Logistic Mixed Effects Models

A GLMM for binary data takes the binomial exponential family, with canonical link being logistic.

We have

Stage 1: $Y_{ij} \sim_{ind}$ Binomial(n_{ij}, p_{ij}) with

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i$$

Stage 2: $\boldsymbol{b}_i \sim_{iid} \mathrm{N}(\boldsymbol{0}, \boldsymbol{D})$.

Marginal moments are not available in closed form.

We initially consider the model with a random intercept only, $b_i \sim \mathrm{N}(0, \sigma_0^2)$.

Parameter Interpretation

For the random intercepts model the conditional parameters $\boldsymbol{\beta}^c$ and marginal parameters $\boldsymbol{\beta}^m$ are approximately linked through

$$\begin{aligned}
\mathrm{E}[Y] &= \frac{\exp(\boldsymbol{x}\boldsymbol{\beta}^m)}{1+\exp(\boldsymbol{x}\boldsymbol{\beta}^m)} = \mathrm{E}_b\{\mathrm{E}[Y|b]\} \\
&= \mathrm{E}_b\left[\frac{\exp(\boldsymbol{x}\boldsymbol{\beta}^c + b)}{1+\exp(\boldsymbol{x}\boldsymbol{\beta}^c + b)}\right] \approx \frac{\exp(\boldsymbol{x}\boldsymbol{\beta}^c/[c^2\sigma_0^2+1]^{1/2})}{1+\exp(\boldsymbol{x}\boldsymbol{\beta}^c/[c^2\sigma_0^2+1]^{1/2})}
\end{aligned}$$

where $c = 16\sqrt{3}/(15\pi)$. Hence the marginal coefficients are attenuated towards zero; Figure 21 illustrates for particular values of $\beta_0, \beta_1, \sigma_0^2$.

For the model

$$\log\left(\frac{\mathrm{E}[Y \mid \boldsymbol{b}]}{1-\mathrm{E}[Y \mid \boldsymbol{b}]}\right) = \boldsymbol{x}\boldsymbol{\beta}^c + \boldsymbol{z}\boldsymbol{b}$$

where $\boldsymbol{b} \sim_{iid} \mathrm{N}_{q+1}(\boldsymbol{0}, \boldsymbol{D})$ we obtain

$$\mathrm{E}[Y] \approx \frac{\exp\left(\boldsymbol{x}\boldsymbol{\beta}^c \mid c^2\boldsymbol{D}\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}} + \boldsymbol{I}_{q+1} \mid^{-(q+1)/2}\right)}{1+\exp\left(\boldsymbol{x}\boldsymbol{\beta}^c \mid c^2\boldsymbol{D}\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}} + \boldsymbol{I}_{q+1} \mid^{-(q+1)/2}\right)}$$

so that

$$\boldsymbol{\beta}^m \approx \mid c^2\boldsymbol{D}\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}} + \boldsymbol{I}_{q+1} \mid^{-(q+1)/2} \boldsymbol{\beta}^c$$

The above follows from approximating the logistic CDF by that of a normal CDF. Specifically,

$$G(x) = (1 + \mathrm{e}^{-x})^{-1}$$

is the CDF of a logistic random variable and

$$G(x) \approx \Phi(cx)$$

where $c = 16\sqrt{3}/(15\pi)$ (Johnson and Kotz, 1970).

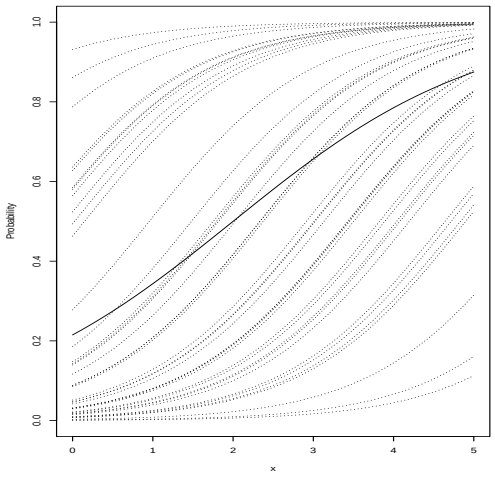

Figure 21: Individual-level curves (dotted lines) from random intercepts logistic GLMM with $\log(\mathrm{E}[Y \mid b])/(1 - \mathrm{E}[Y \mid b])) = \beta_0 + \beta_1 x$, with $\beta_0 = -2, \beta_1 = 1$ and $b \sim_{iid} \mathrm{N}(0, 2^2)$, along with marginal curve (solid curve). Approximate attenuation is 1.54.

Specifying the Parameters of a Wishart Prior

Suppose we have

$$\begin{aligned} \boldsymbol{b}|\boldsymbol{W} &\sim \mathrm{N}_p(\boldsymbol{0}, \boldsymbol{W}) \\ \boldsymbol{W} &\sim \mathrm{Wishart}(r, \boldsymbol{S}) \end{aligned}$$

We derive the marginal distribution for $\boldsymbol{b}$:

$$\begin{aligned} p(\boldsymbol{b}) &= \int p(\boldsymbol{b}|\boldsymbol{W}) \times \pi(\boldsymbol{W}) d\boldsymbol{W} \\ &\propto \int |\boldsymbol{W}|^{1/2} \exp\left(-\frac{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{b}}{2}\right) |\boldsymbol{W}|^{(r+p-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{W}\boldsymbol{S}^{-1})\right) d\boldsymbol{W} \\ &= \int |\boldsymbol{W}|^{(r+p+1-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{W}[\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}} + \boldsymbol{S}^{-1})\right) d\boldsymbol{W} \end{aligned}$$

— the integrand is a $\mathrm{Wishart}_p\{r+1, (\boldsymbol{S}^{-1} + \boldsymbol{b}\boldsymbol{b}^{\mathrm{T}})^{-1}\}$

We know the normalizing constant of a Wishart distribution:

$$\begin{aligned} p(\boldsymbol{b}) &= |\boldsymbol{S}^{-1} + \boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}|^{-(r+1)/2} \\ &\propto |\boldsymbol{I}_p + \boldsymbol{S}\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}}|^{-(r+1)/2} \\ &= (1 + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{b})^{-(r+1)/2} \end{aligned}$$

which is a $T_p\{\boldsymbol{0}, [(r-p+1)\boldsymbol{S}]^{-1}, d = r-p+1\}$ density.

The margins of a multivariate Student's t distribution are t and so we can specify the parameters r and $\boldsymbol{S}$ using the same technique as with the gamma, $\mathrm{Ga}(a, b)$, noting that $a = r/2, b = 1/[2S]$.

Non-linear Mixed Effects Models

We now turn to a class of models that are not GLMs — we begin with a motivating example.

Example: Pharmacokinetics of Theophylline

Twelve subjects given an oral dose of the anti-asthmatic agent theophylline, with 11 concentration measurements obtained from each individual over 25 hours.

Figure 22 shows the concentration-time data.

The curves follow a similar pattern but there is clearly between-subject variability. For these data the the one-compartment model with first-order absorption and elimination is a good starting point for analysis.

The mean concentration at time point t is often modeled as:

$$f(\eta, t) = \frac{D k_e k_a}{Cl(k_a - k_e)} \{\exp(-k_e t) - \exp(-k_a t)\}$$

where D is the initial dose.

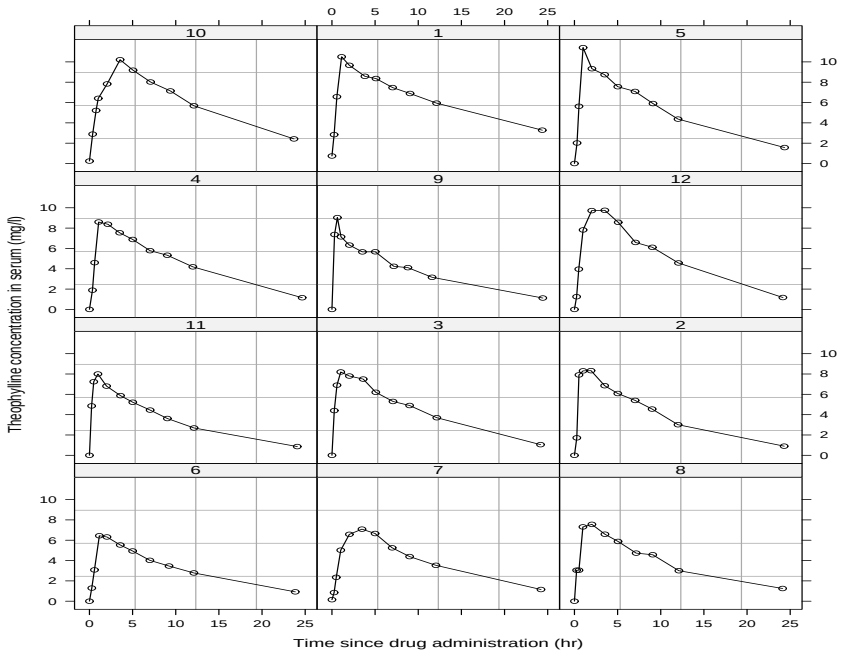

Figure 22: Concentration time data for Theophylline.

Non-Linear Mixed Effects Model Structure

In a nonlinear mixed model (NLMEM) the first stage of a linear mixed model is replaced by a nonlinear form. We describe a specific two-stage form that is useful in many longitudinal situations.

Stage 1: Response model, *conditional* on random effects, $\boldsymbol{b}_i$:

$$\boldsymbol{y}_i = f_{ij}(\eta_{ij}, t_{ij}) + \epsilon_{ij}, \tag{48}$$

where f_{ij} is a nonlinear function and

$$\eta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i,$$

where

- a $(k+1) \times 1$ vector of fixed effects, $\boldsymbol{\beta}$,
- a $(q+1) \times 1$ vector of random effects, $\boldsymbol{b}_i$, with $q \leq k$.
- $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i})^{\mathrm{T}}$, the design matrix for the fixed effect with $\boldsymbol{x}_{ij} = (1, x_{ij1}, ..., x_{ijk})^{\mathrm{T}}$, and
- $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{in_i})^{\mathrm{T}}$, and design matrix for the random effects with $\boldsymbol{z}_{ij} = (1, z_{ij1}, ..., z_{ijq})^{\mathrm{T}}$.

Stage 2: Model for random terms:

$$\begin{aligned} \mathrm{E}[\boldsymbol{\epsilon}_i] &= \mathbf{0}, \quad \mathrm{var}(\boldsymbol{\epsilon}_i) = \boldsymbol{E}_i(\boldsymbol{\alpha}), \\ \mathrm{E}[\boldsymbol{b}_i] &= \mathbf{0}, \quad \mathrm{var}(\boldsymbol{b}_i) = \boldsymbol{D}(\boldsymbol{\alpha}), \\ \mathrm{cov}(\boldsymbol{b}_i, \boldsymbol{\epsilon}_i) &= \mathbf{0} \end{aligned}$$

where $\boldsymbol{\alpha}$ is the vector of variance-covariance parameters.

A common model assumes

$$\boldsymbol{\epsilon}_i \sim_{ind} N(\mathbf{0}, \sigma^2_\epsilon \boldsymbol{I}_{n_i}), \quad \boldsymbol{b}_i \sim_{iid} N(\mathbf{0}, \boldsymbol{D}).$$

Let $\boldsymbol{\alpha}$ represent σ^2_ϵ and the parameters of $\boldsymbol{D}$ and $N = \sum_i n_i$.

Likelihood Inference for the Nonlinear Mixed Effects Model

As with the linear mixed and generalized linear mixed models already considered the likelihood is defined with respect to fixed effects $\boldsymbol{\beta}$ and variance components $\boldsymbol{\alpha}$:

$$p(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\alpha}) = \prod_{i=1}^{m} \int_{\boldsymbol{b}_i} p(\boldsymbol{y}_i|\boldsymbol{b}_i,\boldsymbol{\beta},\sigma^2_\epsilon) \times p(\boldsymbol{b}_i|\boldsymbol{D})\ d\boldsymbol{b}_i$$

The first difficulty is how to calculate the required integrals, which for non-linear models are analytically intractable, recall for linear models they were available in closed form. For nonlinear models even the first two moments are not available in closed form in general:

$$\begin{aligned}
\mathrm{E}[Y_{ij} \mid \boldsymbol{\beta},\boldsymbol{\alpha}] &= \mathrm{E}_{b_i|D}[f_{ij}(\boldsymbol{x}_{ij}\boldsymbol{\beta}+\boldsymbol{z}_{ij}\boldsymbol{b}_i,t_{ij})] \neq f(\boldsymbol{\beta},\boldsymbol{0},\boldsymbol{x}_{ij}) \\
\mathrm{var}(Y_{ij} \mid \boldsymbol{\beta},\boldsymbol{\alpha}) &= \sigma^2_\epsilon + \mathrm{var}_{b_i|D}[f_{ij}(\boldsymbol{x}_{ij}\boldsymbol{\beta}+\boldsymbol{z}_{ij}\boldsymbol{b}_i,t_{ij})] \\
\mathrm{cov}(Y_{ij},Y_{ij'} \mid \boldsymbol{\beta},\boldsymbol{\alpha}) &= \mathrm{cov}_{b_i|D}(f_{ij}(\boldsymbol{x}_{ij}\boldsymbol{\beta}+\boldsymbol{z}_{ij}\boldsymbol{b}_i,t_{ij}), f_{ij'}(\boldsymbol{x}_{ij'}\boldsymbol{\beta}+\boldsymbol{z}_{ij'}\boldsymbol{b}_i,t_{ij'})] \\
\mathrm{cov}(Y_{ij},Y_{i'j'} \mid \boldsymbol{\beta},\boldsymbol{\alpha}) &= 0, \quad i \neq i'
\end{aligned}$$

The data do not in general have a closed-form marginal distribution.

As with the GLMM there are two issues with respect to implementation:

- How do we evaluate the integrals, and
- how do we maximize the resultant likelihood?

As with the LMEM empirical Bayes estimates for the random effects are available, but caution should be given to using these for checking assumptions since they are strongly influenced by the assumption of normality being correct.

If n_i is large then this will be less of a problem.

Identifiabiity Issues

With many nonlinear models care must be taken to ensure the model is identifiability in the sense that if $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, $f(\boldsymbol{\theta}) \neq f(\boldsymbol{\theta}')$. If there is non-identiability then one may either reparameterize the model, or enforce identifiability through the prior. We illustrate the problems with an example.

Example: Pharmacokinetics of Theophylline

The mean concentration at time point t is

$$f(\eta, t) = \frac{D k_e k_a}{Cl(k_a - k_e)} \{\exp(-k_e t) - \exp(-k_a t)\}$$

where D is the initial dose.

This model is known as the "flip-flop" model because there is non-identifiability; the parameters (k_a, k_e, Cl) give the same curve as the parameters (k_e, k_a, Cl).

To enforce identifiability it is typical to assume that $k_a > k_e > 0$. We may or may not enforce this constraint in our parameterization, as we discuss later.

We first fit the above model to each individual, using non-linear least squares, Figure 23 gives the resultant 95% asymptotic confidence intervals, the between-individual variability is evident.

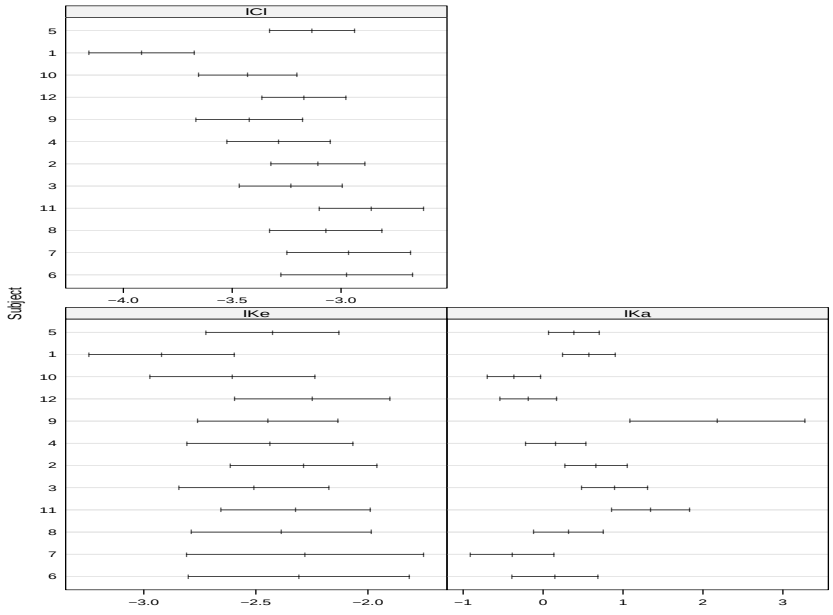

Figure 23: 95% confidence intervals for each of the three parameters and 12 individuals.

NLMEM for Theophylline

The mean concentration at time point t is

$$f(\eta_i, t_{ij}) = \frac{D_i k_{ei} k_{ai}}{Cl_i(k_{ai} - k_{ei})} \{\exp(-k_{ei} t_{ij}) - \exp(-k_{ai} t_{ij})\}$$

where D is the initial dose and

$$\begin{aligned} \log k_{ai} &= \beta_1 + b_{1i} \\ \log k_{ei} &= \beta_2 + b_{2i} \\ \log Cl_i &= \beta_3 + b_{3i} \end{aligned}$$

with $\boldsymbol{b}_i = [b_{1i}, b_{2i}, b_{3i}]^{\mathrm{T}} \sim \mathrm{N}_3(\mathbf{0}, \boldsymbol{D})$.

Exploratory Plots/Analyses

```
> library(nlme); data(Theoph); (Theoph)
> TheoSTS.nls <- nlsList(conc~SSfol(Dose, Time, lKe, lKa, lCl),data=Theoph)
> TheoSTS.nls
 Model: conc ~ SSfol(Dose, Time, lKe, lKa, lCl) | Subject
Coefficients:
         lKe        lKa       lCl
6  -2.307332  0.1516234 -2.973242
7  -2.280370 -0.3860511 -2.964335
8  -2.386437  0.3188339 -3.069111
11 -2.321530  1.3478239 -2.860397
3  -2.508073  0.8975422 -3.229965
2  -2.286108  0.6640568 -3.106317
4  -2.436494  0.1582638 -3.286087
9  -2.446088  2.1821879 -3.420774
12 -2.248326 -0.1828442 -3.170158
10 -2.604148 -0.3631216 -3.428271
1  -2.919614  0.5751612 -3.915857
5  -2.425486  0.3862853 -3.132600
Degrees of freedom: 132 total; 96 residual
Residual standard error: 0.7001921
> plot(intervals(TheoSTS.nls))
> pairs(coef(TheoSTS.nls))
> plot(augPred(TheoSTS.nls))
```

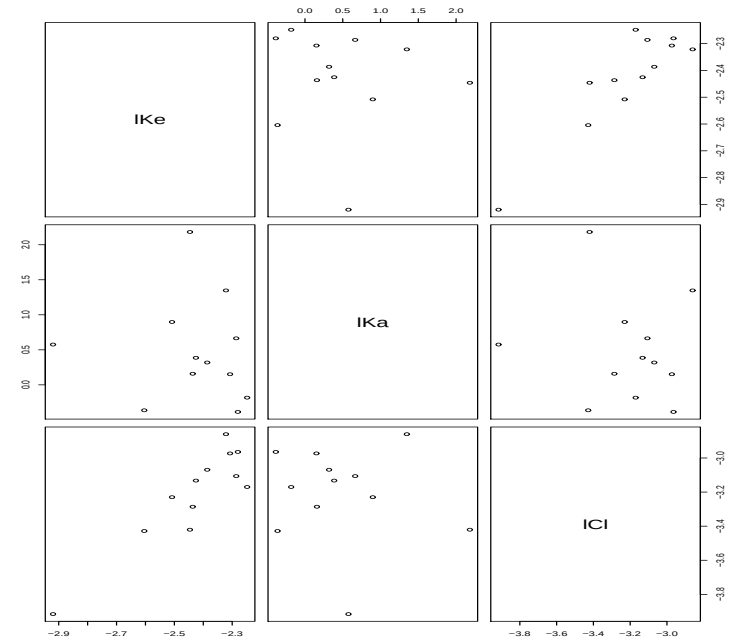

Figure 24: Non-linear LS estimates for 12 individuals.

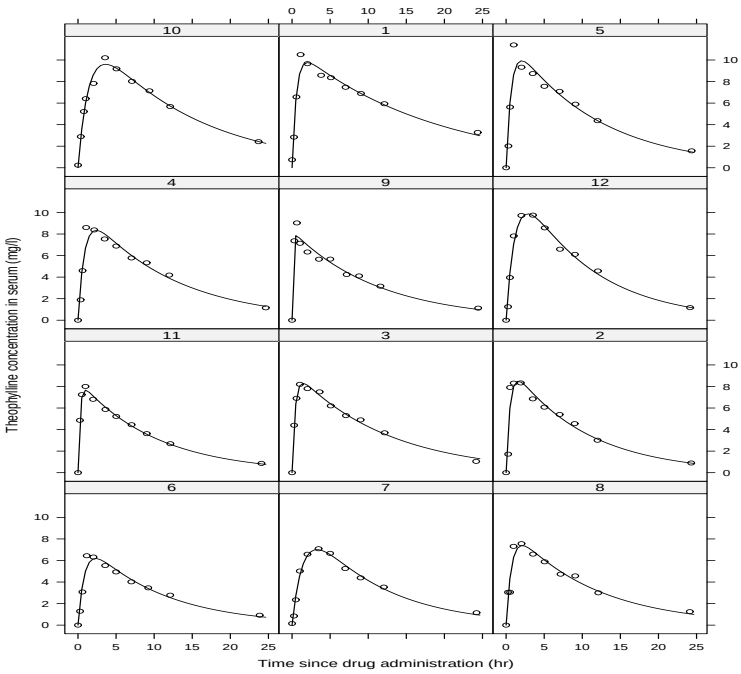

Figure 25: Fitted curves from non-linear LS.

Likelihood Inference for the NLMEM

See Pinheiro and Bates (2000, Chapter 7).

The likelihood is, as usual, obtained by integrating out the random effects:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= (2\pi\sigma_\epsilon^2)^{-N/2}(2\pi)^{-m/2}|\boldsymbol{D}|^{-m/2} \\ &\times \prod_{i=1}^{m} \int \exp\left[-\frac{(\boldsymbol{y}_i - \boldsymbol{f}_i)^{\mathrm{T}}(\boldsymbol{y}_i - \boldsymbol{f}_i)}{2\sigma_\epsilon^2} - \frac{\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{b}_i}{2}\right] d\boldsymbol{b}_i. \end{aligned}$$

where $\boldsymbol{f}_i$ is made up of terms $f(\eta_{ij}, t_{ij})$, $i = 1, ..., m$, $j = 1, ..., n_i$.

Laplace Approximation in the NLMEM

See Pinheiro and Bates, Chapter 7.

We wish to evaluate

$$p(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\alpha}) = (2\pi\sigma^2)^{-n_i/2}(2\pi)^{-(q+1)/2} \mid \boldsymbol{D} \mid^{-1/2} \int \exp\{n_i g(\boldsymbol{b}_i)\}\, d\boldsymbol{b}_i,$$

where

$$-2n_i g(\boldsymbol{b}_i) = [\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{x}_i)]^{\mathrm{T}}[\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{x}_i)]/\sigma_\epsilon^2 + \boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{b}_i.$$

A Laplace approximation is a second-order Taylor series expansion of g about

$$\widehat{\boldsymbol{b}}_i = \arg\min_{\boldsymbol{b}_i} -g(\boldsymbol{b}_i)$$

which will not be available in closed form for a non-linear model.

The `nlme` algorithm

Within `nlme` an algorithm, introduced by Lindstrom and Bates (1990) is used.

The algorithm alternates between two steps:

Penalized Non-linear Least Squares (PNLS)

Condition on the current estimates of $\widehat{\boldsymbol{D}}$ and $\widehat{\sigma}^2_\epsilon$ and then minimize

$$\frac{1}{\widehat{\sigma}^2_\epsilon}\sum_{i=1}^m(\boldsymbol{y}_i-\boldsymbol{f}_i)^{\mathrm{T}}(\boldsymbol{y}_i-\boldsymbol{f}_i)+\boldsymbol{b}_i\widehat{\boldsymbol{D}}^{-1}\boldsymbol{b}_i,$$

to obtain estimates $\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{b}}_1,...,\widehat{\boldsymbol{b}}_m$, which may be viewed as finding the posterior mode for $\boldsymbol{\beta}$ and $\boldsymbol{b}_1,...,\boldsymbol{b}_m$.

Linear Mixed Effects (LME)

Carry out a first-order Taylor series of $\boldsymbol{f}_i$ about $\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{b}}_i$.

This results in a linear mixed effects model which can be maximized to obtain estimates of $\boldsymbol{D}$ and σ^2_ϵ.

We have likelihood

$$L(\boldsymbol{\beta},\boldsymbol{\alpha})=\mid\boldsymbol{D}\mid^{-m/2}\sigma_\epsilon^{-N}\int\exp\left\{-\frac{1}{2}\sum_{i=1}^m(\boldsymbol{y}_i-\boldsymbol{f}_i)^{\mathrm{T}}(\boldsymbol{y}_i-\boldsymbol{f}_i)-\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{b}_i\right\}d\boldsymbol{b}_i$$

where $\boldsymbol{f}_i=\boldsymbol{f}(\boldsymbol{\beta},\boldsymbol{b}_i,\boldsymbol{x}_i)$, $i=1,...,m$.

Carry out a first-order Taylor series expansion of $\boldsymbol{f}_i$ about the estimates, obtained in the PNLS step at iteration k, of $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$, call these $\widehat{\boldsymbol{\beta}}^{(k)}$ and $\widehat{\boldsymbol{b}}_i^{(k)}$.

Specifically

$$\boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i) \approx \boldsymbol{f}_i\left(\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\boldsymbol{b}}_i^{(k)}\right) + \widehat{\boldsymbol{x}}_i^{(k)}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{(k)}\right) + \widehat{\boldsymbol{z}}_i^{(k)}\left(\boldsymbol{b}_i - \widehat{\boldsymbol{b}_i}^{(k)}\right)$$

where

$$\widehat{\boldsymbol{x}}_i^{(k)} = \left.\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\right|_{\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\boldsymbol{b}}_i^{(k)}}$$

$$\widehat{\boldsymbol{z}}_i^{(k)} = \left.\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{b}_i^{\mathrm{T}}}\right|_{\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\boldsymbol{b}}_i^{(k)}}$$

This gives

$$\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i) \approx \boldsymbol{y}_i^{(k)} - \widehat{\boldsymbol{x}}_i^{(k)}\boldsymbol{\beta} - \widehat{\boldsymbol{z}}_i^{(k)}\boldsymbol{b}_i$$

where

$$\boldsymbol{y}_i^{(k)} = \boldsymbol{y}_i - \boldsymbol{f}_i\left(\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\boldsymbol{b}}_i^{(k)}\right) + \widehat{\boldsymbol{x}}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)} + \widehat{\boldsymbol{z}}_i^{(k)}\widehat{\boldsymbol{b}}_i^{(k)}$$

The integral can now be evaluated in closed-form to give the log-likelihood

$$l(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{i=1}^{m}\log|\widehat{\boldsymbol{V}}_i| - \frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{y}_i^{(k)} - \widehat{\boldsymbol{x}}_i^{(k)}\boldsymbol{\beta})^{\mathrm{T}}\widehat{\boldsymbol{V}}_i^{-1}(\boldsymbol{Y}_i - \widehat{\boldsymbol{x}}_i\boldsymbol{\beta})$$

where

$$\widehat{\boldsymbol{V}}_i = \widehat{\boldsymbol{z}}_i^{(k)}\boldsymbol{D}\widehat{\boldsymbol{z}}_i^{(k)\mathrm{T}} + \sigma_\epsilon^2\boldsymbol{I}_i,$$

which may be maximized to give ML estimates. REML estimates are obtained by adding the term

$$-\frac{1}{2}\sum_{i=1}^{m}\log|\,\widehat{\boldsymbol{x}}_i^{(k)\mathrm{T}}\widehat{\boldsymbol{V}}_i(\boldsymbol{\alpha})\widehat{\boldsymbol{x}}_i^{(k)}\,|$$

The Laplace approximation is generally more accurate than the LB algorithm, it is, however, more computationally expensive.

Asymptotic Inference

Under the LB algorithm, the asymptotic distribution of the REML estimator $\widehat{\boldsymbol{\beta}}$ is

$$\left(\sum_{i=1}^{m} \widehat{\boldsymbol{x}}_i^{\mathrm{T}} \widehat{\boldsymbol{V}}_i^{-1} \widehat{\boldsymbol{x}}_i\right)^{1/2} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \ \rightarrow_d \ \mathrm{N}_{p+1}(\boldsymbol{0}, \boldsymbol{I}_{p+1}),$$

where $\widehat{\boldsymbol{x}}_i = \widehat{\boldsymbol{x}}_i^{(k)}$ with k the final iteration, $i = 1, ..., m$

Similarly, the asymptotic distribution of $\boldsymbol{\alpha}$ is based on the information as calculated from the linear approximation to the likelihood.

The LB estimator is inconsistent if the n_i's are fixed and $m \rightarrow \infty$.

Empirical Bayes estimates for the random effects are available, but caution should be given to using these for checking assumptions since they are strongly influenced by the assumption of normality being correct. If n_i is large then this will be less of a problem.

Approaches for NLMEMs

Various other approaches to likelihood inference have been suggested, we briefly summarize.

In general we need to carry out m integrals of dimension $q + 1$ for each likelihood evaluation, so with large m and q this can be computationally expensive.

First-Order Approximation

Let $\boldsymbol{\beta}_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{b}_i$, and then carry out a first-order Taylor series about $\mathrm{E}[\boldsymbol{b}_i] = \boldsymbol{0}$ to give

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}_i) + \boldsymbol{\epsilon}_i \approx \boldsymbol{f}_i(\boldsymbol{x}_i\boldsymbol{\beta}) + \frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{\beta}_i} \frac{\partial \boldsymbol{\beta}_i}{\partial \boldsymbol{b}_i} \boldsymbol{b}_i + \boldsymbol{\epsilon}_i.$$

In contrast to the LB algorithm which considered an expansion about the subject-specific mean, the expansion here is about the population-averaged mean. The first-order estimator is inconsistent and has bias even if n_i and m go to infinity, see Demidenko (2004, Chapter 8)

Adaptive Gaussian quadrature may also be used.

Example: NLMEM for Pharmacokinetics of Theophylline

```
> Theo.nlme <- nlme(TheoSTS.nls,fixed=lKe+lKa+lCl~1,random=lKe+lKa+lCl~1,data=Theoph)
> summary(Theo.nlme)
Nonlinear mixed-effects model fit by maximum likelihood
  Model: conc ~ SSfol(Dose, Time, lKe, lKa, lCl) Random effects:
 Formula: list(lKe ~ 1, lKa ~ 1, lCl ~ 1)
 Structure: General positive-definite, Log-Cholesky parametrization
         StdDev    Corr
lKe      0.1310435 lKe    lKa
lKa      0.6377804  0.012
lCl      0.2511766  0.995 -0.089
Residual 0.6818265
Fixed effects: list(lKe ~ 1, lKa ~ 1, lCl ~ 1)
        Value  Std.Error  DF   t-value p-value
lKe -2.432671 0.06302415 118 -38.59903  0.0000
lKa  0.451410 0.19624487 118   2.30024  0.0232
lCl -3.214452 0.08059540 118 -39.88382  0.0000
 Correlation:
    lKe    lKa
lKa -0.143
lCl  0.854 -0.131
Number of Observations: 132 Number of Groups: 12
> plot(augPred(Theo.nlme))
> plot(compareFits(coef(Theo.nlme), coef(TheoSTS.nls)))
```


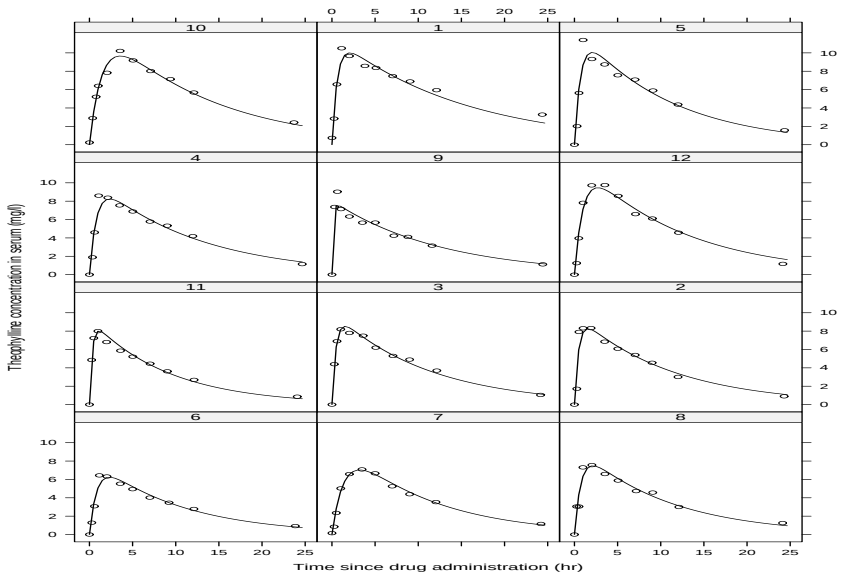

Figure 26: Fitted curves from NLMEM fit.

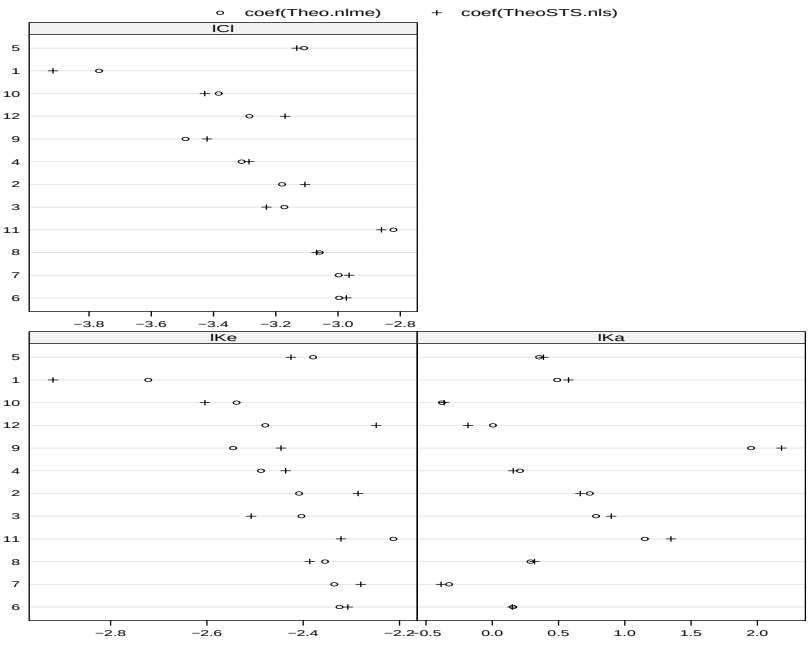

Figure 27: Comparison of non-linear LS and NLMEM estimates.

The following commands produced Figures 28–31.

```
> plot(Theo.nlme,resid(.,type="n")~fitted(.),id=0.05,adj=-1) # id=0.05 gives
# outliers outside of 95% of distn, adj=-1 adjusts the text which
# labels these outliers
> plot(Theo.nlme,resid(.,type="n")~Time,id=0.05,adj=-1)
> qqnorm(Theo.nlme)
> plot(augPred(Theo.nlme,level=0:1)) # Obtain predictions at population and
# individual level of hierarchy
```

The population curves differ here because of the different doses.

There appears to be problems with the assumed mean-variance relationship here.

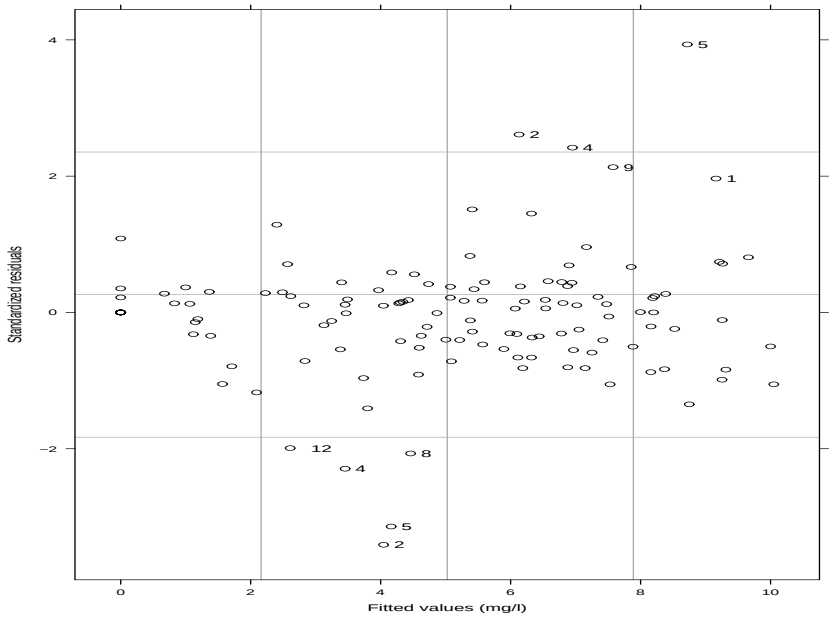

Figure 28: Standardized residuals versus fitted values.

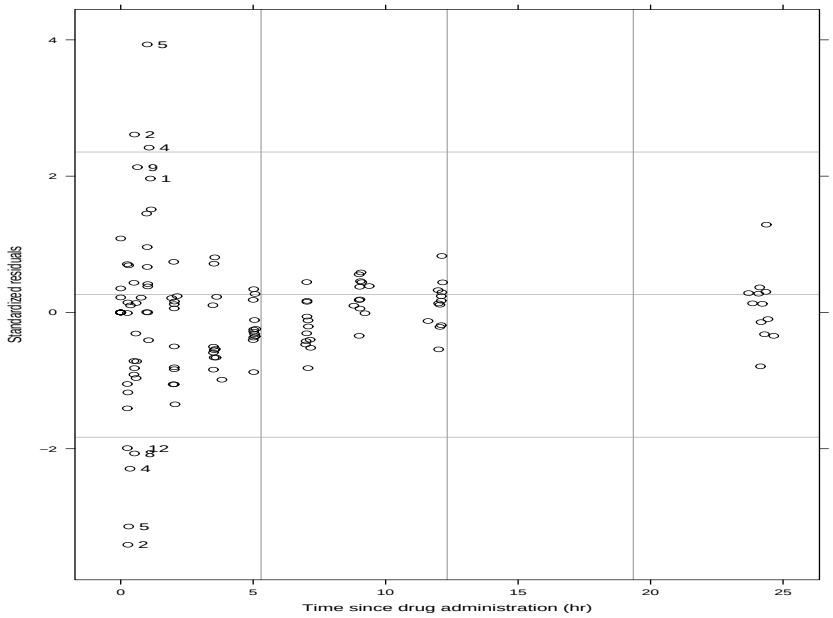

Figure 29: Standardized residuals versus time.

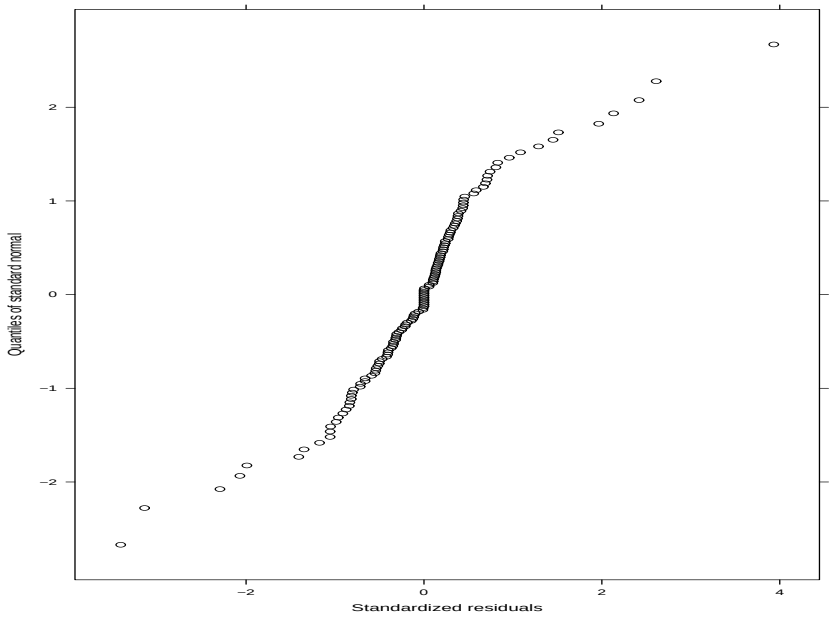

Figure 30: QQ plot of normalized residuals.

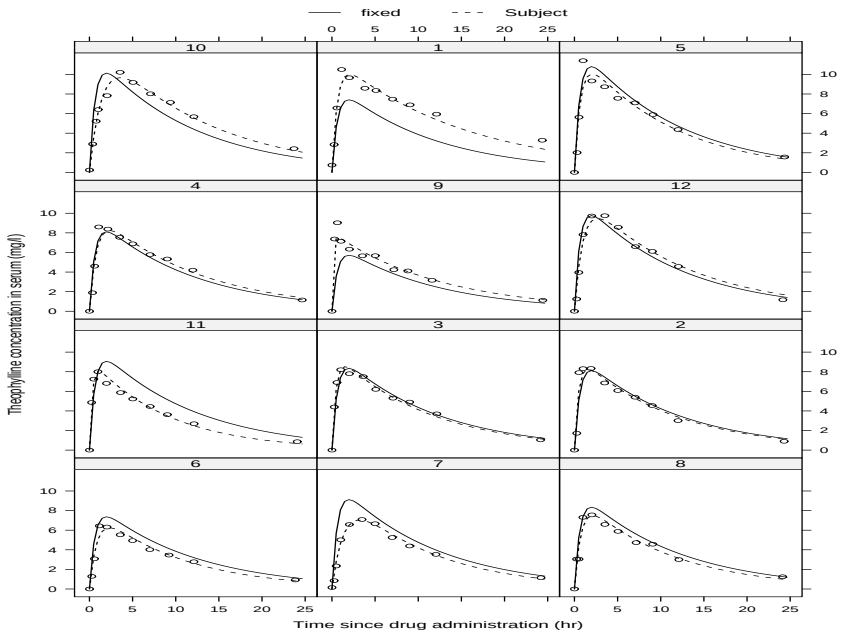

Figure 31: Solid lines are population predictions, dashed lines individual predictions.

Theophylline Data

We now fit the model

$$\log y_{ij} = \log\left[\frac{D_i k_{ai} k_{ei}}{Cl_i(k_{ai}-k_{ei})}\left\{\exp(-k_{ei}t_{ij}) - \exp(-k_{ai}t_{ij})\right\}\right] + \epsilon_{ij}$$

with $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2_\epsilon)$, to attempt to stabilize the variance.

We reparameterize as

$$\begin{aligned} \mathrm{E}[\log y_{ij}] &= \log D_i + \theta_{0i} + \theta_{1i} - \theta_{2i} - \log(\mathrm{e}^{\theta_{0i}} - \mathrm{e}^{\theta_{1i}}) \\ &+ \log[\exp(-\mathrm{e}^{-\theta_{1i}}t_{ij}) - \exp(-\mathrm{e}^{-\theta_{0i}}t_{ij})] \end{aligned}$$

where

$$\begin{aligned} \theta_{0i} &= \log k_{ai} = \beta_0 + b_{i0} \\ \theta_{1i} &= \log k_{ei} = \beta_1 + b_{i1} \\ \theta_{2i} &= \log Cl_i = \beta_2 + b_{i2} \end{aligned}$$

We can no longer use the in-built model function so we build the model from scratch. We also have to remove the observations at $t = 0$ and zero concentrations. Unfortunately the fits don't look a lot better — some model inadequacy here.

Point estimate for $\mathrm{E}[\log k_a]$ is the only parameter that changes substantively.

`nlme` for Theophylline model 2

```
> library(nlme); data(Theoph); (Theoph); Theoph2 <- Theoph[conc>0 & Time>0,]
> logmod <- nlme(model = log(conc) ~ log(Dose) + theta0 + theta1 - theta2 -
log( exp(theta0)-exp(theta1) ) + log( exp(-exp(theta1)*Time) - exp(-exp(theta0)*Time) ),
fixed=~theta0+theta1+theta2~1,data=Theoph2,random=theta0+theta1+theta2~1,
start=c(theta0=0.45,theta1=-2.4,theta2=-3.2) )
> summary(logmod)
Nonlinear mixed-effects model fit by maximum likelihood
  Model: log(conc) ~ log(Dose) + theta0 + theta1 - theta2 - log(exp(theta0) - exp(theta1)) + log(e
Random effects:
 Formula: list(theta0 ~ 1, theta1 ~ 1, theta2 ~ 1)
 Level: Subject
 Structure: General positive-definite, Log-Cholesky parametrization
         StdDev    Corr
theta0   0.6552241 theta0 theta1
theta1   0.1194242 -0.190
theta2   0.2394020 -0.137  0.998
Residual 0.1714818
Fixed effects: ~theta0 + theta1 + theta2 ~ 1
           Value  Std.Error  DF   t-value p-value
theta0  0.231978 0.20309964 106   1.14219  0.2559
theta1 -2.414752 0.04811560 106 -50.18647  0.0000
theta2 -3.211060 0.07290759 106 -44.04287  0.0000
```

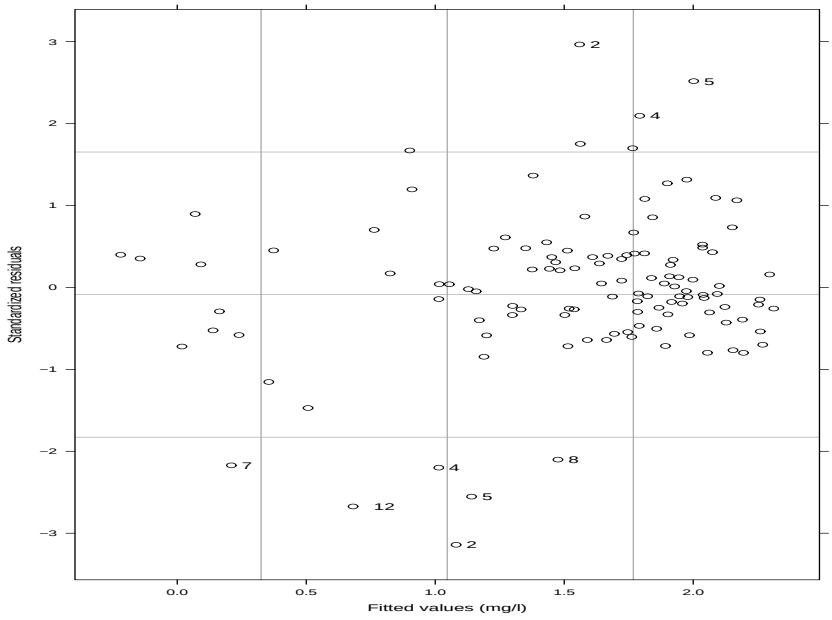

Figure 32: Standardized residuals versus fitted values.

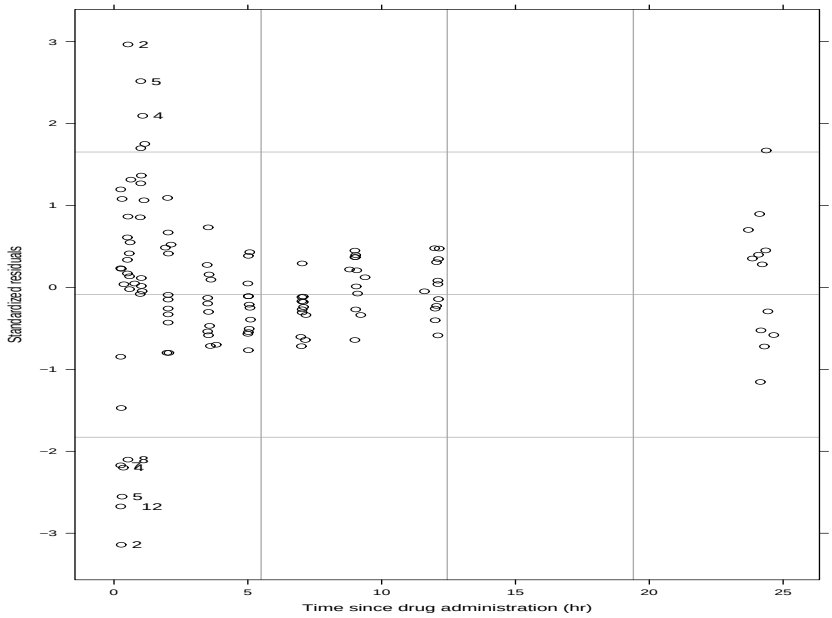

Figure 33: Standardized residuals versus time.

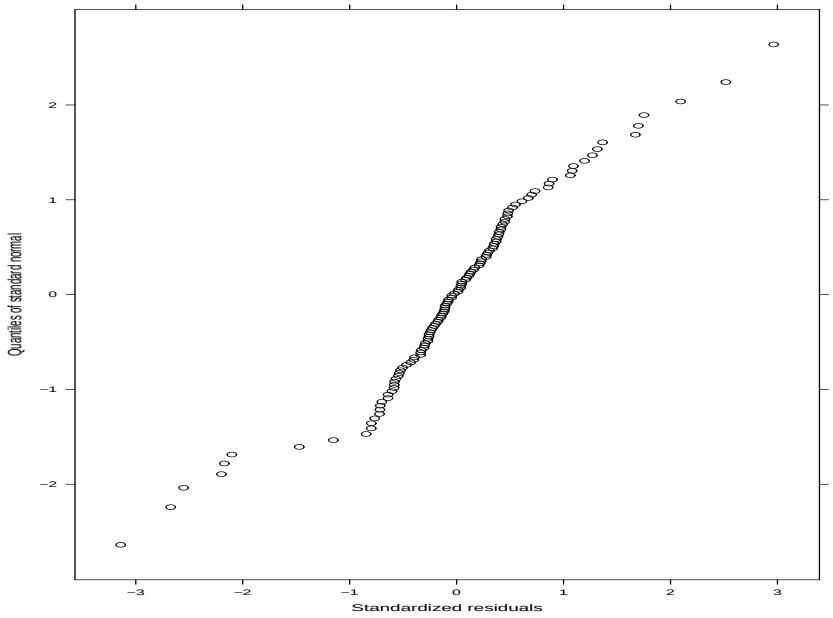

Figure 34: QQ plot of normalized residuals.

Bayesian Approach

A Bayesian approach adds a prior distribution for $\boldsymbol{\beta}, \boldsymbol{\alpha}$, to the likelihood $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$. As with the linear model proper prior is required for the matrix $\boldsymbol{D}$. In general a proper prior is required for $\boldsymbol{\beta}$ also, to ensure the propriety of the posterior distribution. Closed-form inference is unavailable, but MCMC is almost as straightforward as in the LMEM case. The joint posterior is

$$p(\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m, \tau, \boldsymbol{\beta}, \boldsymbol{W}, \boldsymbol{b} \mid \boldsymbol{y}) \propto \prod_{i=1}^{m} \{p(\boldsymbol{y}_i \mid \boldsymbol{\beta}_i, \tau) p(\boldsymbol{\beta}_i \mid \boldsymbol{\beta}, \boldsymbol{W})\} \, \pi(\boldsymbol{\beta}) \pi(\tau) \pi(\boldsymbol{W}).$$

Suppose we have priors:

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathrm{N}_{q+1}(\boldsymbol{\beta}_0, \boldsymbol{V}_0) \\ \tau &\sim \mathrm{Ga}(a_0, b_0) \\ \boldsymbol{W} &\sim \mathrm{W}_{q+1}(r, \boldsymbol{R}^{-1}) \end{aligned}$$

The conditional distributions for $\boldsymbol{\beta}$, τ, $\boldsymbol{W}$ are unchanged from the linear case. There is no closed form conditional distribution for $\boldsymbol{\beta}_i$, which is given by:

$$p(\boldsymbol{\beta}_i \mid \boldsymbol{\beta}, \tau, \boldsymbol{W}, \boldsymbol{y}) \propto p(\boldsymbol{y}_i \mid \boldsymbol{\beta}_i, \tau) \times p(\boldsymbol{\beta}_i \mid \boldsymbol{\beta}, \boldsymbol{W})$$

but a Metropolis-Hastings step can be used.

Example: Theophylline Prior Specification

We begin with the obvious parameterization:

$$\begin{aligned} \log k_{ei} &= \beta_1 + b_{1i} \\ \log k_{ai} &= \beta_2 + b_{2i} \\ \log Cl_i &= \beta_3 + b_{3i} \end{aligned}$$

with $\boldsymbol{b}_i = [b_{1i}, b_{2i}, b_{3i}]^{\mathrm{T}} \sim \mathrm{N}_3(\boldsymbol{0}, \boldsymbol{D})$.

We assume independent normal priors for the elements of $\boldsymbol{\beta}$, centered at 0 and with large variances (recall that we need proper priors).

For $\boldsymbol{D}$ we assume a Wishart$(r, \boldsymbol{R})$ with $r = 3$ and zero off-diagonal elements. We choose the diagonal elements with the following rationale.

Consider a generic "natural" parameter ϕ (for example, k_e, k_a or Cl) for which $\phi \sim \mathrm{LogNormal}(\mu, \sigma^2)$.

Pharmacokinetics have insight into the coefficient of variation for θ, i.e. $\mathrm{CV}(\theta) = \mathrm{sd}(\theta)/\mathrm{E}[\theta]$.

We have

$$\begin{aligned} \mathrm{E}[\theta] &= \exp(\mu + \sigma^2/2) \\ \mathrm{var}(\theta) &= \mathrm{E}[\theta]^2(\mathrm{e}^{\sigma^2} - 1) \\ \mathrm{sd}(\theta) &= \mathrm{E}[\theta]\sqrt{\mathrm{e}^{\sigma^2} - 1} \\ &\approx \mathrm{E}[\theta]\sigma \\ \mathrm{CV}(\theta) &\approx \sigma \end{aligned}$$

Hence $\sqrt{D_{ii}}$ is approximately the coefficient of variation, which allows a prior to be placed. In the Theophylline example we choose a prior guess of 20% CV, i.e. $R_{ii} = 0.04$, $i = 1, 2, 3$.

For inference: $\exp(\mu - \sigma), \exp(\mu), \exp(\mu + \sigma^2/2)$ are the mode, mean and median of the population distribution of θ, and $\exp(\mu \pm 1.96\sigma)$ is a 95% interval for θ in the population.

Functions of Interest

In pharamacokinetics there is interest in quantities such as the terminal half-life

$$t_{1/2} = \frac{\log 2}{k_e}$$

Since $\log k_e \sim \mathrm{N}(\beta_1, D_{11})$,

$$\log t_{1/2} \sim \mathrm{N}(\log[\log 2] - \beta_1, D_{11})$$

Other parameters are not simple linear combinations, e.g. time to maximum

$$t_{\max} = \frac{1}{k_a - k_e} \log\left(\frac{k_a}{k_e}\right)$$

and the maximum concentration

$$\begin{aligned} \mathrm{E}[Y|t_{\max}] &= \frac{Dk_a}{V(k_a - k_e)} \{\exp(-k_e t_{\max}) - \exp(-k_a t_{\max})\} \\ &= \frac{D}{V}\left(\frac{k_a}{k_e}\right)^{k_a/(k_a-k_e)}. \end{aligned}$$

WinBUGS for Theophylline

```
model
{
for( i in 1 : N ) {
for( j in 1 : T ) {
Y[i , j] ~ dnorm(mu[i , j],eps.tau)
mu[i , j] <- Dose[i]*exp(theta[i,1] + theta[i,2] - theta[i,3]) *
                (exp(-exp(theta[i,1])*time[i,j]) - exp(-exp(theta[i,2])*time[i,j]) )/
                (exp(theta[i,2])-exp(theta[i,1]))
}
theta[i, 1:3] ~ dmnorm(beta[1:3], Dinv[1:3, 1:3])
ke[i] <- exp(theta[i,1])
ka[i] <- exp(theta[i,2])
Cl[i] <- exp(theta[i,3])
}
eps.tau <- exp(logtau)
logtau ~ dflat()
sigma <- 1 / sqrt(eps.tau)
beta[1:3] ~ dmnorm(mean[1:3], prec[1:3, 1:3])
kemed <- exp(beta[1])
kamed <- exp(beta[2])
Clmed <- exp(beta[3])
Dinv[1:3, 1:3] ~ dwish(R[1:3, 1:3], 3)
D[1:3, 1:3] <- inverse(Dinv[1:3, 1:3])
```

```
for (i in 1 : 3) {sdD[i] <- sqrt(D[i, i]) }
}
DATA
list( N = 12, T = 11,Dose=c(4.02,4.4,4.53,4.4,5.86,4,4.95,4.53,3.1,5.5,4.92,5.3),
Y = structure(.Data =   c(0.74,2.84,6.57,10.50,...,4.57,1.17),
.Dim = c(12,11)),
time = structure(.Data = c(0.00,0.25, ...,7.07,9.03,12.05,24.15),.Dim = c(12,11)),
mean = c(0,0,0),R = structure(.Data = c(0.2, 0, 0,0, 0.2, 0,0, 0, 0.2), .Dim = c(3, 3)),
prec = structure(.Data = c(1.0E-6,0,0,0,1.0E-6,0,0,0,1.0E-6),.Dim = c(3, 3)))
INITS
list(theta = structure(.Data = c(-2.2,0,3,-2.2,0,3,-2.2,0,3,-2.2,0,3,-2.2,0,3,-2.2,
0,3,-2.2,0,3,-2.2,0,3,-2.2,0,3,-2.2,0,3,-2.2,0,3,-2.2,0,3), .Dim = c(12, 3)),
beta = c(-2, .1, -3), Dinv = structure(.Data = c(1, 0, 0,0, 1, 0,0, 0, 1), .Dim = c(3, 3)),
logtau = 0)
```

The first time this model was run with two chains the second chain flipped between two non-identifiable regions in the parameter space

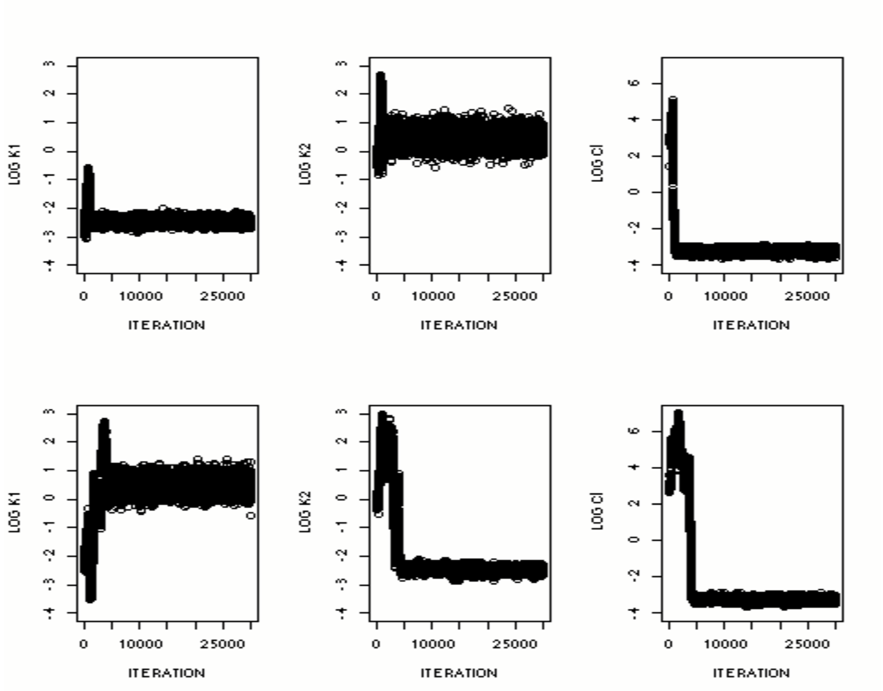

Figure 35: Demonstration of flip-flop behavior.

Bayesian analysis of Theophylline Data

	Likelihood	Bayes
Parameter	Estimate (s.e.)	Estimate (s.d.)
β_1	-2.43 (0.06)	-2.45 (0.09)
β_2	0.45 (0.20)	0.47 (0.21)
β_3	-3.21 (0.08)	-3.22 (0.09)
$\sqrt{D_{11}}$	0.13 (–)	0.22 (0.06)
$\sqrt{D_{22}}$	0.64 (–)	0.69 (0.18)
$\sqrt{D_{33}}$	0.25 (–)	0.29 (0.07)

Table 14: Comparison of likelihood and Bayesian estimation techniques. For the likelihood summaries we report the MLEs and the asymptotic standard errors, while for the Bayesian analysis we report the mean and standard deviation of the posterior distribution.

Inference is very similar between the two approaches (the Bayesian summaries are based on a second run that did not flip-flop).

Theopylline: A Second Model

We now constrain the parameters so that $k_{ai} > k_{ei} > 0$ via the parameterization:

$$\begin{aligned}\theta_{1i} &= \log k_{ei} = \beta_1 + b_{1i} \\ \theta_{2i} &= \log(k_{ai} - k_{ei}) = \beta_2 + b_{2i} \\ \theta_{3i} &= \log Cl_i = \beta_3 + b_{3i}\end{aligned}$$

with $\boldsymbol{b}_i = [b_{1i}, b_{2i}, b_{3i}]^{\mathrm{T}} \sim \mathrm{N}_3(\boldsymbol{0}, \boldsymbol{D})$. Note this is a different model since the prior is different.

Hence

$$\begin{aligned}k_{ei} &= \exp(\theta_{1i}) \\ k_{ai} &= \exp(\theta_{1i}) + \exp(\theta_{2i}) \\ Cl_i &= \exp(\theta_{3i})\end{aligned}$$

Note that:

$$\mathrm{E}[k_a] = \mathrm{E}[\exp(\theta_1) + \exp(\theta_2)] = \mathrm{e}^{\beta_1 + \sqrt{D_{11}}/2} + \mathrm{e}^{\beta_1 + \sqrt{D_{11}}/2}$$

WinBUGS code for second model

```
model
{
for( i in 1 : N ) {
for( j in 1 : T ) {
Y[i , j] ~ dnorm(mu[i , j],eps.tau)
mu[i , j] <- Dose[i]*exp(theta[i,1])*(exp(theta[i,1])+exp(theta[i,2]))*
                exp(- theta[i,3]) * (exp(-exp(theta[i,1])*time[i,j]) - exp(-(exp(theta[i,1])+
                exp(theta[i,2]))*time[i,j]) )/exp(theta[i,2])
}
theta[i, 1:3] ~ dmnorm(beta[1:3], Dinv[1:3, 1:3])
ke[i] <- exp(theta[i,1])
ka[i] <- exp(theta[i,1])+exp(theta[i,2])
Cl[i] <- exp(theta[i,3])
}
eps.tau <- exp(logtau)
logtau ~ dflat()
sigma <- 1 / sqrt(eps.tau)
beta[1:3] ~ dmnorm(mean[1:3], prec[1:3, 1:3])
kemean <- exp(beta[1]+sqrt(sdD[1,1]))
kamean <- exp(beta[1]+sqrt(sdD[1,1]))+exp(beta[2]+sqrt(sdD[2,2]))
Clmean <- exp(beta[3]+sqrt(sdD[2,2]))
kemed <- exp(beta[1])
Clmed <- exp(beta[3])
```



```
Dinv[1:3, 1:3] ~ dwish(R[1:3, 1:3], 3)
D[1:3, 1:3] <- inverse(Dinv[1:3, 1:3])
for (i in 1 : 3) {sdD[i] <- sqrt(D[i, i]) }
}
Results:
 node  mean  sd MC error 2.5% median 97.5% start sample
beta[1] -2.461 0.08436 0.003582 -2.629 -2.461 -2.296 5000 15001
beta[2] 0.413 0.2266 0.003567 -0.037 0.4132 0.868 5000 15001
beta[3] -3.229 0.09176 0.002393 -3.41 -3.23 -3.047 5000 15001
sdD[1] 0.2243 0.06038 0.002094 0.1367 0.2142 0.3687 5000 15001
sdD[2] 0.7277 0.1853 0.004157 0.4518 0.6993 1.168 5000 15001
sdD[3] 0.2839 0.07131 0.002151 0.1779 0.2726 0.457 5000 15001
```

Generalized Estimating Equations

If interest lies in population parameters then we may use the estimator $\widehat{\boldsymbol{\beta}}$ that satisfies

$$\boldsymbol{G}(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}) = \sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0},$$

where $\boldsymbol{D}_i = \frac{\partial \mu_i}{\partial \boldsymbol{\beta}}$, $\boldsymbol{W}_i = \boldsymbol{W}_i(\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}})$ is the working covariance model, $\mu_i = \mu_i(\boldsymbol{\beta})$ and $\widehat{\boldsymbol{\alpha}}$ is a consistent estimator of $\boldsymbol{\alpha}$. Sandwich estimation may be used to obtain an empirical estimate of the variance, $\boldsymbol{V}_\beta$:

$$\left(\sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{D}_i \right)^{-1} \left\{ \sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \mathrm{cov}(\boldsymbol{Y}_i) \boldsymbol{W}_i^{-1} \boldsymbol{D}_i \right\} \left(\sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{D}_i \right)^{-1}.$$

We then have

$$\boldsymbol{V}_\beta^{-1/2} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow_d \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}).$$

In practice an empirical estimator of $\mathrm{cov}(\boldsymbol{Y}_i)$ is substituted to give $\widehat{\boldsymbol{V}}_\beta$.

GEE has not been extensively used in a non-linear (non-GLM) setting. This is probably because in many settings (e.g. pharmacokinetic/pharmacodynamic) interest focuses on understanding between individual-variability, and explaining this in terms of individual-specific covariates.