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

Pierre Barbillon¹, Célia Barthélémy², Adeline Leclercq-Samson³

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

²INRIA, POPIX team / Université Paris-Sud

³Laboratoire 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

1. Mixed model
   - Model
   - Estimation

2. Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3. Simulations

4. Conclusion
Outline

1. Mixed model
   - Model
   - Estimation

2. Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3. Simulations

4. Conclusion
Outline

1. Mixed model
   - Model
   - Estimation

2. Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3. Simulations

4. Conclusion
$\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^t$ 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 $i = 1, \ldots, N, j = 1, \ldots, n_i$:

$$
\begin{align*}
\mathbf{y}_{ij} &= f(t_{ij}, \psi_i) + \sigma \varepsilon_{ij}, \quad \varepsilon_{ij} \sim iid \mathcal{N}(0, 1) \\
\psi_i &\sim iid \mathcal{N}(\mu, \Omega),
\end{align*}
$$

where

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

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

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

**Estimate** $\theta$ by maximum likelihood procedure.

The likelihood of the mixed model is the following

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

$$= \prod_{i=1}^{N} \int \frac{1}{(2\pi \sigma^2) n_i/2} \exp \left( -\frac{1}{2} t(y_i - f(t_i, \psi_i))(\sigma^2 l_{ni})^{-1}(y_i - f(t_i, \psi_i)) \right)$$

$$\times \frac{1}{\sqrt{(2\pi)^N |\Omega|}} \exp \left( -\frac{1}{2} t(\psi_i - \mu) \Omega^{-1}(\psi_i - \mu) \right) d\psi_i$$
Outline

1. Mixed model
   - Model
   - Estimation

2. Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3. Simulations

4. Conclusion
EM principle

[Dempster and al., 1977]

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

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

For fixed \( \theta^{(c)} \), if \( Q(\theta, \theta^{(c)}) \) increases, then \( L(y; \theta) \) 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 $\psi$ according to $p(\cdot | y; \theta^{(c)})$:

- 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 \( p(\cdot | y; \theta^{(k-1)}) \),

2. **Stochastic Approximation step**: update the sufficient statistics \( S_D \)

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

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

\[
\begin{align*}
    \hat{\mu}^{(k)} &= \frac{s_{k,1}}{N}, & \hat{\Omega}^{(k)} &= \frac{s_{k,2}}{N} - \frac{s_{k,1} t s_{k,1}}{N^2} \\
    \hat{\sigma}_\varepsilon^2(k) &= \frac{s_{k,3}}{n_{tot}}
\end{align*}
\]
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 $p(\psi_i|y_i; \theta^{(k-1)})$ as stationary distribution.

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

1. Mixed model
   - Model
   - Estimation

2. Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3. Simulations

4. Conclusion
Outline

1. Mixed model
   - Model
   - Estimation

2. Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3. Simulations

4. Conclusion

P. Barbillon, C. Barthélémy, A. Leclercq-Samson

Parametric estimation of complex mixed models
Gaussian Process Emulator

Sacks et al. (1989).

- Assumption: $f$ realization of a Gaussian process $F$:
  \[
  \forall x = (t, \psi) \in E, \\
  F(x) = \sum_{k=1}^{Q} \beta_k h_k(x) + \zeta(x) = H(x)^T \beta + \zeta(x).
  \]

- Pre-computation step: $y_1 = f(x_1), \ldots, y_n = f(x_n)$ evaluations of $f$ on a design $D$.

- Process $F^D$: Conditioning $F$ to $F(x_1) = y_1, \ldots, F(x_n) = y_n$.
  Gaussian Process with mean $m_D(x)$ and covariance $C_D(x, x') \forall x, x'$.

For all $x \in E$,

- $m_D(x)$ approximates $f(x)$,
- $C_D(x, x)$ uncertainty on this approximation.
Gaussian process emulator: illustration

Figure: Posterior mean and realisations of the conditioned process.
Outline

1 Mixed model
   - Model
   - Estimation

2 Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3 Simulations

4 Conclusion
Mixed meta-model

\( f \) is replaced with \( F^D \):

\[
y_{ij} = \underbrace{F^D(t_{ij}, \psi_i)}_{m_D(t_{ij}, \psi_i) + r(t_{ij}, \psi_i)} + \sigma_x \varepsilon_{ij}, \quad \varepsilon_{ij} \sim_{iid} \mathcal{N}(0, 1)
\]

\( \psi_i \sim_{iid} \mathcal{N}(\mu, \Omega) \)

\( r(t, \psi) = F^D(t, \psi) - m_D(t, \psi) \sim \mathcal{GP}(0, C_D(t, \psi; t, \psi)) \),

\( \Rightarrow r() \) 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

\[ p_D(y; \theta) = \int p(\psi; \theta)p_D(y|\psi; \theta) d\psi , \]

\[ = \int p(\psi; \theta) \frac{1}{(2\pi)^{1/2} \sigma^2 \epsilon I_{n_{tot}} + C_D(t, \psi)^{1/2}} \exp \left( -\frac{1}{2} t (y - m_D(t, \psi))(\sigma^2 \epsilon I_{n_{tot}} + C_D(t, \psi))^{-1} (y - m_D(t, \psi)) \right) d\psi. \]

- Likelihood not explicit, because \( m_D(t_{ij}, \psi_i) \) not linear in \( \psi_i \),
- Likelihood cannot be simplified as a product of individual likelihoods because \( y_i \) not independent (matrix \( C_D(t, \psi) \) is a full matrix),
- Computational burden to invert \( C_D(t, \psi) \) 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.
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**

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

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

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

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 $p(y; \theta)$ and $\tilde{p}_D(y; \theta)$ can be as small as we want for $D$ rich enough.
Outline

1. Mixed model
   - Model
   - Estimation

2. Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3. Simulations

4. Conclusion
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 \( D = 6 \text{ mmol} \) 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 \( \Omega \) with the following diagonal elements: \( \omega_{ke} = \omega_{ka} = \omega_{Cl} = 0.1. \)
- 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.

P. Barbillon, C. Barthélémy, A. Leclercq-Samson

Parametric estimation of complex mixed models
### Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Intermediate meta-model</th>
<th>Simple meta-model</th>
<th>Exact model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n_D$ = 50</td>
<td>$n_D$ = 100</td>
<td>$n_D$ = 50</td>
</tr>
<tr>
<td>$\mu \log k_e$</td>
<td>Bias</td>
<td>0.101</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Cov.</td>
<td>94.2</td>
<td>94.4</td>
</tr>
<tr>
<td>$\mu \log k_a$</td>
<td>Bias</td>
<td>-2.441</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.222</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>Cov.</td>
<td>90.9</td>
<td>95.6</td>
</tr>
<tr>
<td>$\mu \log C_l$</td>
<td>Bias</td>
<td>0.388</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>Cov.</td>
<td>87.6</td>
<td>95.1</td>
</tr>
<tr>
<td>$\omega^2 \log k_e$</td>
<td>Bias</td>
<td>-12.113</td>
<td>-2.745</td>
</tr>
<tr>
<td></td>
<td>Cov.</td>
<td>83.2</td>
<td>91.5</td>
</tr>
<tr>
<td>$\omega^2 \log k_a$</td>
<td>Bias</td>
<td>-20.485</td>
<td>-3.442</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>10.696</td>
<td>5.911</td>
</tr>
<tr>
<td></td>
<td>Cov.</td>
<td>72.3</td>
<td>89.7</td>
</tr>
<tr>
<td>$\omega^2 \log C_l$</td>
<td>Bias</td>
<td>0.375</td>
<td>-1.145</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>5.944</td>
<td>5.726</td>
</tr>
<tr>
<td></td>
<td>Cov.</td>
<td>92.6</td>
<td>92.0</td>
</tr>
<tr>
<td>$\sigma^2 \epsilon$</td>
<td>Bias</td>
<td>-45.262</td>
<td>-0.612</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>20.719</td>
<td>0.232</td>
</tr>
</tbody>
</table>

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

1. Mixed model
   - Model
   - Estimation

2. Meta-modeling / GP emulation
   - Principle
   - Consequence on estimation
   - Convergence results

3. Simulations

4. Conclusion
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

The SEM Algorithm: a Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem.
*Computational Statistics Quarterly*, 2, 73-82.

B. Delyon, M. Lavielle, and E. Moulines.
Convergence of a stochastic approximation version of the EM algorithm.


S. Donnet and A. Samson.
Estimation of parameters in incomplete data models defined by dynamical systems.


Thomas A. Louis.
Finding the observed information matrix when using the EM algorithm.

Designs for Computer Experiments.
*Technometrics* 31, 41-47.

A Monte-Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms.