Mixed models in R using the lme4 package
Part 8: Nonlinear mixed models

Douglas Bates

University of Wisconsin - Madison
and R Development Core Team
<Douglas.Bates@R-project.org>

University of Lausanne

July 3, 2009

Outline

# Nonlinear mixed-effects models (NLMM)

- The LMM and GLMM are powerful data analysis tools.
- The "common denominator" of these models is the expression for the linear predictor. The models require that the fixed effects parameters and the random effects occur linearly in

$$\boldsymbol{\eta} = \boldsymbol{Zb} + \boldsymbol{X\beta} = \boldsymbol{Uu} + \boldsymbol{X\beta}$$

- This is a versatile and flexible way of specifying *empirical* models, whose form is determined from the data.
- In many situations, however, the form of the model is derived from external considerations of the mechanism generating the response. The parameters in such *mechanistic* models often occur nonlinearly.
- Mechanistic models can emulate behavior like the response approaching an asymptote, which is not possible with models that are linear in the parameters.
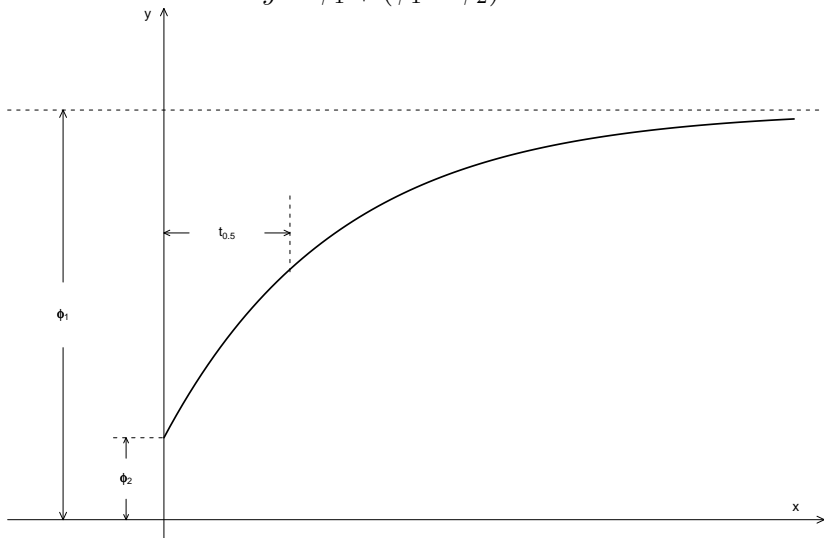
The Michaelis-Menten model, SSmicmen

$$y = \frac{\phi_1 x}{x + \phi_2}$$
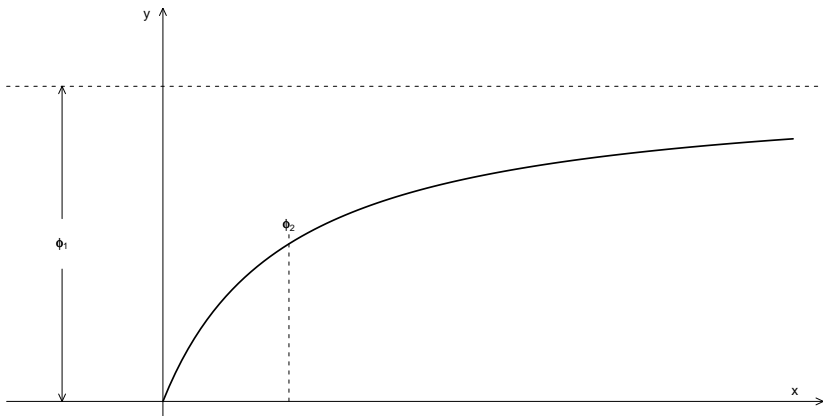
ϕ_1 (called V_m in enzyme kinetics) is the maximum reaction velocity, ϕ_2 (K) is the concentration at which $y = \phi_1/2$.
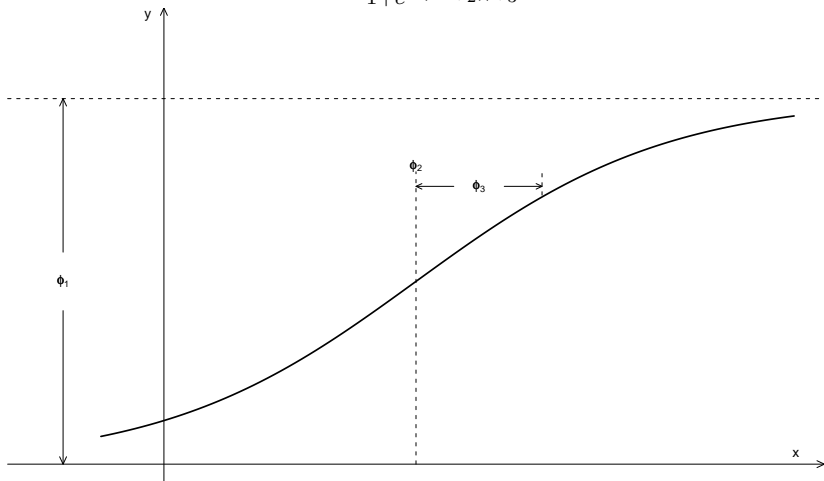
The "asymptotic regression" model, SSasymp

$$y = \phi_1 + (\phi_1 - \phi_2)e^{-\phi_3 x}$$

The logistic growth model, SSlogis

$$y = \frac{\phi_1}{1 + e^{-(x - \phi_2)/\phi_3}}$$

Modeling repeated measures data with a nonlinear model

- Nonlinear mixed-effects models are used extensively with longitudinal pharmacokinetic data.
- For such data the time pattern of an individual's response is determined by pharmacokinetic parameters (e.g. rate constants) that occur nonlinearly in the expression for the expected response.
- The form of the nonlinear model is determined by the pharmacokinetic theory, not derived from the data.

$$d \cdot k_e \cdot k_a \cdot C \frac{e^{-k_e t} - e^{-k_a t}}{k_a - k_e}$$

- These pharmacokinetic parameters vary over the population. We wish to characterize typical values in the population and the extent of the variation.
- Thus, we associate random effects with the parameters, k_a, k_e and C in the nonlinear model.

A simple example - logistic model of growth curves

- The `Orange` data set are measurements of the growth of a sample of five orange trees in a location in California.
- The response is the circumference of the tree at a particular height from the ground (often converted to "diameter at breast height").
- The covariates are `age` (days) and `Tree` (balanced).
- A data plot indicates that the growth patterns are similar but the eventual heights vary.
- One possible growth model is the *logistic* growth model

$$f(t, A, t_0, s) = \frac{A}{1 + e^{-(t-t_0)/s}}$$

which can be seen to be related to the inverse logit link function.

Orange tree growth data
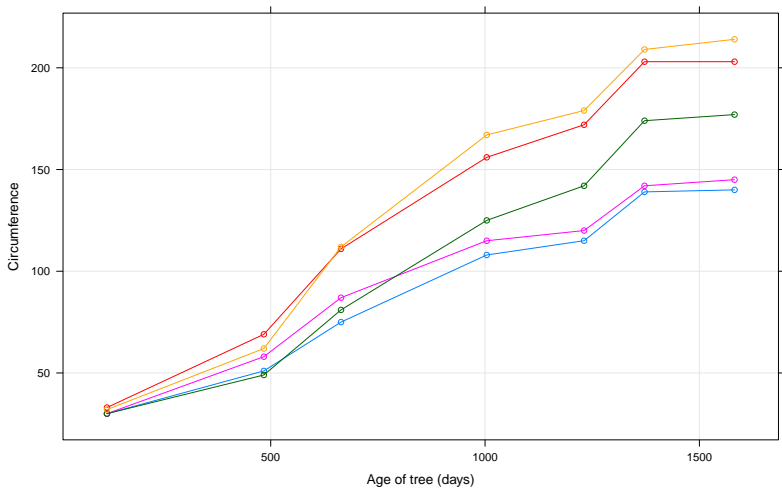
Using nlmer

- The nonlinear mixed-effects model is fit with the `nlmer` function in the `lme4` package.
- The formula argument for `nlmer` is in three parts: the response, the nonlinear model function depending on covariates and a set of nonlinear model (nm) parameters, and the mixed-effects formula.
- There is no longer a concept of an intercept or a `1` term in the mixed-effects model. All terms in the mixed-effects formula incorporate names of nm parameters.
- The default term for the fixed-effects is a separate "intercept" parameter for each nm parameter.
- At present, the nonlinear model must provide derivatives, in addition to the expected response. The `deriv` function can be used to create such a function from an expression.
- The starting values for the fixed effects must also be given. It is safest to phrase these as a named vector.

Model fit for orange tree data

```
> print(nm1 <- nlmer(circumference ~ SSlogis(age,
+     Asym, xmid, scal) ~ Asym | Tree, Orange, start = c(Asym =
+     xmid = 770, scal = 120)), corr = FALSE)
```

```
Nonlinear mixed model fit by the Laplace approximation
Formula: circumference ~ SSlogis(age, Asym, xmid, scal) ~ Asym | Tree
   Data: Orange
  AIC  BIC logLik deviance
 1901 1908 -945.3     1891
Random effects:
 Groups   Name Variance  Std.Dev.
 Tree     Asym 53985.368 232.348
 Residual        52.868   7.271
Number of obs: 35, groups: Tree, 5
Fixed effects:
     Estimate Std. Error t value
Asym   192.04     104.09   1.845
xmid   727.89      31.97  22.771
scal   347.97      24.42  14.252
```
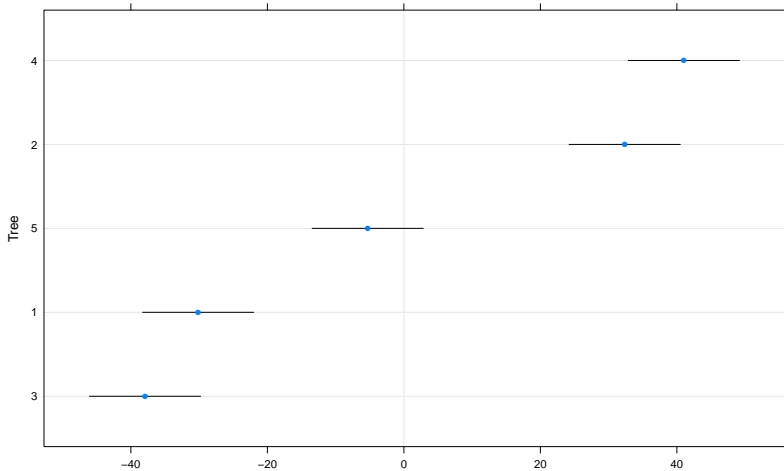
Random effects for trees

Extending the model

- Model `nm1` incorporates random effects for the asymptote only. The asymptote parameter occurs linearly in the model expression. When random effects are associated with only such *conditionally linear* parameters, the Laplace approximation to the deviance is exact.

- We can allow more general specifications of random effects. In practice it is difficult to estimate many variance and covariance parameters when the number of levels of the grouping factor (`Tree`) is small.

- Frequently we begin with independent random effects to see which parameters show substantial variability. Later we allow covariances.

- This is not a fool-proof modeling strategy by any means but it is somewhat reasonable.

Independent random effects for each parameter

```
Nonlinear mixed model fit by the Laplace approximation
Formula: circumference ~ SSlogis(age, Asym, xmid, scal) ~ (Asym | Tree)
   Data: Orange
  AIC  BIC logLik deviance
 1381 1392 -683.6     1367
Random effects:
 Groups   Name Variance    Std.Dev.
 Tree     Asym 34038.004   184.4939
 Tree     xmid 201573.105  448.9689
 Tree     scal 42152.970   205.3119
 Residual           36.817   6.0677
Number of obs: 35, groups: Tree, 5
Fixed effects:
     Estimate Std. Error t value
Asym   192.77      82.69   2.331
xmid   726.14     203.17   3.574
scal   355.44      94.71   3.753
```

Correlated random effects for Asym and scal only

```
Nonlinear mixed model fit by the Laplace approximation
Formula: circumference ~ SSlogis(age, Asym, xmid, scal) ~ (Asym + scal
   Data: Orange
  AIC  BIC logLik deviance
 1573 1584 -779.7     1559
Random effects:
 Groups    Name Variance   Std.Dev. Corr
 Tree      Asym 36734.899 191.6635
           scal 93569.170 305.8908 -0.680
 Residual          42.887   6.5488
Number of obs: 35, groups: Tree, 5
Fixed effects:
     Estimate Std. Error t value
Asym   194.09      85.89   2.260
xmid   735.97      28.75  25.595
scal   365.99     138.73   2.638
```
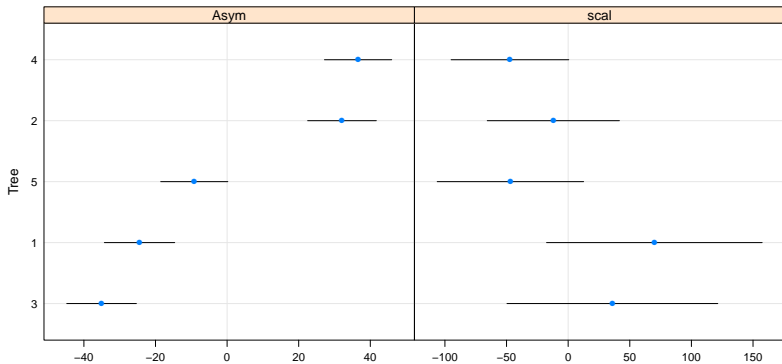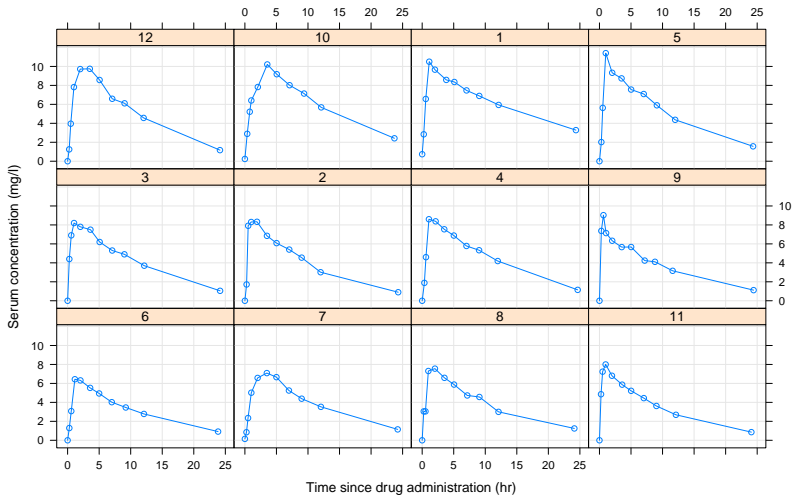
Singular variance-covariance matrix

Theophylline pharmacokinetics

Initial fit of first-order model

```
Nonlinear mixed model fit by the Laplace approximation
Formula: conc ~ SSfol(Dose, Time, lKe, lKa, lCl) ~ (lKe + lKa + lCl |
   Data: Theoph
   AIC   BIC logLik deviance
 152.1 181.0 -66.07    132.1
Random effects:
 Groups   Name Variance Std.Dev. Corr
 Subject  lKe  0.000000 0.00000
          lKa  0.227357 0.47682      NaN
          lCl  0.015722 0.12539      NaN -0.012
 Residual      0.591717 0.76923
Number of obs: 132, groups: Subject, 12
Fixed effects:
    Estimate Std. Error t value
lKe -2.47519    0.05641  -43.88
lKa  0.47414    0.15288    3.10
lCl -3.23550    0.05235  -61.80
```

Remove random effect for lKe

```
Nonlinear mixed model fit by the Laplace approximation
Formula: conc ~ SSfol(Dose, Time, lKe, lKa, lCl) ~ (lKa + lCl | Subject
   Data: Theoph
   AIC   BIC logLik deviance
 146.1 166.3 -66.07    132.1
Random effects:
 Groups    Name Variance Std.Dev. Corr
 Subject   lKa  0.227362 0.47682
           lCl  0.015722 0.12539  -0.012
 Residual       0.591715 0.76923
Number of obs: 132, groups: Subject, 12
Fixed effects:
    Estimate Std. Error t value
lKe -2.47518    0.05641  -43.88
lKa  0.47415    0.15288    3.10
lCl -3.23552    0.05235  -61.80
```

Remove correlation

```
> print(nm6 <- nlmer(conc ~ SSfol(Dose, Time, lKe,
+     lKa, lCl) ~ (lKa | Subject) + (lCl | Subject),
+     Theoph, start = Th.start), corr = FALSE)

Nonlinear mixed model fit by the Laplace approximation
Formula: conc ~ SSfol(Dose, Time, lKe, lKa, lCl) ~ (lKa | Subject) + (l
   Data: Theoph
   AIC    BIC  logLik deviance
 144.1 161.4  -66.07    132.1
Random effects:
 Groups    Name Variance Std.Dev.
 Subject   lKa  0.227493 0.47696
 Subject   lCl  0.015739 0.12545
 Residual       0.591690 0.76921
Number of obs: 132, groups: Subject, 12
Fixed effects:
    Estimate Std. Error t value
lKe -2.47500    0.05641  -43.88
lKa  0.47408    0.15291    3.10
lCl -3.23538    0.05236  -61.79
```
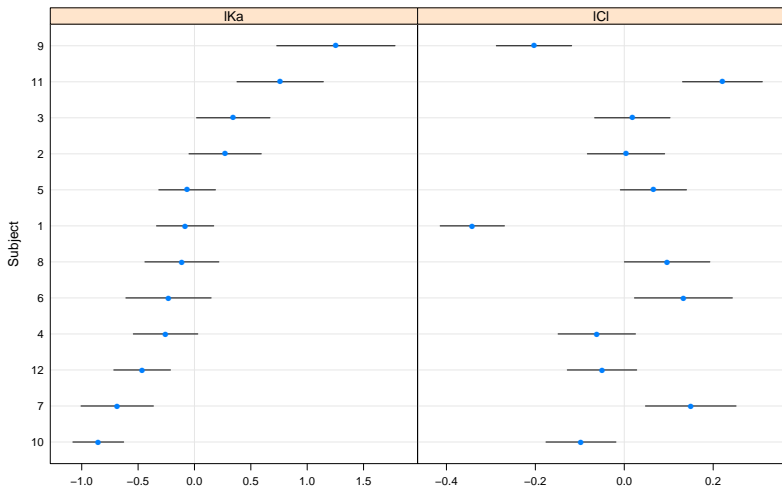
Random effects for clearance and absorption

Methodology

- Evaluation of the deviance is very similar to the calculation for the generalized linear mixed model. For given parameter values $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ the conditional mode $\tilde{\boldsymbol{u}}(\boldsymbol{\theta}, \boldsymbol{\beta})$ is determined by solving a penalized nonlinear least squares problem.

- $r^2(\boldsymbol{\theta}, \boldsymbol{\beta})$ and $|\boldsymbol{L}|^2$ determine the Laplace approximation to the deviance.

- As for GLMMs this can (and will) be extended to an adaptive Gauss-Hermite quadrature evaluation when there is only one grouping factor for the random effects.

- The theory (and, I hope, the implementation) for the generalized nonlinear mixed model (GNLMM) is straightforward, once you get to this point. Map first through the nonlinear model function then through the inverse link function.

From linear predictor to $\boldsymbol{\mu}$

- The main change in evaluating $\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}$ for NLMMs is in the role of the linear predictor. If there are s nonlinear model (nm) parameters and n observations in total then the model matrix X is $n \cdot s \times p$ and the model matrix Z is $n \cdot s \times q$.

- The linear predictor, $\boldsymbol{v} = \boldsymbol{X\beta} + \boldsymbol{Uu}$, of length $n \cdot s$, is rearranged as an $n \times s$ matrix of parameter values $\boldsymbol{\Phi}$. The ith component of the unbounded predictor, $\boldsymbol{\eta}$, is the nonlinear model evaluated for the i set of covariate values with the nonlinear parameters, ϕ, at the ith row of $\boldsymbol{\Phi}$.

$$\boldsymbol{u} \to \boldsymbol{b} \to \boldsymbol{v} \to \boldsymbol{\Phi} \to \boldsymbol{\eta} \to \boldsymbol{\mu}$$
$$\boldsymbol{b} = \boldsymbol{\Lambda}(\boldsymbol{\theta})\boldsymbol{u}$$
$$\boldsymbol{v} = \boldsymbol{X\beta} + \boldsymbol{Zb} = \boldsymbol{X\beta} + \boldsymbol{U}(\boldsymbol{\theta})\boldsymbol{P}'\boldsymbol{u} = \mathrm{vec}(\boldsymbol{\Phi})$$
$$\boldsymbol{\eta} = \boldsymbol{f}(\boldsymbol{t}, \boldsymbol{\Phi})$$
$$\boldsymbol{\mu} = \boldsymbol{g}^{-1}\boldsymbol{\eta}$$

Generalizations of PIRLS

- The reason that the PLS problem for determining the conditional modes is relatively easy is because the standard least squares-based methods for fixed-effects models are easily adapted.

- For linear mixed-models the PLS problem is solved directly. In fact, for LMMs it is possible to determine the conditional modes of the random effects and the conditional estimates of the fixed effects simultaneously.

- Parameter estimates for generalized linear models (GLMs) are (very efficiently) determined by iteratively re-weighted least squares (IRLS) so the conditional modes in a GLMM are determined by penalized iteratively re-weighted least squares (PIRLS).

- Nonlinear least squares, used for fixed-effects nonlinear regression, is adapted as penalized nonlinear least squares (PNLS) or penalized iteratively reweighted nonlinear least squares (PIRNLS) for generalized nonlinear mixed models.