Bayesian ODE-penalized B-spline model with Gaussian mixture as error distribution

Jonathan Jaeger [1] Philippe Lambert [2,1]

[1]Institut de Statistique, Biostatistique et Sciences Actuarielles,
Université Catholique de Louvain, Belgium

[2]Institut des Sciences Humaines et Sociales,
Méthodes Quantitatives en Sciences Sociales,
Université de Liège, Belgium

June 6, 2012

PEDS II, Eindhoven

Motivating example

Theophylline dataset

Data for the kinetics of the anti-asthmatic drug theophylline with 12 subjects, concentration were measured at 11 time point over 25 hours for each subject.

Differential equation

$$\begin{cases} \dfrac{dx_1}{dt}(t) &= -k_a x_1(t), \\ \dfrac{dx_2}{dt}(t) &= \dfrac{k_a}{V} x_1(t) - k_e x_2(t), \\ x_1(0) &= D \\ x_2(0) &= 0. \end{cases}$$

- $x_1(t)$ quantity of drug in stomach,
- $x_2(t)$ concentration of drug in blood.

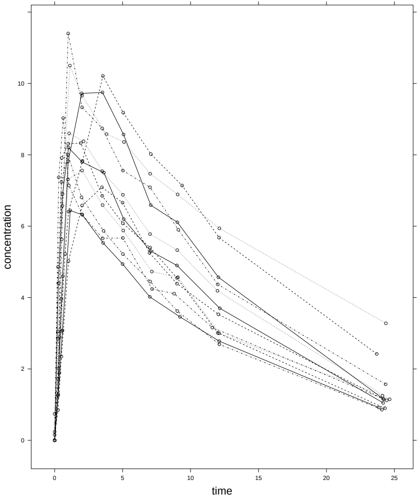


Figure : Plasma concentration of theophylline over time

Introduce the

Bayesian ODE-penalized B-spline approach

in the case of systems of affine differential equations

when the conditional error distribution is a

penalized mixture of Gaussian distributions

Theoretical part

Differential equation and observations

Differential equation

$$\begin{cases} \frac{d\mathbf{x}}{dt}(t) & = & \mathbf{f}(\mathbf{x}(t), \theta) \\ \mathbf{x}(0) & = & \mathbf{x_0} \end{cases}$$

with:
- $(\mathbf{x}(t))^T = (x_1(t), \ldots, x_d(t))$ the set of d state functions and $\mathbf{x_0}$ the set of initial conditions,
- $\theta \in \mathbb{R}^q$ the vector of unknown parameters,
- $\mathbf{f}$ a known affine function of $\mathbf{x}$.

Differential equation and observations

Differential equation

$$\begin{cases} \dfrac{d\mathbf{x}}{dt}(t) & = & \mathbf{f}(\mathbf{x}(t), \theta) \\ \mathbf{x}(0) & = & \mathbf{x_0} \end{cases}$$

with:
- $(\mathbf{x}(t))^T = (x_1(t), \ldots, x_d(t))$ the set of d state functions and $\mathbf{x_0}$ the set of initial conditions,
- $\theta \in \mathbb{R}^q$ the vector of unknown parameters,
- $\mathbf{f}$ a known affine function of $\mathbf{x}$.

Measurement

A subset $\mathcal{J}$ of the d state functions are observed with additive measurement errors:

$$y_{jk} = x_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}$$

with $\mathbb{E}(\epsilon_j) = 0$ and $\mathbb{V}(\epsilon_j) = 1$.

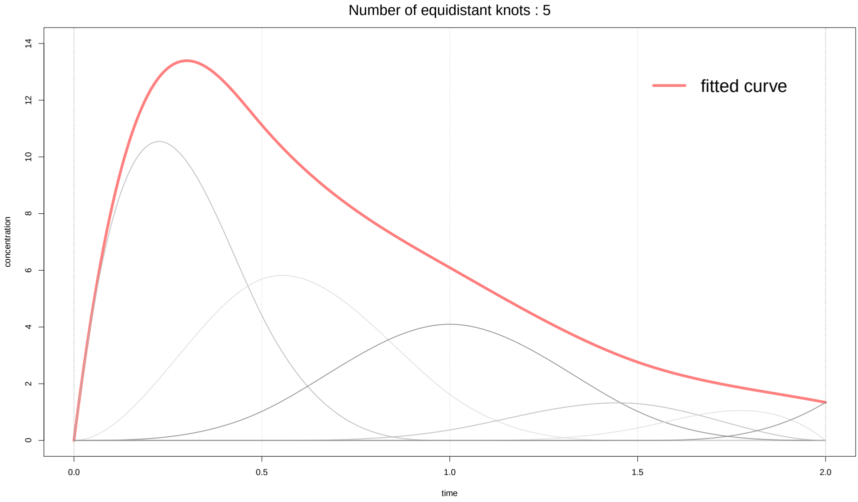
B-spline approximation

Each state function $x_j(t)$ is approximated using a B-spline basis function expansion:

$$
\begin{aligned}
\widetilde{x}_j(t) &= \sum_{k=1}^{K_j} B_{jk}(t)\, c_{jk} \\
&= \left(\mathbf{B_j}(t)\right)^T \mathbf{c_j}
\end{aligned}
$$

where:
- $\mathbf{B_j}(t)$ is the vector of B-spline basis function evaluated at time t,
- $\mathbf{c_j}$ is the vector of spline coefficients.

Influence of the number of knots and the spline coefficients



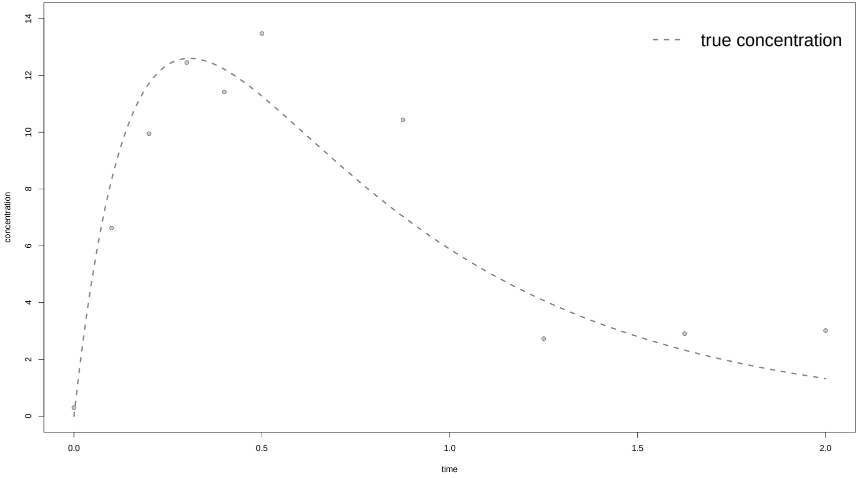
ODE-penalty (Ramsay et al., 2007)

The penalty for the jth equation quantifies the proximity of the approximation $\widetilde{x}_j(t)$ to the jth component of the solution $\mathbf{x}(t)$:

$$\begin{aligned} PEN_j &= \int \left(\frac{d\widetilde{x}_j}{dt}(t) - f_j(\widetilde{\mathbf{x}}(t), \theta) \right)^2 dt \\ PEN &= \sum_{j=1}^{d} \gamma_j PEN_j \\ &= \mathbf{c}^T \mathbf{R}(\theta, \gamma) \mathbf{c} + 2\mathbf{c}^T \mathbf{r}(\theta, \gamma) + l(\theta, \gamma) \end{aligned}$$

where:
- $\mathbf{c}^T = \left(\mathbf{c_1}^T, \ldots, \mathbf{c_d}^T\right)$ is the vector of spline coefficients,
- $\gamma^T = (\gamma_1, \ldots, \gamma_d)$ is the vector of ODE-adhesion parameters.

Influence of the ODE-adhesion parameter



Prior distribution for the spline coefficients & the ODE-adhesion parameters

Prior distribution for the spline coefficients (Jaeger and Lambert, 2011)

$$p(\mathbf{c}|\theta, \gamma) \quad \propto \quad \exp\left(-\frac{1}{2}\sum_{j=1}^{d} \gamma_j PEN_j\right)$$

Prior distribution for the spline coefficients & the ODE-adhesion parameters

Prior distribution for the spline coefficients (Jaeger and Lambert, 2011)

$$p\left(\mathbf{c}|\theta,\gamma\right) \quad \propto \quad \exp\left(-\frac{1}{2}\sum_{j=1}^{d}\gamma_j PEN_j - \frac{1}{2}\left(\mathbf{c}-\mu_{\mathbf{c}}\right)^T \mathbf{P_c}\left(\mathbf{c}-\mu_{\mathbf{c}}\right)\right)$$

with:

- $\mu_{\mathbf{c}}$ and $\mathbf{P_c}$ to express uncertainty w.r.t initial conditions of the states,

Prior distribution for the spline coefficients & the ODE-adhesion parameters

Prior distribution for the spline coefficients (Jaeger and Lambert, 2011)

$$p\left(\mathbf{c}|\theta,\gamma\right) \quad \propto \quad \exp\left(-\frac{1}{2}\left\{\mathbf{c}^T\left(\mathbf{R}\left(\theta,\gamma\right)+\mathbf{P_c}\right)\mathbf{c}-2\mathbf{c}^T\left(-\mathbf{r}\left(\theta,\gamma\right)+\mathbf{P_c}\mu_\mathbf{c}\right)\right\}\right)$$

with:

- $\mu_\mathbf{c}$ and $\mathbf{P_c}$ to express uncertainty w.r.t initial conditions of the states,

Prior distribution for the spline coefficients & the ODE-adhesion parameters

Prior distribution for the spline coefficients (Jaeger and Lambert, 2011)

$$p\left(\mathbf{c}|\theta,\gamma\right) \quad \propto \quad \exp\left(-\frac{1}{2}\left\{\mathbf{c}^T\mathbf{V_1}\mathbf{c}-2\mathbf{c}^T\mathbf{v_1}\right\}\right)$$

with:

- $\mu_{\mathbf{c}}$ and $\mathbf{P_c}$ to express uncertainty w.r.t initial conditions of the states,
- $\mathbf{v_1}=-\mathbf{r}\left(\theta,\gamma\right)+\mathbf{P_c}\mu_{\mathbf{c}}$ and $\mathbf{V_1}=\mathbf{R}\left(\theta,\gamma\right)+\mathbf{P_c}$,

Prior distribution for the spline coefficients & the ODE-adhesion parameters

Prior distribution for the spline coefficients (Jaeger and Lambert, 2011)

$$p(\mathbf{c}|\theta,\gamma) \quad \propto \quad (\det(\mathbf{V_1}))^{1/2} \exp\left(-\frac{1}{2}\left\{\mathbf{c}^T\mathbf{V_1}\mathbf{c} - 2\mathbf{c}^T\mathbf{v_1} + \mathbf{v_1}^T\mathbf{V_1}^{-1}\mathbf{v_1}\right\}\right)$$

with:
- $\mu_{\mathbf{c}}$ and $\mathbf{P_c}$ to express uncertainty w.r.t initial conditions of the states,
- $\mathbf{v_1} = -\mathbf{r}(\theta,\gamma) + \mathbf{P_c}\mu_{\mathbf{c}}$ and $\mathbf{V_1} = \mathbf{R}(\theta,\gamma) + \mathbf{P_c}$,

Prior distribution for the spline coefficients & the ODE-adhesion parameters

Prior distribution for the spline coefficients (Jaeger and Lambert, 2011)

$$p(\mathbf{c}|\theta,\gamma) \quad \propto \quad (\det(\mathbf{V_1}))^{1/2} \exp\left(-\frac{1}{2}\left\{\mathbf{c}^T\mathbf{V_1}\mathbf{c} - 2\mathbf{c}^T\mathbf{v_1} + \mathbf{v_1}^T\mathbf{V_1}^{-1}\mathbf{v_1}\right\}\right)$$

with:
- $\mu_{\mathbf{c}}$ and $\mathbf{P_c}$ to express uncertainty w.r.t initial conditions of the states,
- $\mathbf{v_1} = -\mathbf{r}(\theta,\gamma) + \mathbf{P_c}\mu_{\mathbf{c}}$ and $\mathbf{V_1} = \mathbf{R}(\theta,\gamma) + \mathbf{P_c}$,

Prior distribution for the ODE-adhesion parameters

$$\gamma_j \sim \mathcal{E}xp\left(10^{-8}\right)$$

It indicates the prior confidence on the proposed ODE to model the dynamic of the state $\mathbf{x}(t)$.

Specification of the conditional distribution

Conditional error distribution (Komárek and Lesaffre, 2008)

$$\epsilon_{jk}|\mathbf{d_j} \sim \sum_{l=-L_j}^{L_j} \pi_{jl}\left(\mathbf{d_j}\right)\mathcal{N}\left(\mu_{jl};\sigma_j^2\right)$$

where $(\mu_{jl})_{l=-L_j}^{L_j}$ is a fixed large number of pre-selected equidistant means and σ_j^2 is a fixed variance, with constraints $\pi_{jl} > 0$ and $\sum_l \pi_{jl} = 1$.

Specification of the conditional distribution

Conditional error distribution (Komárek and Lesaffre, 2008)

$$\epsilon_{jk}|\mathbf{d_j} \sim \sum_{l=-L_j}^{L_j} \pi_{jl}\left(\mathbf{d_j}\right) \mathcal{N}\left(\mu_{jl}; \sigma_j^2\right)$$

where $(\mu_{jl})_{l=-L_j}^{L_j}$ is a fixed large number of pre-selected equidistant means and σ_j^2 is a fixed variance, with constraints $\pi_{jl} > 0$ and $\sum_l \pi_{jl} = 1$.

Generalized logit transformation of the mixture weights

$$\pi_{jl} = \frac{\exp\left(a_{j,l}\right)}{\sum_{k=-L_j}^{L_j} \exp\left(a_{j,k}\right)}$$

with identifiability constraints $a_{j,L_j} = 0$ for all $j \in \mathcal{J}$.
Only $\mathbf{d_j} = \mathrm{vec}\left(a_{j,l}; l = -L_j, \ldots, L_j - 1\right)$, $j \in \mathcal{J}$ have to be estimated to estimate the error distributions.

Fitted error densities

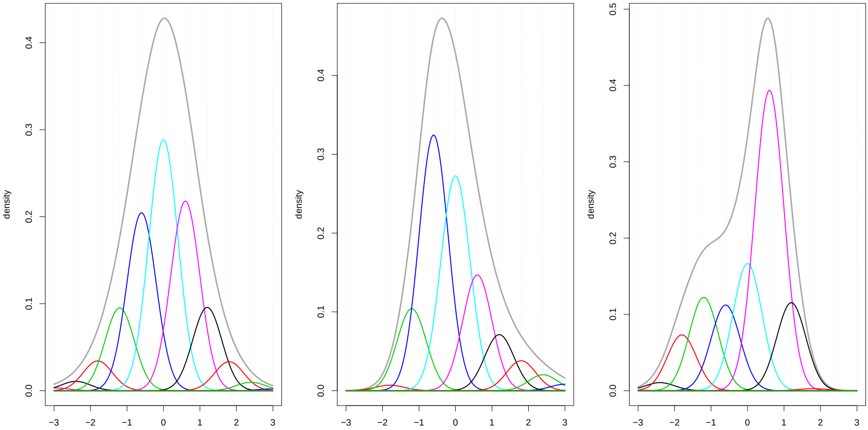


Figure : Fitted error densities (Student, extreme value and mixture of 2 Gaussians)

Prior distribution for the (transformed) weights and the roughness penalty parameters

Finite difference penalty on the transformed weights

$$
\begin{aligned}
Q\left(\mathbf{a_j} \mid \lambda_j\right) &= -\frac{\lambda_j}{2} \sum_{l} \left(a_{j,l} - 3a_{j,l-1} + 3a_{j,l-2} - a_{j,l-3}\right)^2 \\
&= -\frac{\lambda_j}{2} \left(\mathbf{D a_j}\right)^T \mathbf{D a_j}
\end{aligned}
$$

Prior distribution for the (transformed) weights and the roughness penalty parameters

Finite difference penalty on the transformed weights

$$\begin{aligned} Q\left(\mathbf{a_j}|\lambda_j\right) &= -\frac{\lambda_j}{2}\sum_l \left(a_{j,l} - 3a_{j,l-1} + 3a_{j,l-2} - a_{j,l-3}\right)^2 \\ &= -\frac{\lambda_j}{2}\left(\mathbf{D}\mathbf{a_j}\right)^T \mathbf{D}\mathbf{a_j} \\ \Rightarrow Q\left(\mathbf{d_j}|\lambda_j\right) &= -\frac{\lambda_j}{2}\mathbf{d_j}^T\mathbf{P}\mathbf{d_j} \end{aligned}$$

where:
- λ_j, $j \in \mathcal{J}$ is the roughness penalty parameter,
- $\mathbf{P}$ is the submatrix of $\mathbf{D}^T\mathbf{D}$ without the last line and last column.

Prior distribution for the (transformed) weights and the roughness penalty parameters

Prior distribution for the transformed weights $\mathbf{d_j}$

$$p\left(\mathbf{d_j}|\lambda_j\right) \propto \exp\left(-\frac{\lambda_j}{2}\mathbf{d_j}^T\mathbf{P}\mathbf{d_j}\right)$$

Prior distribution for the (transformed) weights and the roughness penalty parameters

Prior distribution for the transformed weights $\mathbf{d_j}$

$$p\left(\mathbf{d_j}|\lambda_j\right) \propto \lambda_j^{L_j-1} \exp\left(-\frac{\lambda_j}{2}\mathbf{d_j}^T\mathbf{P}\mathbf{d_j}\right)$$

Prior distribution for the (transformed) weights and the roughness penalty parameters

Prior distribution for the transformed weights $\mathbf{d_j}$

$$p\left(\mathbf{d_j}|\lambda_j\right) \propto \lambda_j^{L_j-1} \exp\left(-\frac{\lambda_j}{2}\mathbf{d_j}^T\mathbf{P}\mathbf{d_j}\right)$$

Prior distribution for the roughness penalty parameters

$$\lambda_j \sim \mathcal{E}xp\left(10^{-8}\right)$$

It indicates the prior confidence on the Gaussian distribution for the error distribution.

Bayesian model (Jaeger and Lambert, 2012a)

$$
\left\{
\begin{array}{rcl}
y_{jk} & = & \widetilde{x}_j\left(t_{jk}\right) + \tau_j^{-1/2}\epsilon_{jk} \\
\epsilon_{jk}|\mathbf{d_j} & \sim & \displaystyle\sum_{l=-L_j}^{L_j} \pi_{jl}\left(\mathbf{d_j}\right)\mathcal{N}\left(\mu_{jl} + \alpha_j; \left(\frac{\sigma_j}{\beta_j}\right)^2\right) \\
p\left(\mathbf{c}|\theta,\gamma\right) & \propto & \left(\det\left(\mathbf{V_1}\right)\right)^{1/2}\exp\left(-\frac{1}{2}\left\{\mathbf{c}^T\mathbf{V_1}\mathbf{c} - 2\mathbf{c}^T\mathbf{v_1} + \mathbf{v_1}^T\mathbf{V_1}^{-1}\mathbf{v_1}\right\}\right) \\
\gamma_j & \sim & \mathcal{E}xp\left(10^{-8}\right) \\
p\left(\mathbf{d_j}|\lambda_j\right) & \propto & \lambda_j^{L_j-1}\exp\left(-\frac{\lambda_j}{2}\mathbf{d_j}^T\mathbf{P_{d,j}}\mathbf{d_j}\right) \\
\lambda_j & \sim & \mathcal{E}xp\left(10^{-8}\right) \\
\tau_j & \sim & \mathcal{G}\left(a_{\tau_j}; b_{\tau_j}\right) \\
\theta & \sim & p\left(\theta\right).
\end{array}
\right.
$$

ODE-model validation (Jaeger and Lambert, 2012b)

If the ODE model is adequate, then one can show that:

$$\lim_{\gamma \to \infty} \left(\log p\left(\gamma | \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{y}\right) - \log p\left(\gamma\right)\right) \doteq 0$$

when no prior information is used for the state functions and one common ODE-adhesion parameter γ for the complete ODE, where:

- $\log\left(p\left(\gamma | \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{y}\right)\right)$ is the log conditional posterior distribution for γ (marginalized w.r.t the spline coefficients),
- $\log\left(p\left(\gamma\right)\right)$ is the log prior distribution for γ.

Application

Theophylinne dataset

Differential equation

$$\begin{cases} \frac{dx_1}{dt}(t) &= -k_a x_1(t), \\ \frac{dx_2}{dt}(t) &= \frac{k_a}{V} x_1(t) - k_e x_2(t), \\ x_1(0) &= D \\ x_2(0) &= 0. \end{cases}$$

- $x_1(t)$ quantity of drug in stomach,
- $x_2(t)$ concentration of drug in blood.

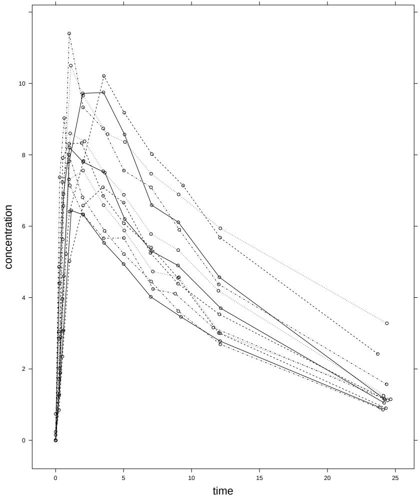


Figure : Plasma concentration of theophylline over time

Theophylinne dataset

Prior information

- The drug cannot be eliminated more quickly that it is absorbed: $k_a > k_e$,
- Total confidence on the initial condition of the state function

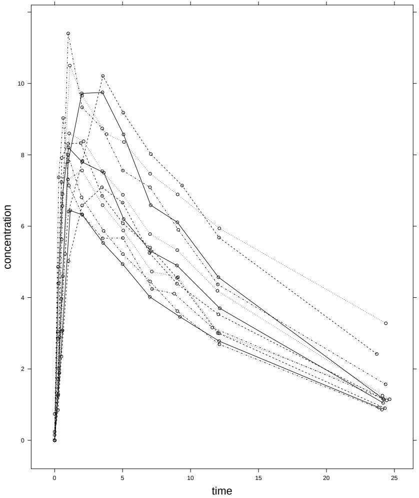


Figure : Plasma concentration of theophylline over time

Fitted error density

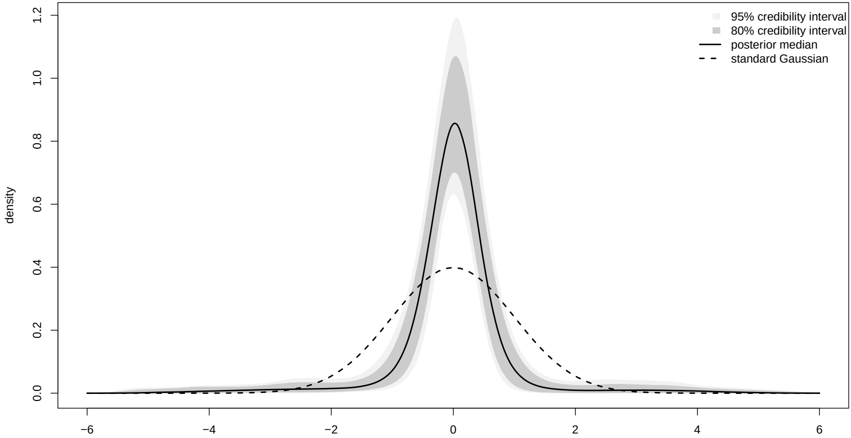


Figure : Fitted error densities with pointwise 80% and 95% credibility interval

Suitability of the ODE-model

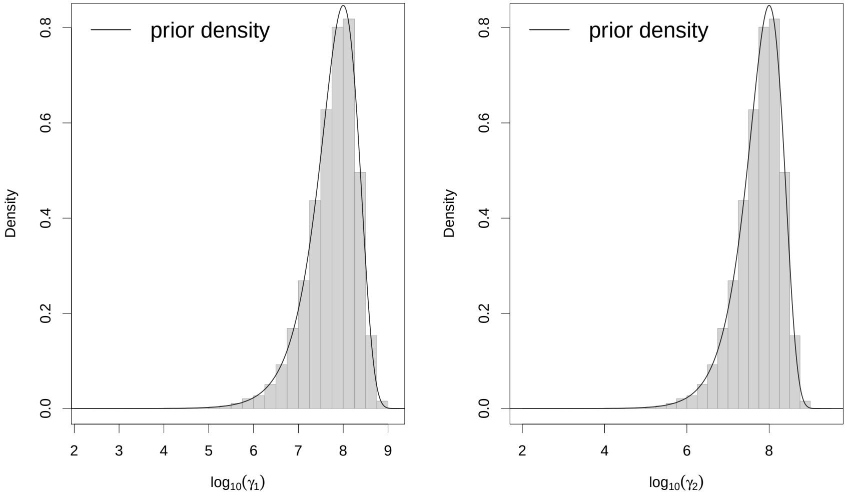


Figure : Histograms of the posterior distribution for γ_1 and γ_2 with prior density (log10 scale)

Posterior credibility intervals for the quantity of drug in stomach

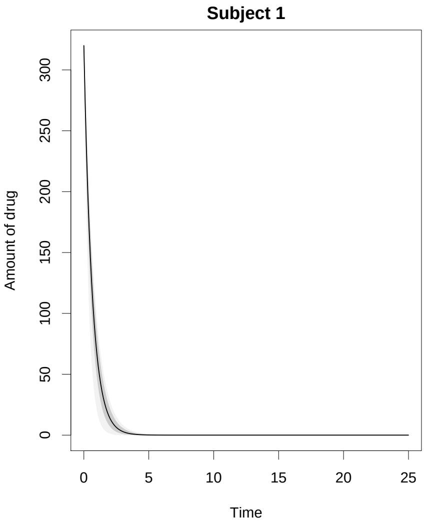


Figure : Pointwise 80% (dark grey) and 95% (light grey) posterior credibility intervals for the quantity of drug in stomach with posterior median

Posterior credibility intervals for the drug concentration in the plasma

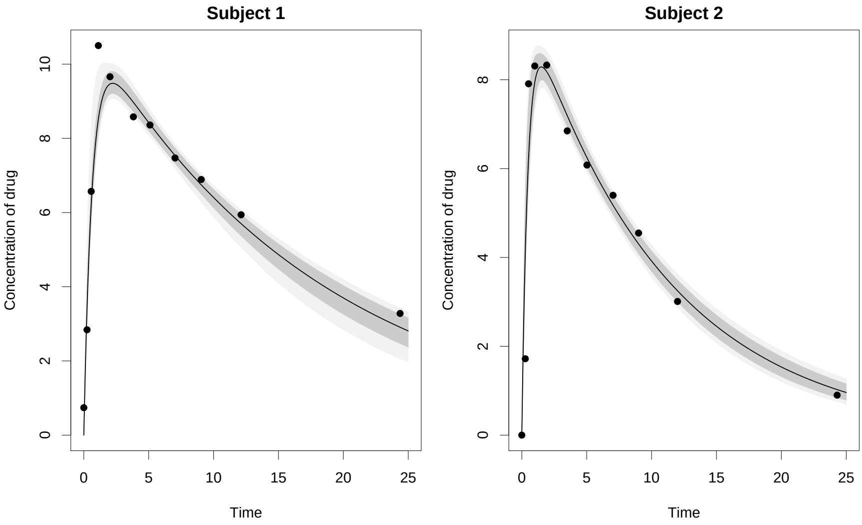


Figure : Pointwise 80% (dark grey) and 95% (light grey) posterior credibility intervals for the drug concentration with posterior median

Conclusion

Conclusion and further work

Conclusion

- Powerful tool that overcomes solving the DE using a numerical method,
- Convenient implementation of the Bayesian ODE-penalized B-spline approach,
- Simple method to include prior information about ODE parameters,
- Possibility to express uncertainty with respect to initial conditions,
- Automatic selection of the ODE-adhesion parameters and of the roughness penalty parameters,
- ODE-model validation,
- Flexible distributional assumption.

Conclusion and further work

Conclusion

- Powerful tool that overcomes solving the DE using a numerical method,
- Convenient implementation of the Bayesian ODE-penalized B-spline approach,
- Simple method to include prior information about ODE parameters,
- Possibility to express uncertainty with respect to initial conditions,
- Automatic selection of the ODE-adhesion parameters and of the roughness penalty parameters,
- ODE-model validation,
- Flexible distributional assumption.

Current/further work

- How to deal with non-homogeneous (non) Gaussian data distribution?
- How to deal with input functions?
- Generalize this approach for nonlinear differential equations.

References

J Jaeger and P Lambert. Bayesian generalized profiling estimation in hierarchical linear dynamic systems. Discussion Paper DP2011/01, Institut de Statistique, Biostatistique et Sciences Actuarielles, 2011.

J Jaeger and P Lambert. Bayesian penalized smoothing approaches in models specified usingaffine differential equations with unknown error distributions. Discussion Paper DP2012/17, Institut de Statistique, Biostatistique et Sciences Actuarielles, 2012a.

J Jaeger and P Lambert. On the use of adhesion parameters to validate models specified using systems of affine differential equations. Discussion Paper DP2012/18, Institut de Statistique, Biostatistique et Sciences Actuarielles, 2012b.

A Komárek and E Lesaffre. Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association*, 103:523–533, 2008.

J Ramsay, G Hooker, D Campbell, and J Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):741–796, 2007.