Now assume a non-diagonal $\boldsymbol{D}$ for all four parameters.

```
> nlme2.indo2 <- update( nlme.indo, random=A1+lrc1+A2+lrc2~1)
> summary(nlme.indo2)
Nonlinear mixed-effects model fit by maximum likelihood
  Model: conc ~ SSbiexp(time, A1, lrc1, A2, lrc2)
Random effects:
 Formula: list(A1 ~ 1, lrc1 ~ 1, A2 ~ 1, lrc2 ~ 1)
 Level: Subject
 Structure: General positive-definite, Log-Cholesky parametrizati
         StdDev     Corr
A1       0.77583020 A1    lrc1  A2
lrc1     0.26863662 0.963
A2       0.38707000 0.459 0.682
lrc2     0.48253192 0.153 0.414 0.948
Residual 0.06962038
Fixed effects: list(A1 ~ 1, lrc1 ~ 1, A2 ~ 1, lrc2 ~ 1)
          Value Std.Error DF   t-value p-value
A1    2.8531611 0.3485825 57  8.185039   0e+00
lrc1  0.8755645 0.1253269 57  6.986245   0e+00
A2    0.6357872 0.1715520 57  3.706091   5e-04
lrc2 -1.2757709 0.2161119 57 -5.903288   0e+00
 Correlation:
     A1    lrc1  A2
lrc1 0.907
A2   0.411 0.676
lrc2 0.108 0.378 0.912
```


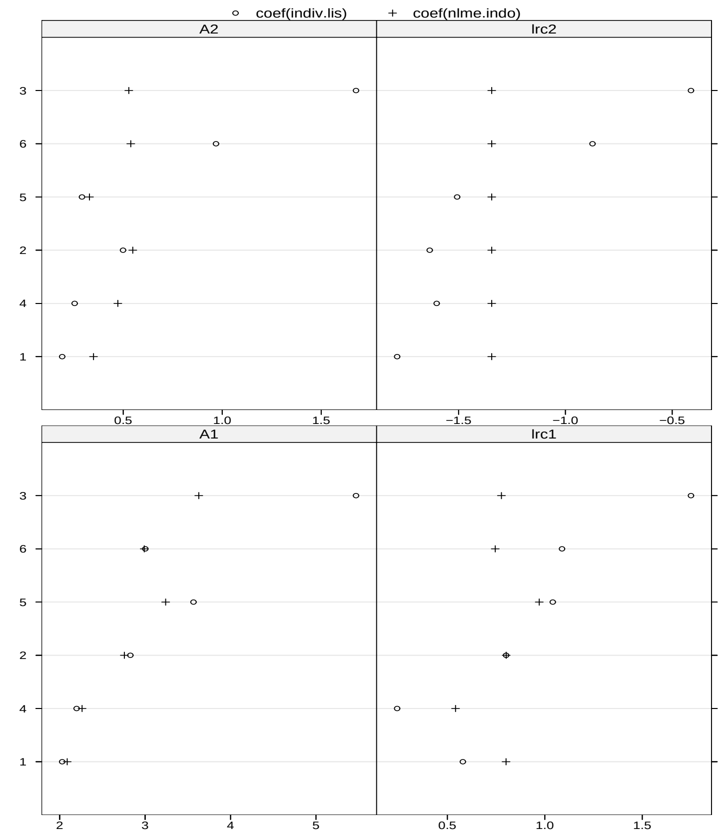

Figure 33: Comparison of non-linear LS and nlme estimates, with the latter from the model `nlme.indo` Created using the command `plot(compareFits(coef(indiv.is),coef(nlme.indo)))`.

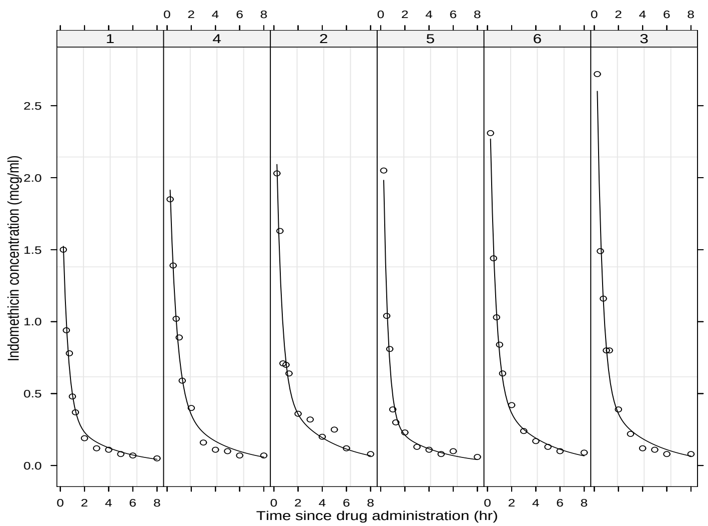

Figure 34: Data with fitted curves from nlme analysis superimposed from the model `nlme.indo`. Created with the command `plot(augPred(nlme.indo),aspect=''xy'',grid=T)`.

Analytic Approximations

Various 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 Apprximation

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

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

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)

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

Laplace Approximation (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

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

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

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

The Laplace approximation is generally more accurate than the LB algorithm since the expansion is about $\boldsymbol{b}_i$ only.

It is, however, more computationally expensive.

Penalized Quasi-Likelihood

Breslow and Clayton (1993) used a Laplace approximation in the context of GLMMs, and then use a variant of Fisher scoring to find estimates.

Adaptive Quadrature

Adaptive Gaussian quadrature may also be used – increasing the number of points in the integration rule allows some assessment of the accuracy of the approximation.

Bayesian Approach

A Bayesian approach adds a prior distribution for $\boldsymbol{\beta}, \boldsymbol{\alpha}$.

A proper prior is required for the matrix $\boldsymbol{D}$ and, in general a proper prior is required for $\boldsymbol{\beta}$ also, both 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 only difference is that a Gibbs sampling strategy will not yield closed form conditional distributions for $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ and so Metropolis-Hastings steps (for example), are required.

Generalized Linear Mixed Models

A GLMM is defined by

1. *Random Component*: $Y_{ij}|\theta_{ij},\phi \sim p(\cdot)$ where $p(\cdot)$ is a member of the exponential family, that is

$$p(y_{ij}|\theta_{ij},\alpha) = \exp[\{y_{ij}\theta_{ij} - b(\theta_{ij})\}/a(\alpha) + c(y_{ij},\alpha)],$$

   for $i = 1, ..., m$ units, and $j = 1, ..., n_i$, measurements per unit.

2. *Systematic Component*: If $\mu_{ij} = \mathrm{E}[Y_{ij}|\theta_{ij},\alpha]$ then we have a link function $g(\cdot)$, with

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i,$$

   so that we have introduced random effects into the linear predictor. The above defines the *conditional* part of the model. The random effects are then assigned a distribution, and in a GLMM this is assumed to be

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

   We also have

$$\mathrm{var}(Y_{ij}|\theta_{ij},\alpha) = \alpha v(\mu_{ij}).$$

Marginal Moments

Mean:

$$\begin{aligned}\mathrm{E}[Y_{ij}] &= \mathrm{E}\{\mathrm{E}[Y_{ij}|\boldsymbol{b}_i]\}\\ &= \mathrm{E}[\mu_{ij}] = \mathrm{E}_{\boldsymbol{b}}[g^{-1}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i)].\end{aligned}$$

Variance:

$$\begin{aligned}\mathrm{var}(Y_{ij}) &= \mathrm{E}[\mathrm{var}(Y_{ij}|\boldsymbol{b}_i)] + \mathrm{var}(\mathrm{E}[Y_{ij}|\boldsymbol{b}_i])\\ &= \alpha\mathrm{E}_{\boldsymbol{b}}[v\{g^{-1}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i)\}]\\ &+ \mathrm{var}_{\boldsymbol{b}}[g^{-1}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i)].\end{aligned}$$

Covariance:

$$\begin{aligned}\mathrm{cov}(Y_{ij},Y_{ik}) &= \mathrm{E}[\mathrm{cov}(Y_{ij},Y_{ik}|\boldsymbol{b}_i)] + \mathrm{cov}(\mathrm{E}[Y_{ij}|\boldsymbol{b}_i],\mathrm{E}[Y_{ik}|\boldsymbol{b}_i])\\ &= \mathrm{cov}\{g^{-1}(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i), g^{-1}(\boldsymbol{x}_{ik}\boldsymbol{\beta} + \boldsymbol{z}_{ik}\boldsymbol{b}_i)\}\\ &\neq 0,\end{aligned}$$

for $j \neq k$ due to shared random effects, and

$$\mathrm{cov}(Y_{ij},Y_{lk}) = 0,$$

for $i \neq l$, is there are no shared random effects.

Example: Log-Linear Poisson Regression

Stage 1: $Y_{ij}|\boldsymbol{\beta}, b_i \sim_{ind} \text{Poisson}(\mu_{ij})$, with

$$g(\mu_{ij}) = \log \mu_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + b_i.$$

Hence

$$\text{E}[Y_{ij}|b_i] = \mu_{ij} = \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + b_i),$$

and

$$\text{var}(Y_{ij}|b_i) = \mu_{ij}.$$

Stage 2: $b_i \sim_{iid} N(0, \sigma^2)$.

The marginal mean is given by

$$\text{E}[Y_{ij}] = \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \sigma^2/2),$$

and the marginal median by

$$\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta}).$$

The marginal variance is given by

$$\begin{aligned} \text{var}(Y_{ij}) &= \text{E}[\mu_{ij}] + \text{var}(\mu_{ij}) \\ &= \text{E}[Y_{ij}]\{1 + \text{E}[Y_{ij}](\text{e}^{\sigma^2} - 1)\} \\ &= \text{E}[Y_{ij}](1 + \text{E}[Y_{ij}] \times \kappa) \end{aligned}$$

where $\kappa = \text{e}^{\sigma^2} - 1 > 0$ illustrating that we have excess-Poisson variation which increases as σ^2 increases.

For the marginal covariance

$$\begin{aligned} \text{cov}(Y_{ij}, Y_{ik}) &= \text{cov}\{\exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + b_i), \exp(\boldsymbol{x}_{ik}\boldsymbol{\beta} + b_i)\} \\ &= \exp(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{x}_{ik}\boldsymbol{\beta}) \times \text{e}^{\sigma^2}\{\text{e}^{\sigma^2} - 1\} \\ &= \text{E}[Y_{ij}]\text{E}[Y_{ik}]\kappa. \end{aligned}$$

Hence for individual i we have variance-covariance matrix

$$\begin{bmatrix} \mu_{i1} + \mu_{i1}^2\kappa & \mu_{i1}\mu_{i2}\kappa & \dots & \mu_{i1}\mu_{in_i}\kappa \\ \mu_{i2}\mu_{i1}\kappa & \mu_{i2} + \mu_{i2}^2\kappa & \dots & \mu_{i2}\mu_{in_i}\kappa \\ \dots & \dots & \dots & \dots \\ \mu_{in_i}\mu_{i1}\kappa & \mu_{in_i}\mu_{i2}\kappa & \dots & \mu_{in_i} + \mu_{in_i}^2\kappa \end{bmatrix},$$

where $\kappa = \text{e}^{\sigma^2} - 1 > 0$.

Likelihood Inference

As with the linear mixed effects model (LMEM) we maximize $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$ where $\boldsymbol{\alpha}$ denote the variance components in $\boldsymbol{D}$, and

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod_{i=1}^{m} \int p(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{b}_i) \times p(\boldsymbol{b}_i | \boldsymbol{\alpha}) \ \mathrm{d}\boldsymbol{b}_i.$$

Unlike the LMEM the required integrals are not available in closed form and so some sort of analytical or numerical approximation is required.

Example: Log-linear Poisson regression GLMM

With a single random effect we have $\boldsymbol{\alpha} = \sigma^2$.

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \prod_{i=1}^{m} \int \prod_{j=1}^{n_i} \frac{\exp(-\mu_{ij}) \mu_{ij}^{y_{ij}}}{y_{ij}!} \times (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2} b_i^2\right) \ \mathrm{d}b_i \\ &= \prod_{i=1}^{m} (2\pi\sigma^2)^{-1/2} \exp\left(\sum_{i=1}^{n_i} y_{ij} x_{ij} \boldsymbol{\beta}\right) \\ &\times \int \exp\left(-\mathrm{e}^{b_i} \sum_{j=1}^{n_i} \mathrm{e}^{\boldsymbol{x}_{ij}\boldsymbol{\beta}} + \sum_{j=1}^{n_i} y_{ij} b_i - \frac{1}{2\sigma^2} b_i^2\right) \ \mathrm{d}b_i \\ &= \prod_{i=1}^{m} \exp\left(\sum_{i=1}^{n_i} y_{ij} x_{ij} \boldsymbol{\beta}\right) \times \int h(b_i) \frac{\exp\{-b_i^2/(2\sigma^2)\}}{(2\pi\sigma^2)^{-1/2}} \ \mathrm{d}b_i, \end{aligned}$$

an integral with respect to a normal random variable (which is analytically intractable).

Likelihood Inference fo loglinear GLMM

In general there are two approaches to inference from a likelihood perspective:

1. Carry out conditional inference in order to eliminate the random effects.
2. Make a distributional assumption for $\boldsymbol{b}_i$, and then carry out likelihood inference (using some form of approximation to evaluate the required integrals).

We first consider the first approach. For simplicity we assume the canonical link function,

$$g(\mu_{ij}) = \theta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i$$

to give likelihood

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \exp\left\{\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}\boldsymbol{x}_{ij}\boldsymbol{\beta} + y_{ij}\boldsymbol{z}_{ij}\boldsymbol{b}_i - b(\boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i)\right\},$$

so that we have sufficient statistics $\boldsymbol{t}_1 = \sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}\boldsymbol{x}_{ij}$ for $\boldsymbol{\beta}$ and $\boldsymbol{t}_{2i} = \sum_{j=1}^{n_i} y_{ij}\boldsymbol{z}_{ij}$ for $\boldsymbol{b}_i$.

Recall the definition of conditional likelihood. Suppose the distribution of the data may be factored as

$$p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{b}) = cp(\boldsymbol{t}_1 \mid \boldsymbol{t}_2, \boldsymbol{\beta}) \times p(\boldsymbol{t}_2 \mid \boldsymbol{\beta}, \boldsymbol{b}),$$

where we choose to ignore the second term and consider the conditional likelihood

$$l_c(\boldsymbol{\beta}) = p(\boldsymbol{t}_1 \mid \boldsymbol{t}_2, \boldsymbol{\beta}) \propto \frac{p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{b})}{p(\boldsymbol{t}_2 \mid \boldsymbol{\beta}, \boldsymbol{b})}.$$

In the context of GLMs we have

$$l_c(\boldsymbol{\beta}) = \prod_{i=1}^{m} p(\boldsymbol{t}_{1i} \mid \boldsymbol{t}_{2i}, \boldsymbol{\beta}) \propto \frac{p(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{b})}{p(\boldsymbol{t}_{2i} \mid \boldsymbol{\beta}, \boldsymbol{b})}.$$

In lectures: development of conditional likelihood for GLMMs.