
- 100 iterations of SAEM with 15 iterations of MCMC at each SAEM S step,
- $n_D = 50$ and $n_D = 100$ tested, a meta-model computed for each time t. Original domain set to $[-4; -1] \times [0; 2] \times [-4.5; 2]$
- Meta-models with linear regression function and Gaussian covariance matrix.

Results

Parameter		Intermediate meta-model		Simple meta-model		Exact model
	n_D	50	100	50	100	
$\mu_{\log k_e}$	Bias	0.101	0.007	-0.320	0.007	0.003
	RMSE	0.004	0.005	0.005	0.005	0.005
	Cov.	94.2	94.4	90.6	94.6	93.9
$\mu_{\log k_a}$	Bias	-2.441	0.001	-8.380	0.008	-0.220
	RMSE	0.222	0.162	0.910	0.160	0.160
	Cov.	90.9	95.6	59.6	95.3	95.6
$\mu_{\log C_l}$	Bias	0.388	0.036	0.160	0.036	-0.004
	RMSE	0.004	0.003	0.003	0.003	0.003
	Cov.	87.6	95.1	93.4	94.7	94.9
$\omega^2_{\log k_e}$	Bias	-12.113	-2.745	-23.200	-2.780	-3.400
	RMSE	7.131	6.404	9.730	6.530	6.460
	Cov.	83.2	91.5	65.7	90.5	90.3
$\omega^2_{\log k_a}$	Bias	-20.485	-3.442	20.900	-3.320	-2.440
	RMSE	10.696	5.911	13.500	5.930	6.050
	Cov.	72.3	89.7	96.9	89.2	90.2
$\omega^2_{\log C_l}$	Bias	0.375	-1.145	-8.100	-1.100	-2.660
	RMSE	5.944	5.726	5.810	5.690	5.650
	Cov.	92.6	92.0	87.5	92.8	91.1
σ^2_ϵ	Biais	-45.262	-0.612	16.000	-0.009	-0.023
	RMSE	20.719	0.232	2.950	0.220	0.220

One compartment simulations: relative bias (%), relative MSE (%) and coverage rate (%) computed over 1000 simulations, with the intermediate meta-, the simple meta- and the exact mixed models. Meta-models are built with either $n_D = 50$ or $n_D = 100$ design points. Coverage rate (Cov.) is the coverage rate of the 95% confidence interval based on the stochastic approximation of the Fisher matrix.

Outline

Concluding remarks and further works

What is done:

- Replacing an expensive solution of ODE/PDE with meta-model to make MCMC-SAEM cheaper/possible,
- integrating in modeling uncertainties due to the use of meta-model,
- controlling the distance between the MLEs with exact and approximated models.

To be continued:

- in case of complete mixed-model, adapt MCMC algorithms,
- adaptive numerical designs of experiments,
- theoretical results with an adaptive design.

References

Celeux, G. and Diebolt, J. (1985).
The SEM Algorithm: a Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem.
Computational Statistics Quaterly **2**, 73-82.

B. Delyon, M. Lavielle, and E. Moulines.
Convergence of a stochastic approximation version of the EM algorithm.
Ann. Statist., 27:94–128, 1999.

Dempster, E. J., Laird, N.M. and Rubin, D.B. (1977).
Maximum likelihood from incomplete data via EM algorithm. *Annals of the Royal Statistical Society, Series B*, **39**, 1-38.

S. Donnet and A. Samson.
Estimation of parameters in incomplete data models defined by dynamical systems.
J. Statist. Plann. Inference, 137:2815–2831, 2007.

Kuhn, E. and Lavielle, M. (2004).
Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM P&S*, **8**, 115-131.

Thomas A. Louis.
Finding the observed information matrix when using the EM algorithm.
Journal of the Royal Statistical Society. Series B, 44(2):226–233, 1982.

Sacks, Jerome and Schiller, Susannah B. and Welch, William J. (1989).
Designs for Computer Experiments.
Technometrics **31**, 41-47.

WEI, G. and TANNER, M. (1990).
A Monte-Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms.
J. Amer. Statist. Assoc. 85 699–704.