# The SAEM algorithm: a powerful stochastic algorithm for population pharmacology modeling

Marc Lavielle[1,2]
& many collaborators...

[1]INRIA Saclay [2]University Paris Sud Orsay,

MAS 2010, Bordeaux

# Outline

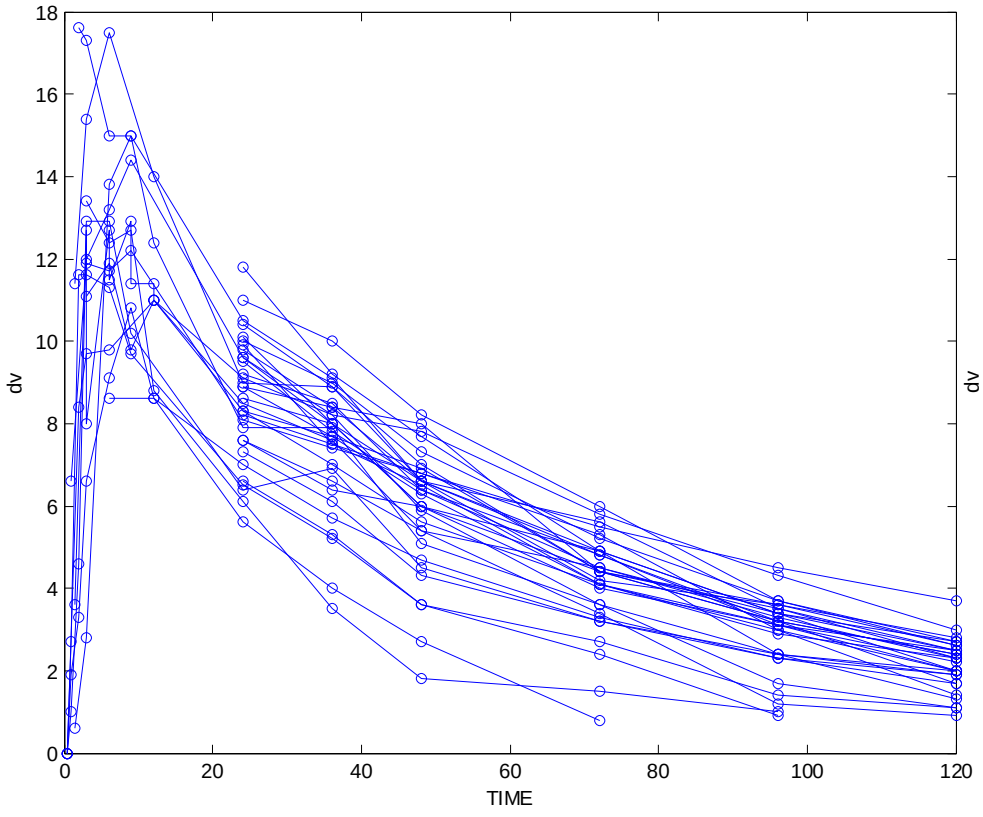
# Outline

# Some examples of data

## Pharmacokinetics/Pharmacodynamics of warfarin

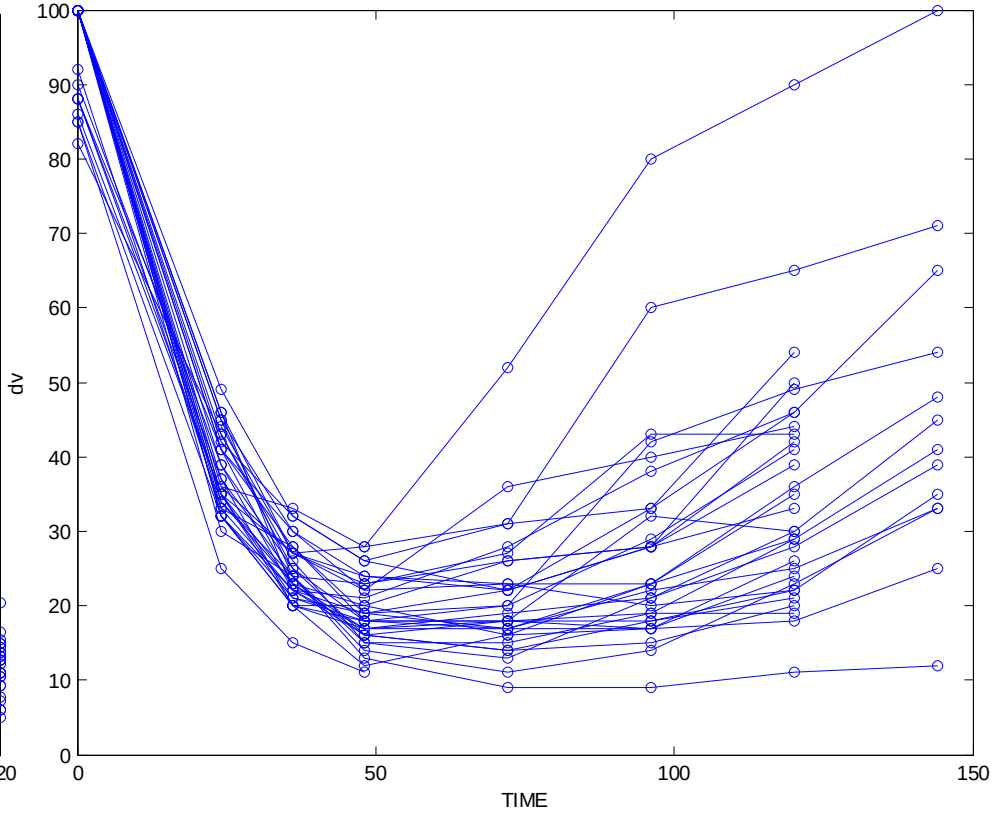
**PK : what the body does to the drug**

warfarin plasma concentration




**PD: what the drug does to the body**

prothrombin complex activity



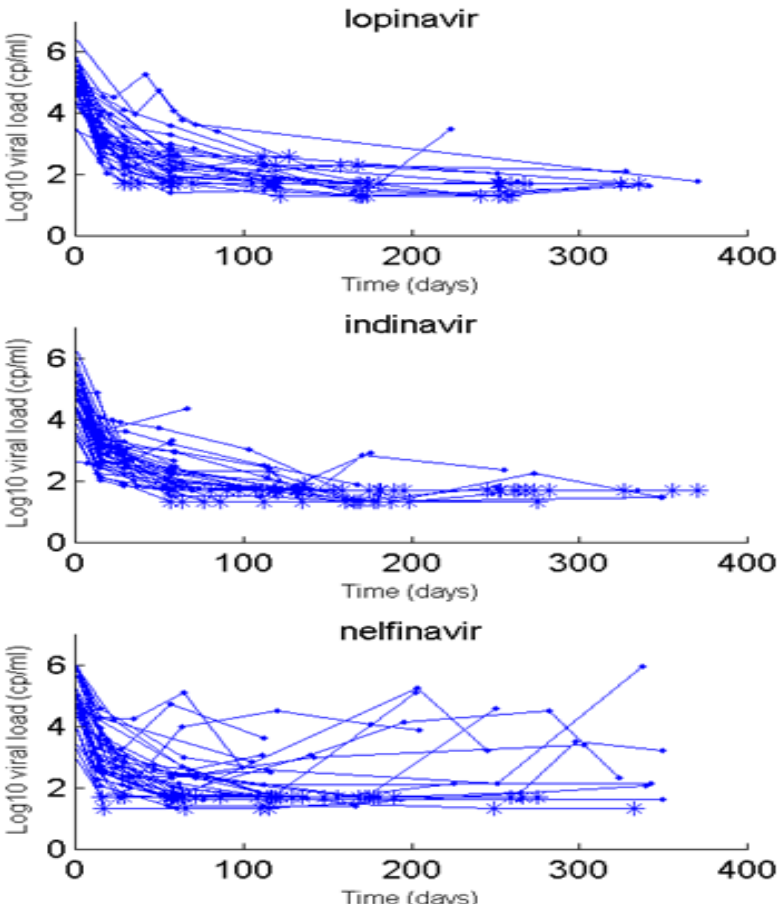
# Some examples of data
## Viral loads and CD4 counts (HIV)

**COPHAR2 TRIAL (ANRS)**

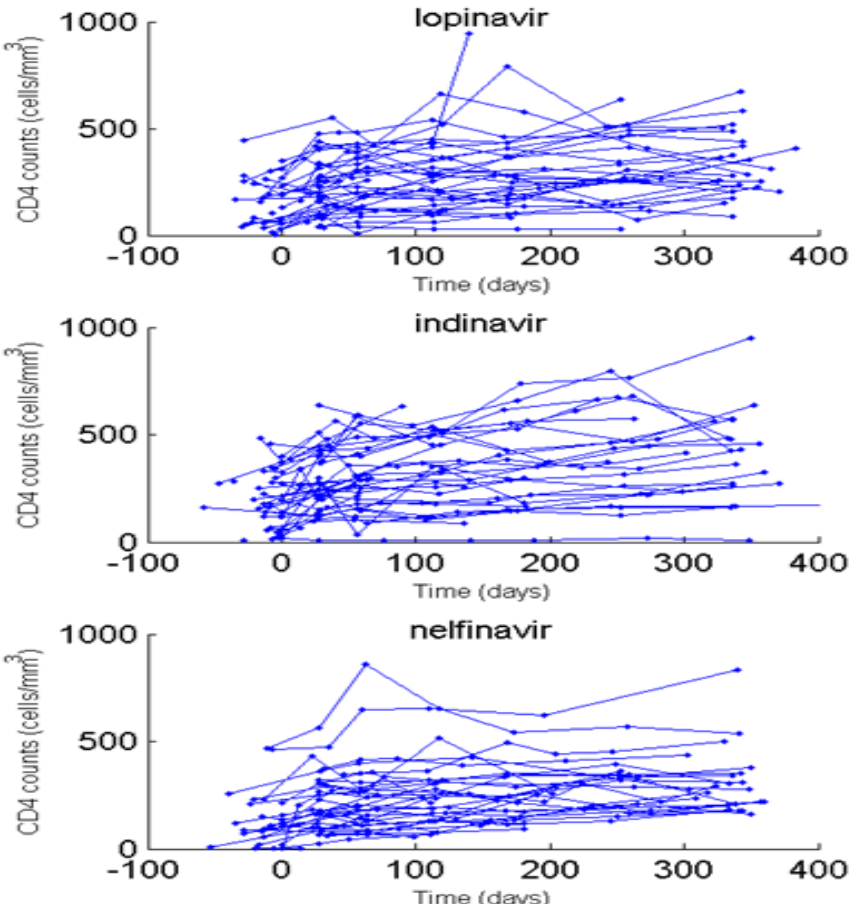
115 patients HIV, starting a tritherapy (2 NRTI + 1 PI)

3 different protease inhibitors

### Viral load

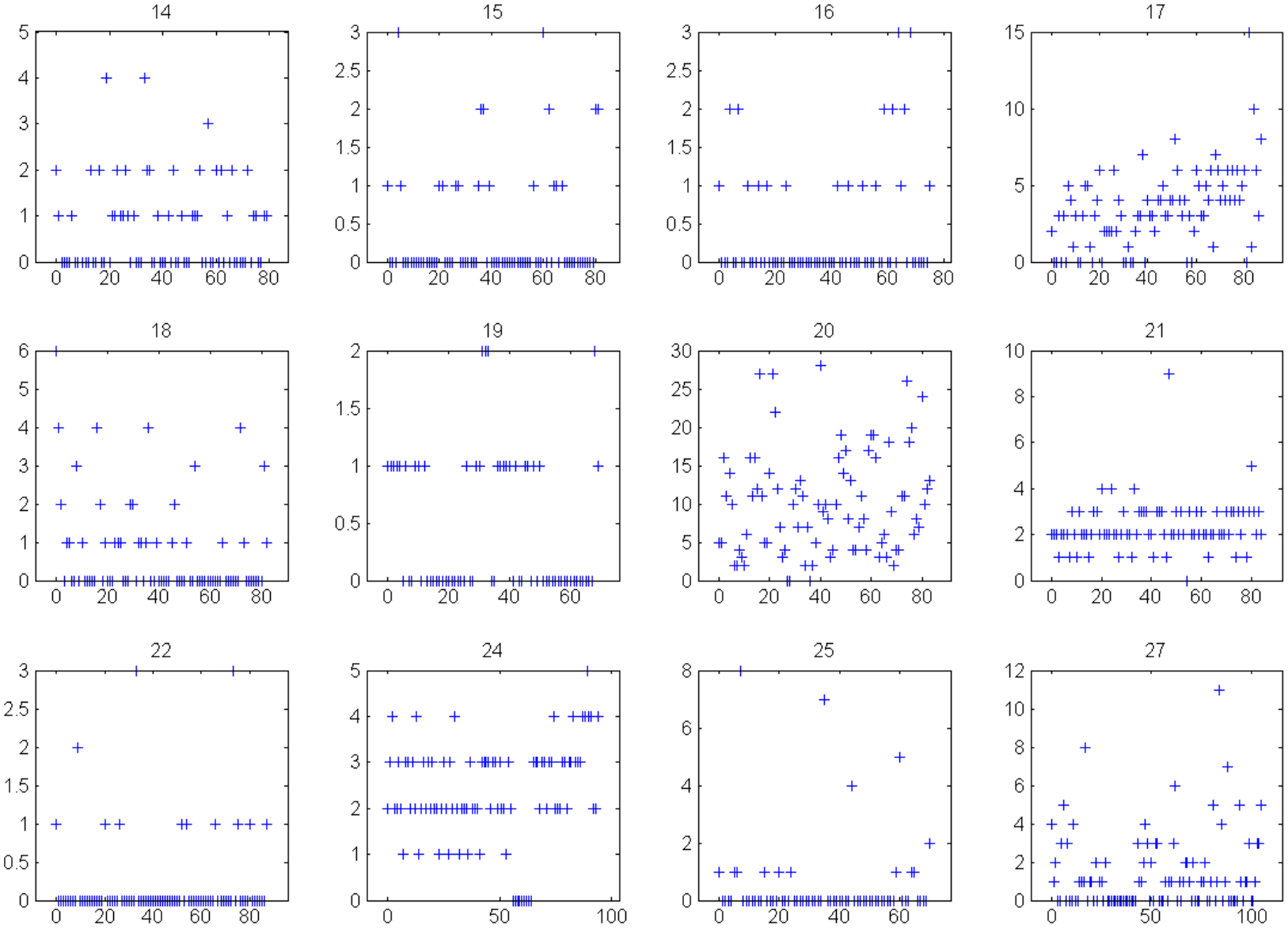



### CD4 count



# Some examples of data

## Daily seizure counts (epilepsy)



# The population approach

Individual approach:

- $y = (y_j, 1 \leq j \leq n)$ : measurements for *a single subject*

$$y \sim h(\,\cdot\,;\psi)$$

- $\psi$ : vector of parameter

Population approach:

- $N$ subjects
- $y_i = (y_{ij}, 1 \leq j \leq n_i)$ : measurements for subject $i$ (observed)

$$y_i \sim h(\,\cdot\,;\psi_i)$$

- $\psi_i$ : individual parameters for subject $i$ (unknown)

$$\psi_i \sim \pi(\,\cdot\,;\theta)$$

- $\theta$ : population parameters of the model (unknown)

# The population approach

Individual approach:

- $y = (y_j, 1 \leq j \leq n)$ : measurements for *a single subject*

$$y \sim h(\,\cdot\,; \psi)$$

- $\psi$ : vector of parameter

Population approach:

- $N$ subjects
- $y_i = (y_{ij}, 1 \leq j \leq n_i)$ : measurements for subject $i$ (observed)

$$y_i \sim h(\,\cdot\,; \psi_i)$$

- $\psi_i$ : individual parameters for subject $i$ (unknown)

$$\psi_i \sim \pi(\,\cdot\,; \theta)$$

- $\theta$ : population parameters of the model (unknown)

# The mixed effects model

- $y_i = (y_{ij}, 1 \leq j \leq n_i)$ : measurements for subject $i$ (observed)

$$y_i \sim h(\,\cdot\,; \psi_i)$$

- $\psi_i$ : individual parameters for subject $i$ (unknown)

$$\psi_i \sim \pi(\,\cdot\,; \theta)$$

In a mixed effects model, $\psi_i$ is decomposed into fixed and random effects:

$$\psi_i = g(\beta, C_i, \eta_i)$$

$C_i$: known vector of individual covariates
$\beta$: unknown vector of fixed effects
$\eta_i$: unknown vector of random effects, usually $\mathcal{N}(0, \Omega)$

# The mixed effects model

- $y_i = (y_{ij}, 1 \leq j \leq n_i)$ : measurements for subject $i$ (observed)

$$y_i \sim h(\,\cdot\,; \psi_i)$$

- $\psi_i$ : individual parameters for subject $i$ (unknown)

$$\psi_i \sim \pi(\,\cdot\,; \theta)$$

In a mixed effects model, $\psi_i$ is decomposed into fixed and random effects:

$$\psi_i = g(\beta, C_i, \eta_i)$$

$C_i$: known vector of individual covariates
$\beta$: unknown vector of fixed effects
$\eta_i$: unknown vector of random effects, usually $\mathcal{N}(0, \Omega)$

# The mixed effects model

- $y_i = (y_{ij}, 1 \leq j \leq n_i)$ : measurements for subject $i$ (observed)

$$y_i \sim h(\,\cdot\,; \psi_i)$$

- $\psi_i$ : individual parameters for subject $i$ (unknown)

$$\psi_i \sim \pi(\,\cdot\,; \theta)$$

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

**Maximum Likelihood Estimation:** maximize $p(y;\theta)$,
**Fisher Information Matrix:** compute $\partial^2_\theta \log p(y;\hat{\theta})$
**Model Selection:** compute $p(y;\hat{\theta})$,
**Estimation of the indiv. parameters:** maximize $p(\psi_i|y_i;\hat{\theta})$.

# Outline

# The EM algorithm (Expectation-Maximization)
## (Dempster, Laird et Rubin, JRSSB, 1977)

Since $\psi$ is not observed, $\log p(y, \psi; \theta)$ cannot be used for estimating $\theta$. Then

Iteration $k$ of the algorithm:

- step E : evaluate the quantity

$$\boxed{Q_k(\theta) = \mathbb{E}[\log p(y, \psi; \theta) | y; \theta_{k-1}]}$$

- step M : update the estimation of $\theta$:

$$\boxed{\theta_k = Argmax \;\; Q_k(\theta)}$$

# The EM algorithm (Expectation-Maximization)
## (Dempster, Laird et Rubin, JRSSB, 1977)

Since $\psi$ is not observed, $\log p(y, \psi; \theta)$ cannot be used for estimating $\theta$. Then

Iteration $k$ of the algorithm:

- step E : evaluate the quantity

$$\boxed{Q_k(\theta) = \mathbb{E}[\log p(y, \psi; \theta) | y; \theta_{k-1}]}$$

- step M : update the estimation of $\theta$:

$$\boxed{\theta_k = Argmax \;\; Q_k(\theta)}$$

# The EM algorithm (Expectation-Maximization)
## (Dempster, Laird et Rubin, JRSSB, 1977)

Since $\psi$ is not observed, $\log p(y, \psi; \theta)$ cannot be used for estimating $\theta$. Then

Iteration $k$ of the algorithm:

- step E : evaluate the quantity

$$\boxed{Q_k(\theta) = \mathbb{E}[\log p(y, \psi; \theta) | y; \theta_{k-1}]}$$

- step M : update the estimation of $\theta$:

$$\boxed{\theta_k = Argmax \;\; Q_k(\theta)}$$

# The SAEM algorithm (Stochastic Approximation of EM)

Delyon, Lavielle and Moulines (the Annals of Statistics, 1999)

Iteration $k$ of the algorithm:

- step E :
  - *Simulation:* draw the non observed data $\psi^{(k)}$ with the conditional distribution $p(\psi \,|y;\theta_{k-1})$
  - *Stochastic approximation:*

$$\boxed{Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k \left[\log p(y, \psi^{(k)}; \theta) - Q_{k-1}(\theta)\right]}$$

    $(\gamma_k)$ is a decreasing sequence: $\sum \gamma_k = +\infty$, $\sum \gamma_k^2 < +\infty$.

- step M:
  - *Maximisation:* update the estimation of $\theta$

$$\boxed{\theta_k = Argmax \;\; Q_k(\theta)}$$

# The SAEM algorithm (Stochastic Approximation of EM)

Delyon, Lavielle and Moulines (the Annals of Statistics, 1999)

Iteration $k$ of the algorithm:

- step E :
  - *Simulation:* draw the non observed data $\psi^{(k)}$ with the conditional distribution $p(\psi \mid y; \theta_{k-1})$
  - *Stochastic approximation:*

$$\boxed{Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k \left[\log p(y, \psi^{(k)}; \theta) - Q_{k-1}(\theta)\right]}$$

    $(\gamma_k)$ is a decreasing sequence: $\sum \gamma_k = +\infty$, $\sum \gamma_k^2 < +\infty$.

- step M:
  - *Maximisation:* update the estimation of $\theta$

$$\boxed{\theta_k = Argmax \;\; Q_k(\theta)}$$

# Coupling SAEM with MCMC

Kuhn and Lavielle, ESAIM P&S, 2004

Let $\Pi_\theta$ be the transition probability of an *ergodic Markov Chain* with limiting distribution $p_{\psi|Y}(\cdot|y;\theta)$.

Iteration $k$ of the algorithm:

- *Simulation* : draw $\psi^{(k)}$ according to the transition probability $\Pi_{\theta_{k-1}}(\psi^{(k-1)},\cdot)$.
- *Stochastic approximation:*

$$\boxed{Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k \left[\log p(y,\psi^{(k)};\theta) - Q_{k-1}(\theta)\right]}$$

- *Maximization:*

$$\boxed{\theta_k = Argmax \;\; Q_k(\theta)}$$

# The main convergence Theorem

## Theorem

*Under very general technical conditions, the SAEM sequence* $(\theta_k)$ *converges a.s. to some (local) maximum of the observed likelihood* $p(y;\theta)$.

## Proof.

1. Delyon, Lavielle & Moulines *The Annals of Statistics* (1999)
   Exact simulation assumed, compactness of $(\psi_i^{(k)})$ not required
2. Kuhn & Lavielle *ESAIM P&S* (2004)
   Markovian perturbation allowed, compactness of $(\psi_i^{(k)})$ required
3. Allassonnière, Kuhn & Trouvé *Bernoulli* (2010)
   Markovian perturbation allowed, compactness of $(\psi_i^{(k)})$ not required

□

# The main convergence Theorem

## Theorem

*Under very general technical conditions, the SAEM sequence* $(\theta_k)$ *converges a.s. to some (local) maximum of the observed likelihood* $p(y;\theta)$.

## Proof.

1. Delyon, Lavielle & Moulines *The Annals of Statistics* (1999)
   Exact simulation assumed, compactness of $(\psi_i^{(k)})$ not required
2. Kuhn & Lavielle *ESAIM P&S* (2004)
   Markovian perturbation allowed, compactness of $(\psi_i^{(k)})$ required
3. Allassonnière, Kuhn & Trouvé *Bernoulli* (2010)
   Markovian perturbation allowed, compactness of $(\psi_i^{(k)})$ not required

□

# Outline

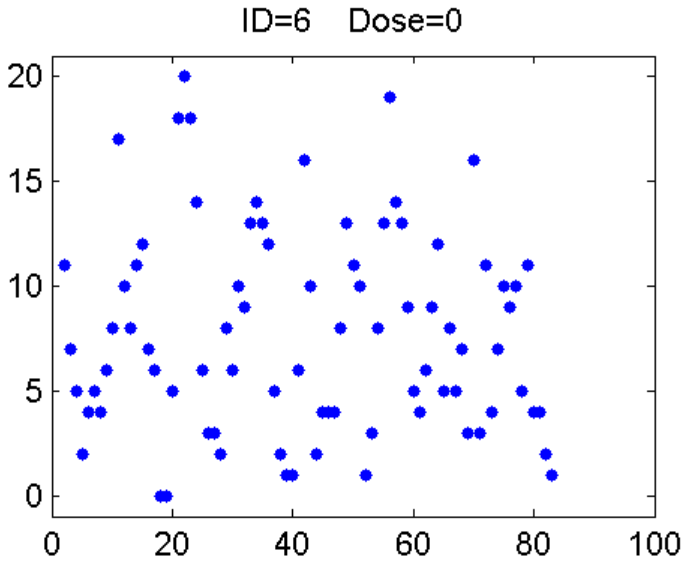
**The data:** sequences of daily seizures in a sample of 788 epileptic patients during a 12 weeks active treatment phase which represents a total of 41198 seizure counts.

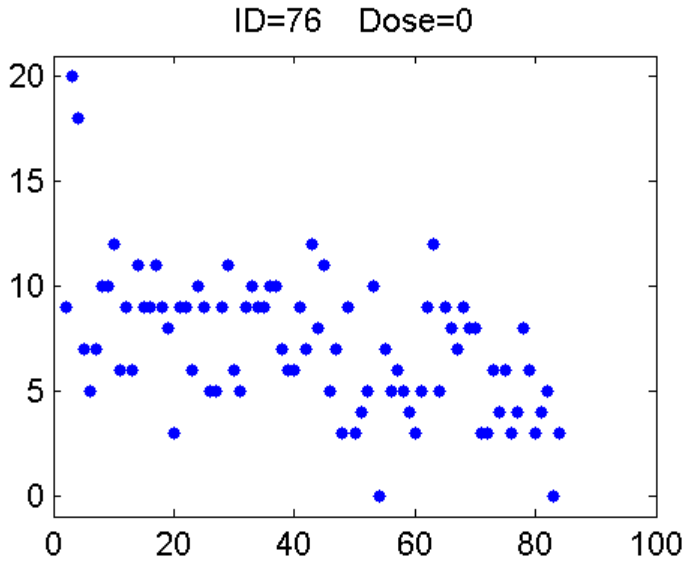
**The design:** one placebo group and one treatment group

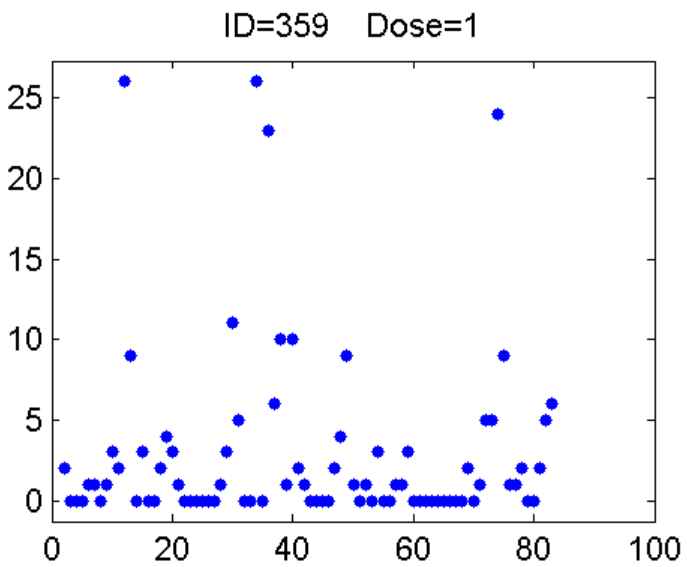
Joint work with Maud Delattre, Rada Savic, Mats Karlsson,
Collaboration with Pfizer.

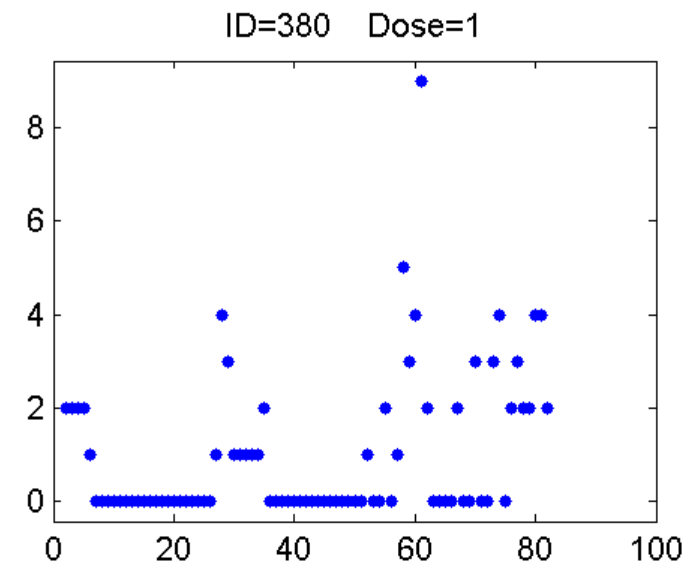
# Inference in Mixed HMM

## Observed seizure counts of 4 typical subjects



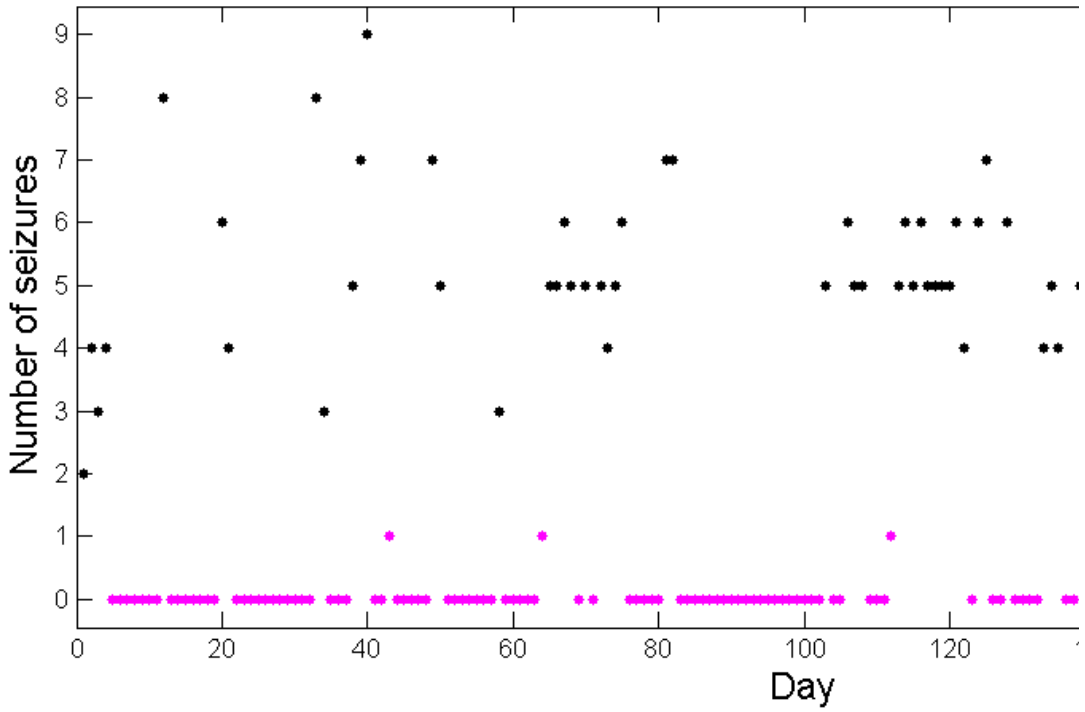



# Inference in Mixed HMM

## A mixture of 2 Poisson distributions

***A Poisson mixture model*** assumes that there exists a sequence of ***hidden states*** ($z_{ij}$)




$z_{ij} = 2$ : high activity

$y_{ij}$ : Poisson($\lambda_{2,i}$)

$P(z_{ij} = 2) = 1-\pi_i$

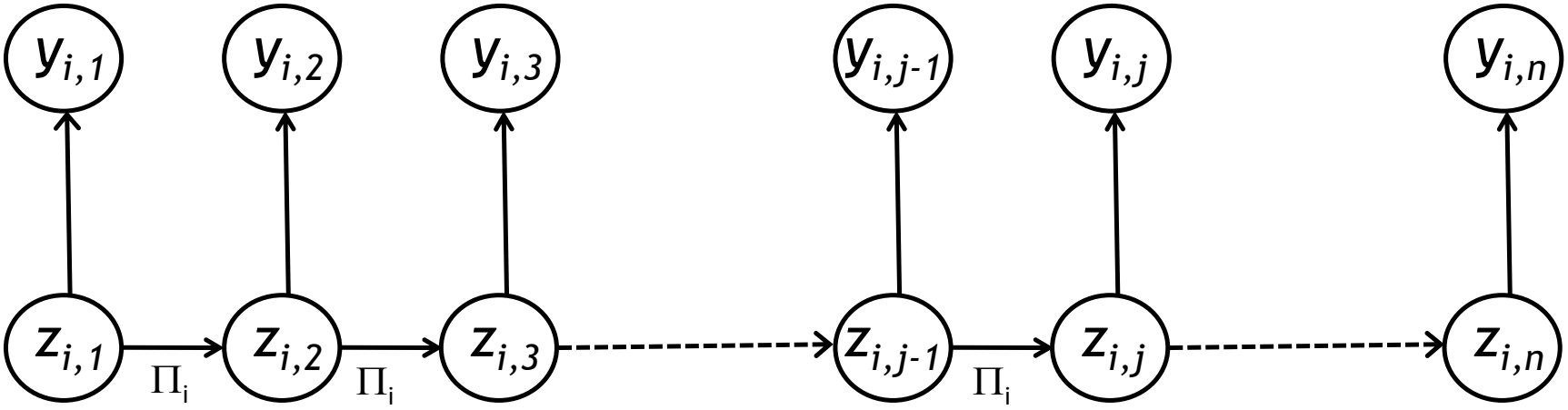
$z_{ij} = 1$ : low activity

$y_{ij}$ : Poisson($\lambda_{1,i}$)

$P(z_{ij} = 1) = \pi_i$

# Inference in Mixed HMM

## A Markovian dynamics




$(z_{i,j})$ is a Markov Chain with transition matrix $\Pi_i = \begin{pmatrix} p_{11,i} & p_{12,i} \\ p_{21,i} & p_{22,i} \end{pmatrix}$

where $p_{\ell m,i} = \mathbb{P}(z_{ij} = m | z_{i,j-1} = \ell)$

$$\begin{aligned} &\text{if } z_{ij} = 1\,, \quad y_{ij} \sim \text{Poisson}(\lambda_{1i}) \\ &\text{if } z_{ij} = 2\,, \quad y_{ij} \sim \text{Poisson}(\lambda_{2i}) \end{aligned}$$

$$\begin{cases} x_i = 0 & \texttt{if i belongs to the } \textit{placebo} \texttt{ group} \\ x_i = 1 & \texttt{if i belongs to the } \textit{treatment} \texttt{ group} \end{cases}$$

$$\begin{aligned} \log(\lambda_{1i}) &= \mu_1 + \beta_1 x_i + \eta_{1i} \\ \log(\alpha_i) &= \mu_2 + \beta_2 x_i + \eta_{2i} \quad ; \quad \lambda_{2i} = \lambda_{1i} + \alpha_i \\ \text{logit}(p_{11,i}) &= \mu_3 + \beta_3 x_i + \eta_{3i} \\ \text{logit}(p_{21,i}) &= \mu_4 + \beta_4 x_i + \eta_{4i} \end{aligned}$$

$$\eta_i = (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i})' \backsim \mathcal{N}(0, \Omega).$$

**Estimation of the pop. param.:** $\hat{\theta}$ maximizes $p(y;\theta)$,
**Estimation of the indiv; param.:** $\hat{\psi}_i$ maximizes $p(\psi|y_i;\hat{\theta})$,
**Estimation of the hidden states:** $\hat{z}_i$ maximizes $p(z|y_i,\hat{\psi}_i;\hat{\theta})$.

Simulation of $\psi$ with the conditional distribution $p(\psi|y;\theta)$ requires to compute

$$\begin{aligned} p(y,\psi;\theta) &= p(y|\psi;\theta)p(\psi;\theta) \\ &= p(y|\psi)p(\psi;\theta) \\ &= \int p(y,z|\psi)dz \; p(\psi;\theta) \end{aligned}$$

The **Baum-Welch algorithm** allows to compute the conditional distribution of the observations $p(y|\psi)$.

Then, the *Simulation-step* reduces to simulate the individual parameters $\psi$:

We don't need to simulate the hidden states $z$ for the *Simulation-step*!

Simulation of $\psi$ with the conditional distribution $p(\psi|y;\theta)$ requires to compute

$$\begin{aligned} p(y,\psi;\theta) &= p(y|\psi;\theta)p(\psi;\theta) \\ &= p(y|\psi)p(\psi;\theta) \\ &= \int p(y,z|\psi)dz \; p(\psi;\theta) \end{aligned}$$

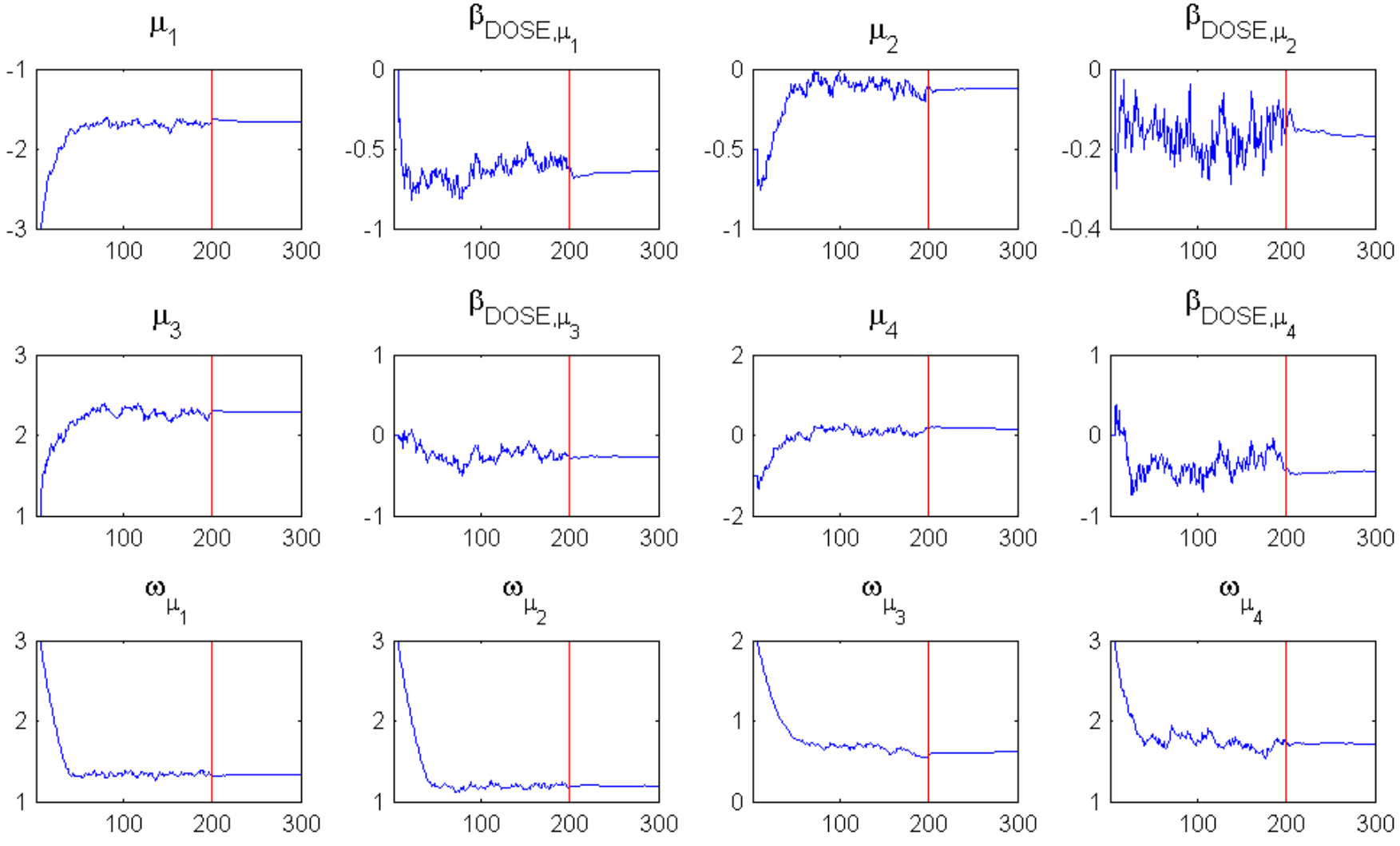
The **Baum-Welch algorithm** allows to compute the conditional distribution of the observations $p(y|\psi)$.

Then, the *Simulation-step* reduces to simulate the individual parameters $\psi$:

We don't need to simulate the hidden states $z$ for the *Simulation-step*!

# Inference in Mixed HMM

## Convergence of SAEM




Elapsed time is 389 seconds.

# Inference in Mixed HMM

## Estimation of the population parameters

**Estimation of the population parameters**

| | | parameter | s.e. (s.a.) | r.s.e.(%) | p-value |
|---|---|---|---|---|---|
| mu1 | : | -1.66 | 0.09 | 5 | |
| beta_mu1(DOSE_G2) | : | -0.643 | 0.16 | 24 | 4.4e-005 |
| mu2 | : | -0.121 | 0.087 | 72 | |
| beta_mu2(DOSE_G2) | : | -0.17 | 0.13 | 76 | 0.19 |
| mu3 | : | 2.29 | 0.081 | 4 | |
| beta_mu3(DOSE_G2) | : | -0.263 | 0.13 | 49 | 0.041 |
| mu4 | : | 0.157 | 0.17 | 109 | |
| beta_mu4(DOSE_G2) | : | -0.442 | 0.27 | 61 | 0.1 |
| | | | | | |
| omega_mu1 | : | 1.34 | 0.062 | 5 | |
| omega_mu2 | : | 1.2 | 0.058 | 5 | |
| omega_mu3 | : | 0.616 | 0.058 | 9 | |
| omega_mu4 | : | 1.71 | 0.14 | 8 | |

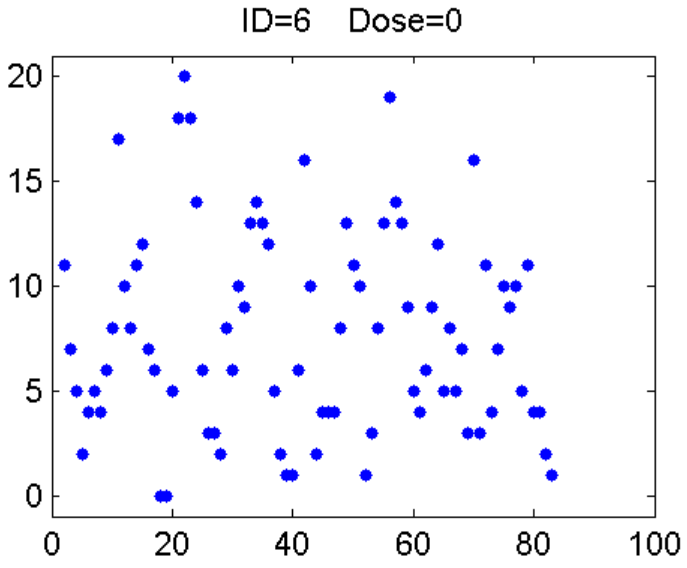
1. For $i = 1, \ldots, N$, the MAP estimate $\hat{\psi}_i$ maximizes the conditional distribution $p(\psi_i | y_i; \hat{\theta})$.

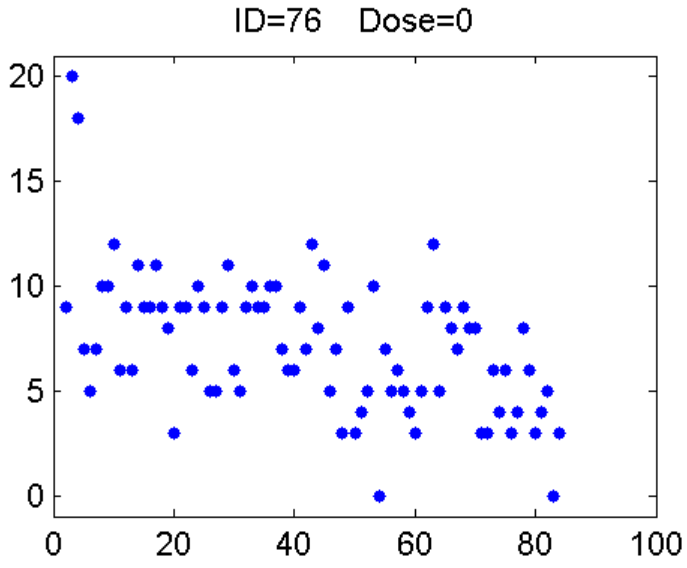
2. For $i = 1, \ldots, N$, the MAP estimate $\hat{z}_i = (\hat{z}_{ij}; 1 \leq j \leq n_i)$ maximizes the conditional distribution $p(z_i | y_i, \hat{\psi}_i)$.

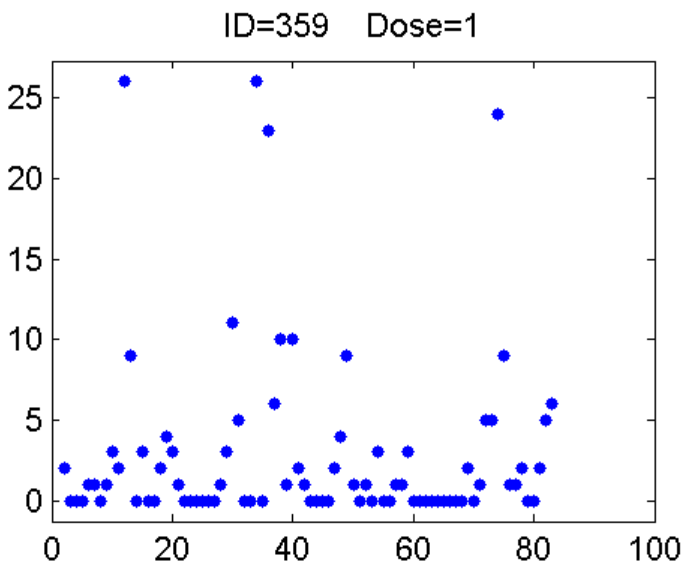
   **Remark:** $\hat{z}_i$ can be computed thanks to the **Viterbi algorithm**.

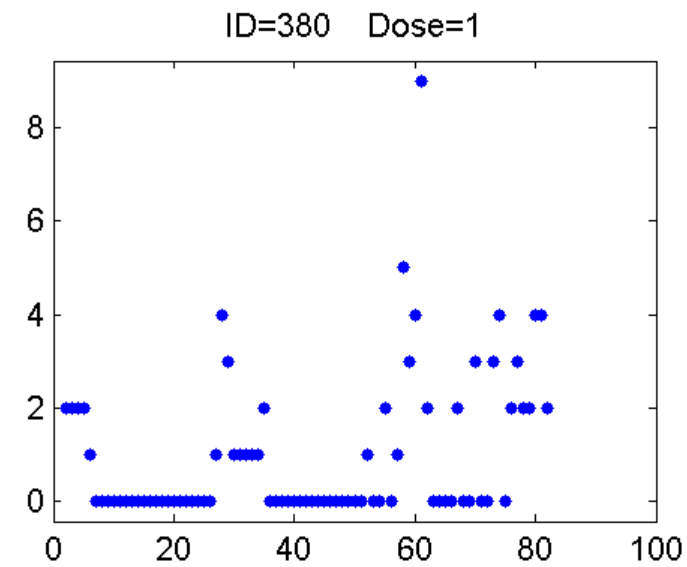
# Inference in Mixed HMM

## Observed seizure counts of 4 typical subjects



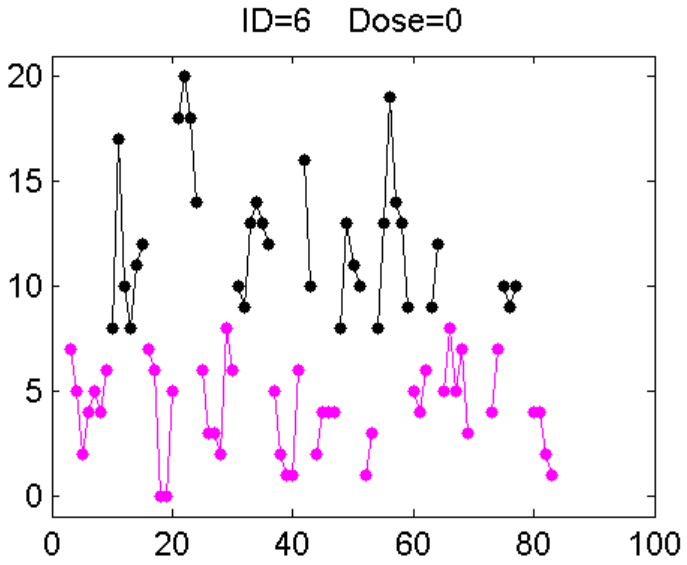



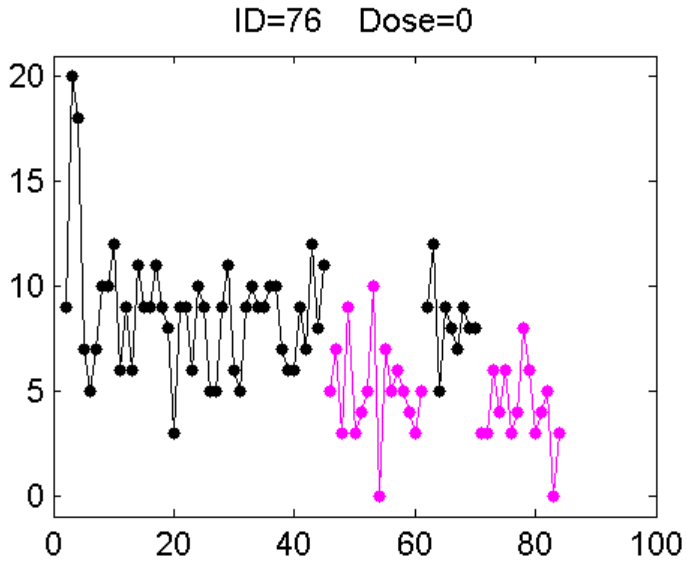



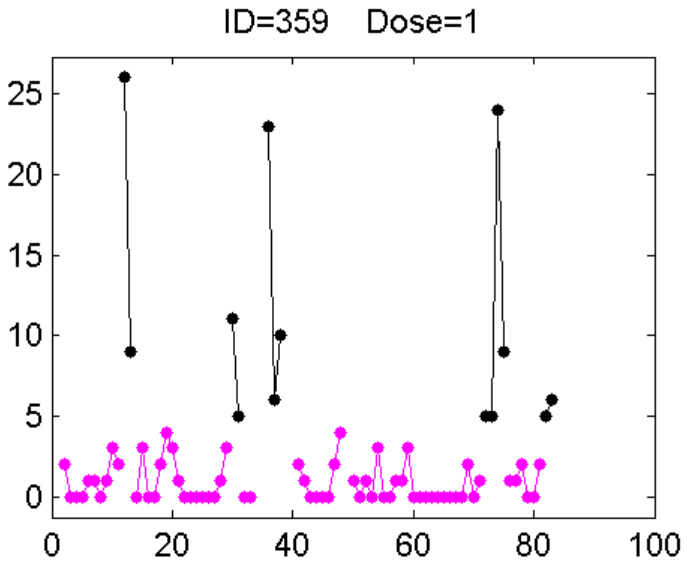



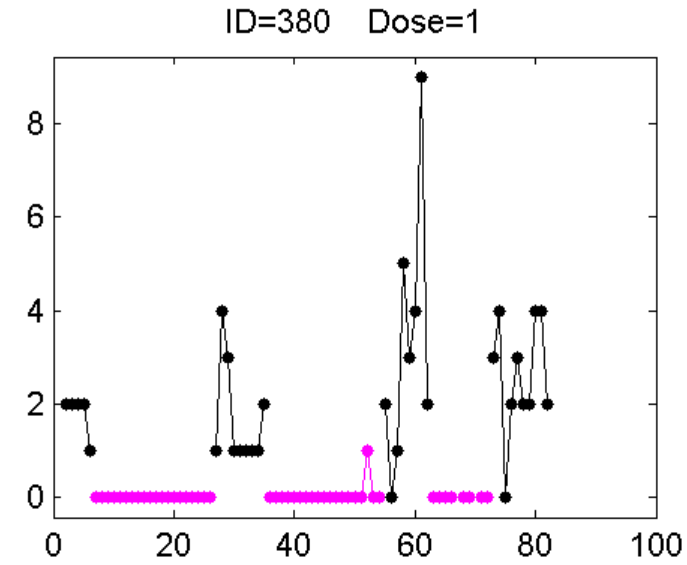
# Inference in Mixed HMM

## Seizure counts and estimated states of 4 typical subjects



# Outline

# Extension to (Markovian) state-space models

$$\begin{aligned} X_{i,j+1} &= F_{i,j+1}(X_{ij}, U_{i,j+1}; \psi_i) \\ y_{ij} &= C_{ij}(X_{ij}, E_{ij}; \psi_i) \end{aligned}$$

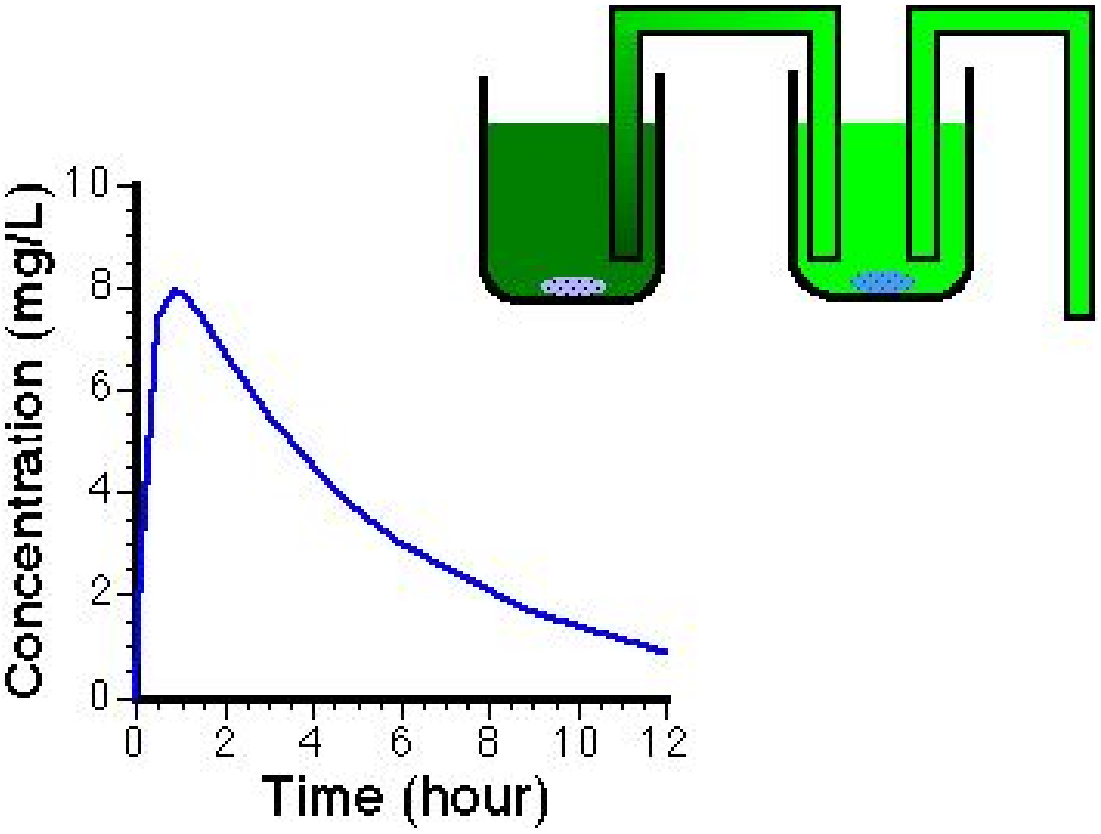
The proposed methodology for NLMEM requires to compute the conditional densities

$$p(y_i|\psi_i) = \int p(y_i, X_i|\psi_i)\, dX_i$$

- HMM: Baum-Welch algorithm
- Linear state-space system: Kalman filter
- Non linear system: Extended Kalman filter, particle filters,

# A pharmacokinetic (PK) example

## oral administration, one compartment



# A pharmacokinetic (PK) example

## 1st order oral absorption model with one compartment and linear elimination

Dose $D$ at time t=0

absorption (rate $k_a$) $\rightarrow$ DRUG AMOUNT $Qc(t)$ $\rightarrow$ elimination (rate $k_e$)

$$\begin{aligned} \frac{dQa}{dt}(t) &= -k_a Qa(t) \quad ; \quad Qa(0) = D \\ \frac{dQc}{dt}(t) &= k_a Qa(t) - k_e Qc(t) \quad ; \quad Qc(0) = 0 \end{aligned}$$

$Qa(t)$: amount at absorption site,
$Qc(t)$: amount in central compartment,
$Cc(t) = Qc(t)/V$: concentration in central compartment.

$(k_a, k_e, V)$: PK parameters

# A pharmacokinetic (PK) example

## The population model assuming a deterministic dynamics (ODEs)

$$
\begin{aligned}
dQa_i(t) &= -ka_i\, Qa_i(t)\, dt \\
dCc_i(t) &= \frac{ka_i}{V_i}\, Qa_i(t)\, dt - ke_i\, Cc_i(t) \\
y_{ij} &= Cc_i(t_{ij}) + \sigma_i e_{ij}
\end{aligned}
$$

$$e_{ij} \sim_{i.i.d.} \mathcal{N}(0,1)$$

$$\psi_i = (ka_i, V_i, Cl_i, \sigma_i)$$

$$\log(\psi_i) \sim \mathcal{N}(\log(\psi_{\text{pop}}), \Omega)$$

# A pharmacokinetic (PK) example

## The population model assuming a stochastic dynamics (SDEs)

$$
\begin{aligned}
dQa_i(t) &= -ka_i\, Qa_i(t)\, dt + \gamma_{i1}\, dW_{i1}(t) \\
dCc_i(t) &= \frac{ka_i}{V_i}\, Qa_i(t)\, dt - ke_i\, Cc_i(t) - \frac{\gamma_{i1}}{V_i}\, dW_{i1}(t) + \gamma_{i2}\, dW_{i2}(t) \\
y_{ij} &= Cc_i(t_{ij}) + \sigma_i e_{ij}
\end{aligned}
$$

$\{W_{i1}(t)\}$ and $\{W_{i2}(t)\}$ are independent Wiener processes

$e_{ij} \sim_{i.i.d.} \mathcal{N}(0,1)$

$\psi_i = (ka_i, V_i, Cl_i, \sigma_i, \gamma_{i1}, \gamma_{i2})$

$\log(\psi_i) \sim \mathcal{N}(\log(\psi_{\text{pop}}), \Omega)$

Let $X_i = (Qa_i, Cc_i)'$. Then the model reduces to

$$\begin{aligned} dX_i(t) &= A_i X(t)\, dt + \Gamma_i dW_i(t) \\ y_{ij} &= [0\ 1] X_i(t_{ij}) + \sigma_i e_{ij} \end{aligned}$$

where

$$\begin{aligned} A_i &= A(\psi_i) \\ \Gamma_i &= \Gamma(\psi_i) \end{aligned}$$

# A pharmacokinetic (PK) example
## Continuous and discrete time representations of the model

Continuous time representation of the model:

$$\begin{aligned} dX_i(t) &= A_i X_i(t)\, dt + \Gamma_i dW_i(t) \\ y_{ij} &= [0\ 1] X_i(t_{ij}) + \sigma_i e_{ij} \end{aligned}$$

Discrete time representation of the model:

$$\begin{aligned} X_{i,j+1} &= e^{A_i(t_{i,j+1}-t_{ij})} X_{ij} + U_{i,j+1} \\ y_{ij} &= [0\ 1] X_{ij} + \sigma_i e_{ij} \end{aligned}$$

$$U_{i,j+1} \sim \mathcal{N}\left(0, \int_0^{t_{i,j+1}-t_{ij}} \left(e^{A_i\, t} \Gamma_i\right)\left(e^{A_i\, t} \Gamma_i\right)' \right) dt$$

Use the Kalman filter for computing

$$p(y_i|\psi_i) = \int p(y_i, X_i|\psi_i)\, dX_i$$

# A pharmacokinetic (PK) example
## Continuous and discrete time representations of the model

Continuous time representation of the model:

$$\begin{aligned} dX_i(t) &= A_i X_i(t)\, dt + \Gamma_i dW_i(t) \\ y_{ij} &= [0\ 1] X_i(t_{ij}) + \sigma_i e_{ij} \end{aligned}$$

Discrete time representation of the model:

$$\begin{aligned} X_{i,j+1} &= e^{A_i(t_{i,j+1}-t_{ij})} X_{ij} + U_{i,j+1} \\ y_{ij} &= [0\ 1] X_{ij} + \sigma_i e_{ij} \end{aligned}$$

$$U_{i,j+1} \sim \mathcal{N}\left(0, \int_0^{t_{i,j+1}-t_{ij}} \left(e^{A_i\, t}\Gamma_i\right)\left(e^{A_i\, t}\Gamma_i\right)' \right) dt$$

Use the Kalman filter for computing

$$p(y_i|\psi_i) = \int p(y_i, X_i|\psi_i)\, dX_i$$

# A pharmacokinetic (PK) example
## Continuous and discrete time representations of the model

Continuous time representation of the model:

$$\begin{aligned} dX_i(t) &= A_i X_i(t)\, dt + \Gamma_i dW_i(t) \\ y_{ij} &= [0\ 1] X_i(t_{ij}) + \sigma_i e_{ij} \end{aligned}$$

Discrete time representation of the model:

$$\begin{aligned} X_{i,j+1} &= e^{A_i(t_{i,j+1}-t_{ij})} X_{ij} + U_{i,j+1} \\ y_{ij} &= [0\ 1] X_{ij} + \sigma_i e_{ij} \end{aligned}$$

$$U_{i,j+1} \sim \mathcal{N}\left(0, \int_0^{t_{i,j+1}-t_{ij}} \left(e^{A_i\, t}\Gamma_i\right)\left(e^{A_i\, t}\Gamma_i\right)' \right) dt$$

Use the Kalman filter for computing

$$p(y_i|\psi_i) = \int p(y_i, X_i|\psi_i)\, dX_i$$

# A pharmacokinetic (PK) example

## A simulated numerical example

$$
\begin{aligned}
dQa_i(t) &= -ka_i\, Qa_i(t)\, dt + \gamma_{i1}\, dW_{i1}(t) \\
dCc_i(t) &= \frac{ka_i}{V_i}\, Qa_i(t)\, dt - \frac{Cl_i}{V_i}\, Cc_i(t) - \frac{\gamma_{i1}}{V_i}\, dW_{i1}(t) + \gamma_{i2}\, dW_{i2}(t) \\
y_{ij} &= [0\ 1] X_i(t_{ij}) + \sigma_i\, e_{ij}
\end{aligned}
$$

$$
N = 50
$$

$$
(t_{ij}) = (1, 2, 3, \ldots, 24)
$$

$$
ka_{\text{pop}} = 0.5 \quad , \quad V_{\text{pop}} = 0.5 \quad , \quad Cl_{\text{pop}} = 0.1
$$

$$
\omega_{ka} = 0.3 \quad , \quad \omega_V = 0.1 \quad , \quad \omega_{Cl} = 0.3
$$

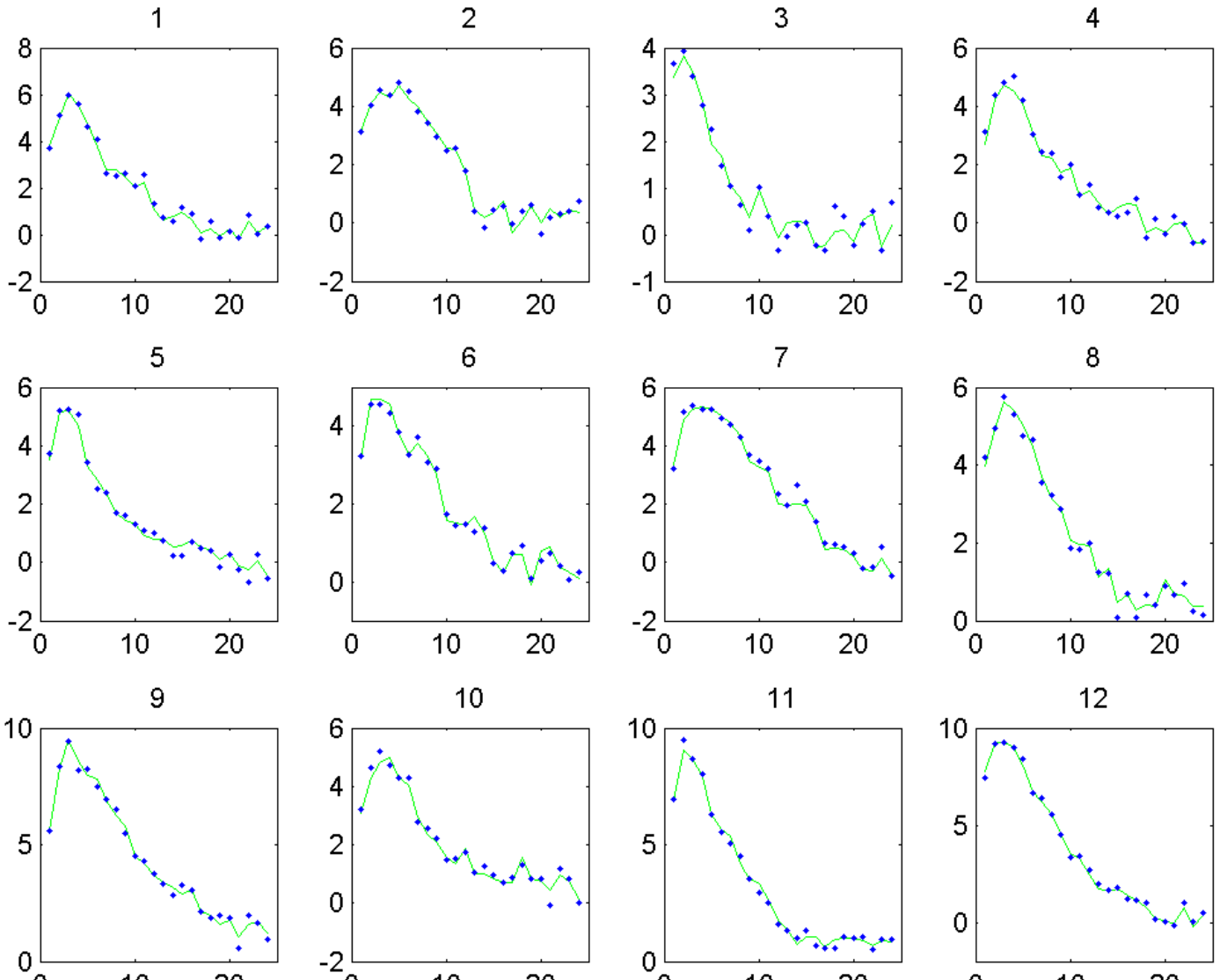
$$
\gamma_1 = 0.2 \quad , \quad \gamma_1 = 0.15 \quad , \quad \sigma = 0.2
$$

$$
\omega_{\gamma_1} = \omega_{\gamma_2} = \omega_\sigma = 0
$$

# A pharmacokinetic (PK) example

## A simulated numerical example

— true concentrations ($Cc_i$) ; · observed concentrations ($y_{ij}$)



# A pharmacokinetic (PK) example

## Estimation of the population parameters

Estimation of the population parameters

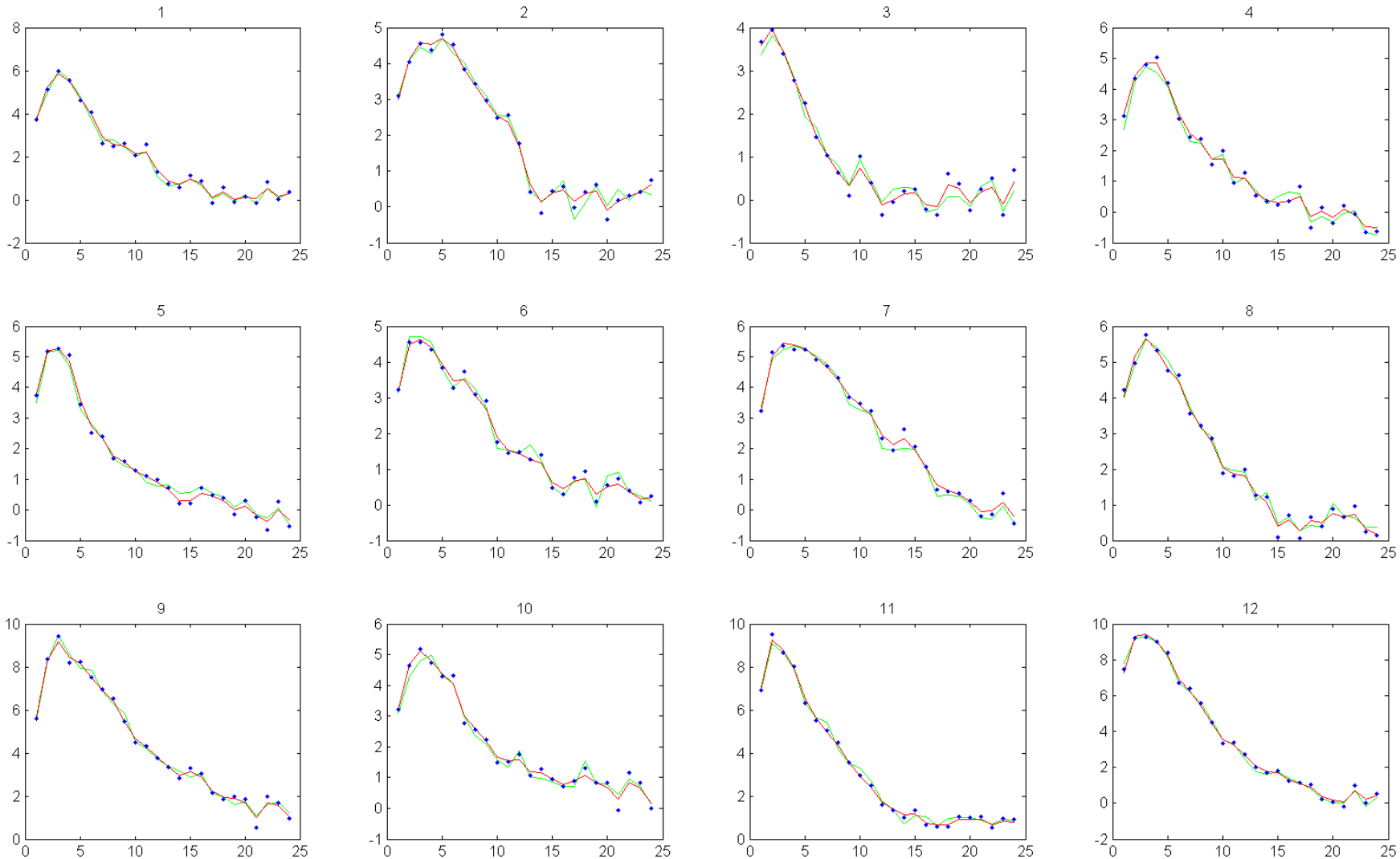
| | | parameter | s.e. (s.a.) | r.s.e.(%) |
|---|---|---|---|---|
| ka | : | 0.522 | 0.028 | 5 |
| V | : | 0.511 | 0.012 | 2 |
| Cl | : | 0.0962 | 0.0043 | 4 |
| gamma_1 | : | 0.184 | 0.025 | 13 |
| gamma_2 | : | 0.169 | 0.028 | 16 |
| sigma | : | 0.23 | 0.022 | 10 |
| omega_ka | : | 0.309 | 0.043 | 14 |
| omega_V | : | 0.0964 | 0.04 | 42 |
| omega_Cl | : | 0.308 | 0.033 | 11 |
| omega_gamma_1 | : | 0 | - | - |
| omega_gamma_2 | : | 0 | - | - |
| omega_sigma | : | 0 | - | - |

Computation time: 2' 30"

# A pharmacokinetic (PK) example

## Estimation of the dynamical system

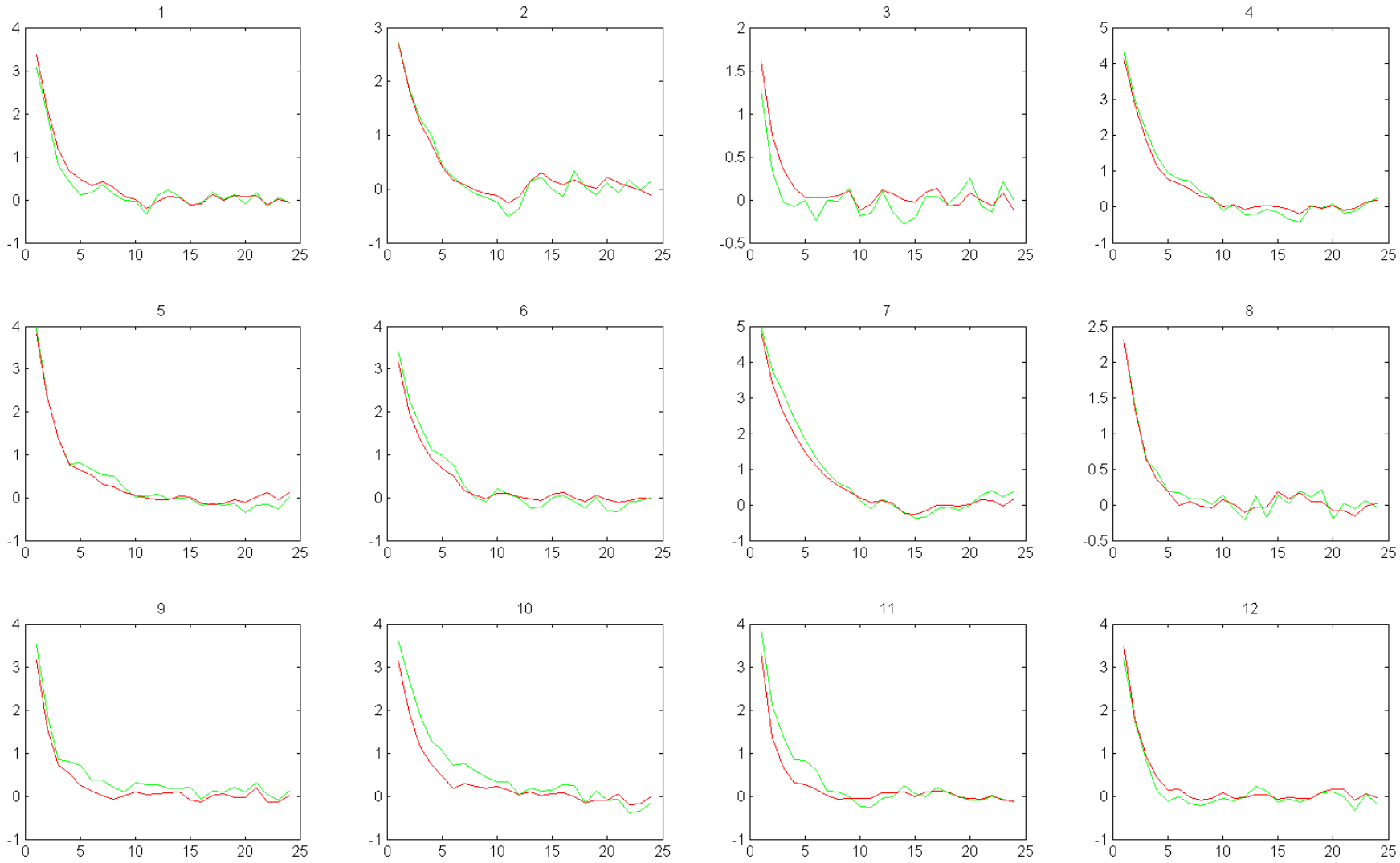
– true concentrations $(Cc_i)$ ; – estimated concentrations $(\widehat{Cc_i})$



# A pharmacokinetic (PK) example

## Estimation of the dynamical system

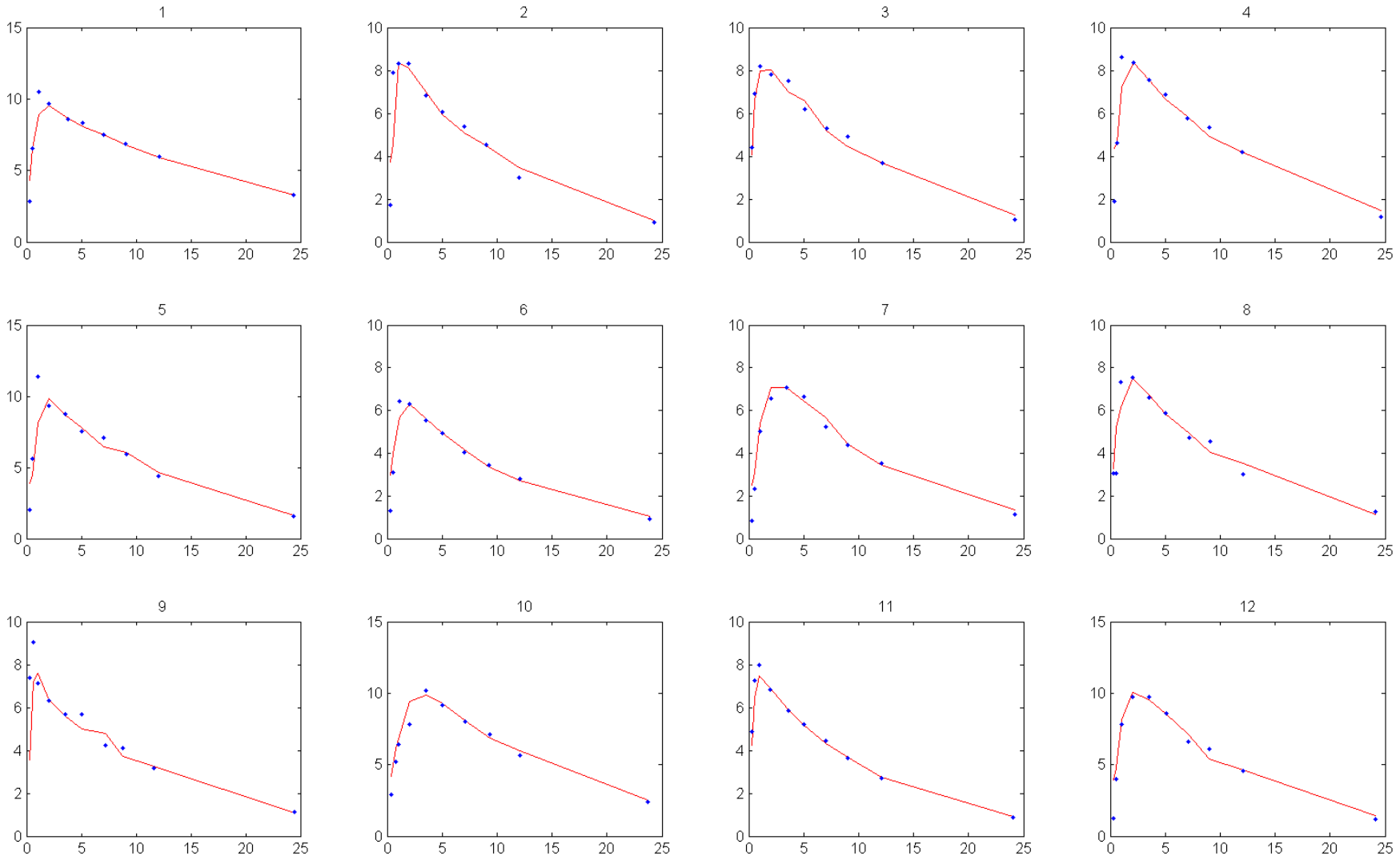
– true amounts ($Qa_i$) ; – estimated amounts ($\widehat{Qa_i}$)



# Application to the PK of theophylline

## Estimation of the concentration

· observed concentrations ($y_{ij}$) ; – estimated concentrations ($\widehat{Cc}_i$)



# A pharmacokinetic (PK) example

## Extension to non linear state-space model

$$\begin{aligned} \log(X_{i,j+1}) &= \log\left(e^{A_i(t_{i,j+1}-t_{ij})}X_{ij}\right) + U_{i,j+1} \\ \log(y_{ij}) &= [0\ 1]\log(X_{ij}) + \sigma_i e_{ij} \end{aligned}$$

where

$$X_{ij} = (Qa_{ij}, Cc_{ij})'$$

Let

$$\begin{aligned} \tilde{X}_{ij} &= \log(X_{i,j+1}), \\ \tilde{y}_{ij} &= \log(y_{ij}), \\ F_{ij} &= e^{A_i(t_{i,j+1}-t_{ij})}. \end{aligned}$$

Then,

$$\begin{aligned} \tilde{X}_{i,j+1} &= \log\left(F_{ij}e^{\tilde{X}_{ij}}\right) + U_{i,j+1} \\ \tilde{y}_{ij} &= [0\ 1]\tilde{X}_{ij} + \sigma_i e_{ij} \end{aligned}$$

# A pharmacokinetic (PK) example

## Extension to non linear state-space model

$$\begin{aligned} \tilde{X}_{i,j+1} &= \log\left(F_{ij}\, e^{\tilde{X}_{ij}}\right) + U_{i,j+1} \\ \tilde{y}_{ij} &= [0\ 1]\tilde{X}_{ij} + \sigma_i\, e_{ij} \end{aligned}$$

$N = 50$
$(t_{ij}) = (1, 2, 3, \ldots, 24)$

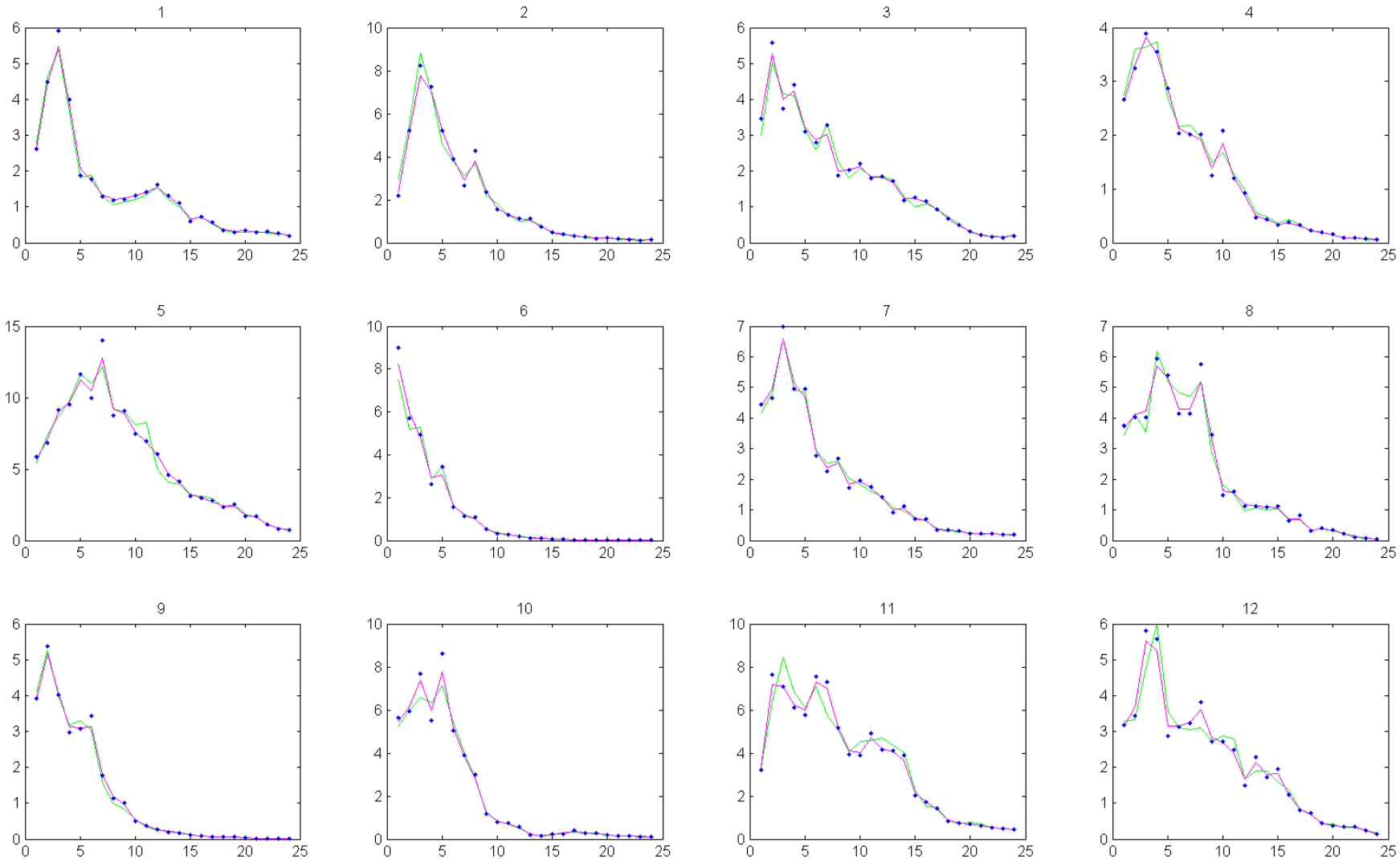
$ka_{\text{pop}} = 0.5 \quad , \quad V_{\text{pop}} = 0.5 \quad , \quad Cl_{\text{pop}} = 0.1$
$\omega_{ka} = 0.3 \quad , \quad \omega_V = 0.1 \quad , \quad \omega_{Cl} = 0.3$

$\gamma_1 = 0.1 \quad , \quad \gamma_2 = 0.2 \quad , \quad \sigma = 0.1$
$\omega_{\gamma_1} = \omega_{\gamma_2} = \omega_\sigma = 0$

# A pharmacokinetic (PK) example

## Estimation of the dynamical system

– true concentrations ($Cc_i$) ; – estimated concentrations ($\widehat{Cc_i}$)



# Outline

# The MONOLIX Group

MOdèles NOn LInéaires à effets miXtes

Multi-disciplinary group led by France Mentré (INSERM & Univ. Paris-Diderot) and Marc Lavielle (INRIA & Univ. Paris-Sud)

The objectives of the group are multiple:

- develop new methodologies for mixed effects models **(NLMEM defined by ODEs or SDEs, categorical data, count data, HMM, time-to-event,...)**
- apply these methodologies to realistic problems **(PKPD, viral dynamics, epilepsy,...)**
- implement these methodologies in the MONOLIX software, **a free software available to the whole community**.

# The Monolix Software

## Main Menu

- Home
- News
- V 3.1 New Features
- The Monolix Project
- Trainings
- The Monolix Group
- Press Room
- Contact the Team

## Resources

- Download Monolix 3.1
- Evaluation
- Demos & Tutorials
- Some Papers

## The Community

- Forum

**MONOLIX** is a free software dedicated to the analysis of non linear mixed effects models. The objective of this software is to perform:

**Parameter estimation,**

**Model selection,**

**Goodness of fit plots,**

**Data simulation.**

*More...*

*This software was developed by INRIA (with the valuable help of several members of the MONOLIX Group!). More...*

## News !

AAPS 2010 : Outstanding Manuscript Award in Modeling and Simulation.

**AstraZeneca** joins the **Monolix** project!

2nd Annual Population PK/PD, 23-24 Sept. 2010


Members of the Monolix Software Project:

Johnson&Johnson PHARMACEUTICAL RESEARCH & DEVELOPMENT · Roche · sanofi aventis · Exprimo · NOVARTIS · AstraZeneca

# SAEM, a reference algorithm for NLMEM

SAEM was first implemented in the MONOLIX software.

This stochastic algorithm becomes a reference algorithm in the field of population pharmacology modelling.

Indeed, SAEM is now available in:

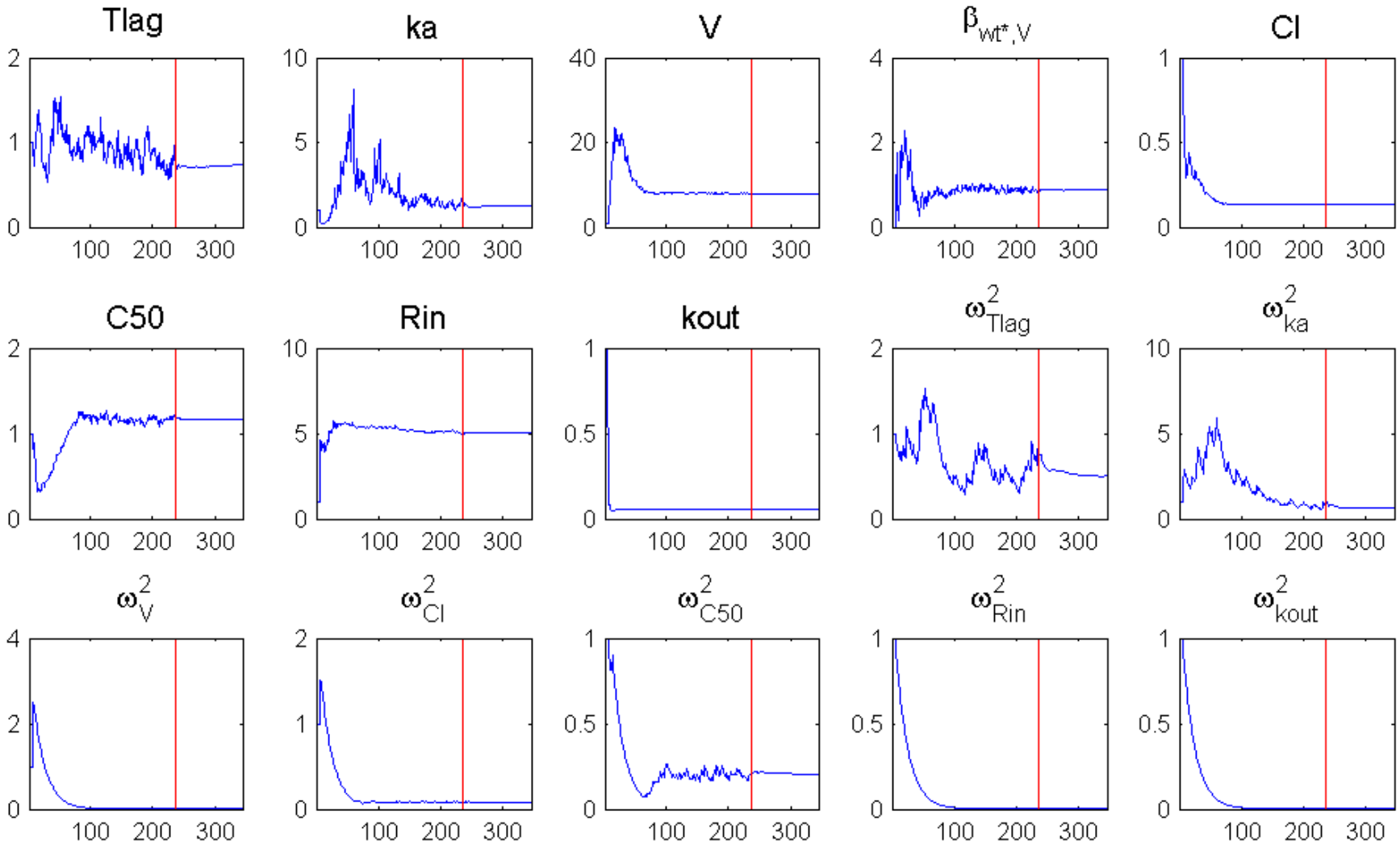
- NONMEM 7 (the gold standard in this field)
- PHOENIX NLME
- MATLAB 10a (nlmefitsa.m, Statistics toolbox)

A R version of SAEM will be available soon.

# Outline

# Convergence of SAEM in the practice

looking at the sequence $(\theta_k)$



# Convergence of SAEM in the practice

$$
\begin{aligned}
\psi^{(k)} &\sim p(\psi|y;\theta_{k-1}) \\
Q_k(\theta) &= Q_{k-1}(\theta) + \gamma_k \left( \log p(y,\psi^{(k)};\theta) - Q_{k-1}(\theta) \right) \\
\theta_k &= \underset{\theta}{\text{Arg max}} \; Q_k(\theta)
\end{aligned}
$$

The algorithm is decomposed into 2 stages:

1) $\gamma_\mathbf{k} = \mathbf{1}$ **during** $\mathbf{K_1}$ **iterations.** Then, SAEM reduces to

$$\theta_k = \underset{\theta}{\text{Arg max}} \; \log p(y,\psi^{(k)};\theta)$$

$(\theta_k)$ is a Markov chain which "quickly converges" to a neighborhood of the solution, "avoiding" local maxima.

2) $\gamma_\mathbf{k} = \mathbf{1}/(\mathbf{k} - \mathbf{K1})$ **during** $\mathbf{K_2}$ **iterations.** $(\theta_k)$ "quickly converges" *a.s.* to the solution, starting from "not too far".

# Convergence of SAEM in the practice

## First stage of the algorithm

$$\begin{aligned} \psi^{(k)} &\sim p(\psi|y;\theta_{k-1}) \\ \theta_k &= \underset{\theta}{\text{Arg max}}\ \log p(y,\psi^{(k)};\theta) \end{aligned}$$

This algorithm is very efficient in a **population context**:

1) draw $\psi_1^{(k)}, \psi_2^{(k)}, \ldots, \psi_N^{(k)}$ with the $N$ different conditional distributions $p(\psi|y_1;\theta_{k-1})$, $p(\psi|y_2;\theta_{k-1})$, $\ldots p(\psi|y_N;\theta_{k-1})$,
2) consider these $N$ simulated vectors of individual parameters as *i.i.d.* realizations of the same distribution $p(\psi;\theta_k)$.

"fixed-point search algorithm"

$$\begin{aligned} p(\psi;\theta^\star) &= \int p(\psi|y;\theta^\star)p(y;\theta^\star)dy \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} p(\psi|y_i;\theta^\star) \end{aligned}$$

# Convergence of SAEM in the practice

## First stage of the algorithm

$$\begin{aligned} \psi^{(k)} &\sim p(\psi|y;\theta_{k-1}) \\ \theta_k &= \underset{\theta}{\text{Arg max}}\ \log p(y,\psi^{(k)};\theta) \end{aligned}$$

This algorithm is very efficient in a **population context**:

1) draw $\psi_1^{(k)}, \psi_2^{(k)}, \ldots, \psi_N^{(k)}$ with the $N$ different conditional distributions $p(\psi|y_1;\theta_{k-1})$, $p(\psi|y_2;\theta_{k-1})$, $\ldots p(\psi|y_N;\theta_{k-1})$,
2) consider these $N$ simulated vectors of individual parameters as *i.i.d.* realizations of the same distribution $p(\psi;\theta_k)$.

"fixed-point search algorithm"

$$\begin{aligned} p(\psi;\theta^\star) &= \int p(\psi|y;\theta^\star)p(y;\theta^\star)dy \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} p(\psi|y_i;\theta^\star) \end{aligned}$$

# Convergence of SAEM in the practice

## First stage of the algorithm

$$\begin{aligned} \psi^{(k)} &\sim p(\psi|y;\theta_{k-1}) \\ \theta_k &= \underset{\theta}{\text{Arg max}}\ \log p(y,\psi^{(k)};\theta) \end{aligned}$$

This algorithm is very efficient in a **population context**:

1) draw $\psi_1^{(k)}, \psi_2^{(k)}, \ldots, \psi_N^{(k)}$ with the $N$ different conditional distributions $p(\psi|y_1;\theta_{k-1})$, $p(\psi|y_2;\theta_{k-1})$, $\ldots p(\psi|y_N;\theta_{k-1})$,
2) consider these $N$ simulated vectors of individual parameters as *i.i.d.* realizations of the same distribution $p(\psi;\theta_k)$.

"fixed-point search algorithm"

$$\begin{aligned} p(\psi;\theta^\star) &= \int p(\psi|y;\theta^\star)p(y;\theta^\star)dy \\ &\simeq \frac{1}{N}\sum_{i=1}^{N} p(\psi|y_i;\theta^\star) \end{aligned}$$

# Convergence of SAEM in the practice
## Second stage of the algorithm

Assumptions on the model:

i) The complete model belongs to the (curved) exponential family:

$$\log p(y, \psi; \theta) = a(y, \psi) + b(\theta) + \langle S(y, \psi); h(\theta) \rangle$$

ii) There exists a function $\hat{\theta}$: $\mathcal{S} \to \Theta$

$$\hat{\theta}(s) = \operatorname{Arg} \max_{\theta} \left( b(\theta) + \langle s; h(\theta) \rangle \right)$$

Example:

$$\begin{aligned} y_i &\sim p(y_i|\psi_i) \\ \psi_i &\sim \mathcal{N}(\beta, \Omega) \end{aligned}$$

Here, $\theta = (\beta, \Omega)$ and

$$\log p(y, \psi; \theta) = \sum_{i=1}^{N} \log p(y_i, \psi_i) - \frac{N}{2} \log |\Omega| + \sum_{i=1}^{N} (\psi_i - \beta)' \Omega^{-1} (\psi_i - \beta)$$

# Convergence of SAEM in the practice
## Second stage of the algorithm

Then, stochastic approximation and maximization steps

$$\begin{aligned} Q_k(\theta) &= Q_{k-1}(\theta) + \gamma_k \left( \log p(y, \psi^{(k)}; \theta) - Q_{k-1}(\theta) \right) \\ \theta_k &= \operatorname{Arg}\max_{\theta} \; Q_k(\theta) \end{aligned}$$

reduce to

$$\begin{aligned} s_k &= s_{k-1} + \gamma_k \left( S(y, \psi^{(k)}) - s_{k-1} \right) \\ \theta_k &= \hat{\theta}(s_k) \end{aligned}$$

# Convergence of SAEM in the practice
## Second stage of the algorithm

Example:

$$\begin{aligned} y_i &\sim p(y_i|\psi_i) \\ \psi_i &\sim \mathcal{N}(\beta, \Omega) \end{aligned}$$

$$\log p(y, \psi; \theta) = \sum_{i=1}^{N} \log p(y_i, \psi_i) - \frac{N}{2} \log |\Omega| - \frac{1}{2} \sum_{i=1}^{N} (\psi_i - \beta)' \Omega^{-1} (\psi_i - \beta)$$

Here,

$$\begin{aligned} s_{1,k} &= s_{1,k-1} + \gamma_k \left( \sum_{i=1}^{N} \psi_i - s_{1,k-1} \right) \\ s_{2,k} &= s_{2,k-1} + \gamma_k \left( \sum_{i=1}^{N} \psi_i \psi' - s_{2,k-1} \right) \end{aligned}$$

and

$$\hat{\beta}_k = \frac{s_{1,k}}{N} \quad ; \quad \hat{\Omega}_k = \frac{s_{2,k}}{N} - \left( \frac{s_{1,k}}{N} \right) \left( \frac{s_{1,k}}{N} \right)'$$

# Convergence of SAEM in the practice

## Second stage of the algorithm

"Realistic" models usually do not belong to the exponential family.

Example (continuous data):

$$\begin{aligned} y_i &\sim \mathcal{N}(f(\psi_i;\alpha),\Sigma) \\ \psi_i &\sim \mathcal{N}(g(C_i;\beta),\Omega) \end{aligned}$$

- $f$ is a (non linear) function of some unknown random individual parameters $\psi_i$ and some **fixed** parameters $\alpha$
- $g$ is a (non linear) function of some known individual covariates $C_i$ and some **fixed** parameters $\beta$

Here, $\theta = (\alpha, \beta, \Omega, \Sigma)$ and

$$\log p(y,\psi;\theta) = a(\theta) - \frac{1}{2}\sum_{i=1}^{N} \|y_i - f(\psi_i;\alpha)\|^2_{\Sigma^{-1}} - \frac{1}{2}\sum_{i=1}^{N} \|\psi_i - g(C_i;\beta))\|^2_{\Omega^{-1}}$$

# Convergence of SAEM in the practice
## Second stage of the algorithm

Stochastic approximation and maximization steps

$$\begin{aligned} Q_k(\theta) &= Q_{k-1}(\theta) + \gamma_k \left( \log p(y, \psi^{(k)}; \theta) - Q_{k-1}(\theta) \right) \\ \theta_k &= \operatorname{Arg}\max_{\theta} \; Q_k(\theta) \end{aligned}$$

are replaced by

$$\begin{aligned} \hat{\theta}(y, \psi^{(k)}) &= \operatorname{Arg}\max_{\theta} \; \log p(y, \psi^{(k)}; \theta) \\ \theta_k &= \theta_{k-1} + \gamma_k \left( \hat{\theta}(y, \psi^{(k)}) - \theta_{k-1} \right) \end{aligned}$$

Open problem: *what can we say about the convergence of this algorithm?*

# Convergence of SAEM in the practice
## Second stage of the algorithm

Stochastic approximation and maximization steps

$$\begin{aligned} Q_k(\theta) &= Q_{k-1}(\theta) + \gamma_k \left( \log p(y, \psi^{(k)}; \theta) - Q_{k-1}(\theta) \right) \\ \theta_k &= \operatorname{Arg}\max_{\theta} \; Q_k(\theta) \end{aligned}$$

are replaced by

$$\begin{aligned} \hat{\theta}(y, \psi^{(k)}) &= \operatorname{Arg}\max_{\theta} \; \log p(y, \psi^{(k)}; \theta) \\ \theta_k &= \theta_{k-1} + \gamma_k \left( \hat{\theta}(y, \psi^{(k)}) - \theta_{k-1} \right) \end{aligned}$$

Open problem: *what can we say about the convergence of this algorithm?*

# Conclusion

- SAEM has shown to be very useful in the field of population pharmacology (continuous PKPD data, viral dynamics, count data, categorical data, HMM, survival data, ...)
- SAEM was successfully used in several other applications (deformable models, animal breeding, agronomy, signal processing, ...),
- SAEM becomes a reference algorithm for non linear mixed effects models (usually "better" than linearization, Laplace, Gaussian quadrature, ...),
- SAEM can be fast (requires an optimal implementation of several stochastic algorithms: MCMC, simulated annealing, stochastic approximation,...),
- Convergence of SAEM demonstrated under "restrictive conditions" (exponential family). Theoretical results for more general models would be welcome!