Generalized Linear and Nonlinear Mixed-Effects Models

Douglas Bates

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

University of Potsdam
August 8, 2008

Outline

Outline

Outline

Outline

Outline

Generalized Linear Mixed Models

- When using linear mixed models (LMMs) we assume that the response being modeled is on a continuous scale.
- Sometimes we can bend this assumption a bit if the response is an ordinal response with a moderate to large number of levels. For example, the Scottish secondary school test results were integer values on the scale of 1 to 10.
- However, an LMM is not suitable for modeling a binary response, an ordinal response with few levels or a response that represents a count. For these we use generalized linear mixed models (GLMMs).
- To describe GLMMs we return to the representation of the response as an n-dimensional, vector-valued, random variable, $\mathcal{Y}$, and the random effects as a q-dimensional, vector-valued, random variable, $\mathcal{B}$.

Parts of LMMs carried over to GLMMs

- Random variables
  - $\mathcal{Y}$ the response variable
  - $\mathcal{B}$ the (possibly correlated) random effects
  - $\mathcal{U}$ the orthogonal random effects
- Parameters
  - $\boldsymbol{\beta}$ - fixed-effects coefficients
  - σ - the common scale parameter (not always used)
  - $\boldsymbol{\theta}$ - parameters that determine $\mathrm{Var}(\mathcal{B}) = \sigma^2(\boldsymbol{TS})(\boldsymbol{TS})'$
- Some matrices
  - $\boldsymbol{X}$ the $n \times p$ model matrix for $\boldsymbol{\beta}$
  - $\boldsymbol{Z}$ the $n \times q$ model matrix for $\boldsymbol{b}$
  - $\boldsymbol{P}$ fill-reducing $q \times q$ permutation (from $\boldsymbol{Z}$)
  - $\boldsymbol{S}(\boldsymbol{\theta})$ non-negative $q \times q$ diagonal scale matrix
  - $\boldsymbol{T}(\boldsymbol{\theta})$ $q \times q$ unit lower-triangular matrix
  - $\boldsymbol{A}(\boldsymbol{\theta}) = (\boldsymbol{Z}\boldsymbol{T}(\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{\theta}))'$

The conditional distribution, $\mathcal{Y}|\mathcal{U}$

- For GLMMs, the marginal distribution, $\mathcal{B} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\boldsymbol{\theta}))$ is the same as in LMMs except that σ^2 is omitted. We define $\mathcal{U} \sim (\mathbf{0}, \boldsymbol{I}_q)$ such that $\mathcal{B} = \boldsymbol{T}(\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{\theta})\boldsymbol{P}'\mathcal{U}$.
- For GLMMs we retain some of the properties of the conditional distribution

  $$(\mathcal{Y}|\mathcal{U} = \boldsymbol{u}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}, \sigma^2 \boldsymbol{I}\right) \text{ where } \boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}'\boldsymbol{P}'\boldsymbol{u}$$

  Specifically
  - The distribution $\mathcal{Y}|\mathcal{U} = \boldsymbol{u}$ depends on $\boldsymbol{u}$ only through the conditional mean, $\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u})$.
  - Elements of $\mathcal{Y}$ are *conditionally independent*. That is, the distribution of $\mathcal{Y}|\mathcal{U} = \boldsymbol{u}$ is completely specified by the univariate, conditional distributions, $\mathcal{Y}_i|\mathcal{U}, i = 1, \dots, n$.
  - These univariate, conditional distributions all have the same form. They differ only in their means.
- GLMMs differ from LMMs in the form of the univariate, conditional distributions and in how $\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u})$ depends on $\boldsymbol{u}$.

Some choices of univariate conditional distributions

- Typical choices of univariate conditional distributions are:
  - The *Bernoulli* distribution for binary (0/1) data, which has probability mass function

$$p(y|\mu) = \mu^y (1-\mu)^{1-y}, \quad 0 < \mu < 1, \quad y = 0, 1$$

  - Several independent binary responses can be represented as a *binomial* response, but only if all the Bernoulli distributions have the same mean.
  - The *Poisson* distribution for count $(0, 1, \dots)$ data, which has probability mass function

$$p(y|\mu) = e^{-y} \frac{\mu^y}{y!}, \quad 0 < \mu, \quad y = 0, 1, 2, \dots$$

- All of these distributions are completely specified by the conditional mean. This is different from the conditional normal (or Gaussian) distribution, which also requires the common scale parameter, σ.

The link function, g

- When the univariate conditional distributions have constraints on μ, such as $0 < \mu < 1$ (Bernoulli) or $0 < \mu$ (Poisson), we cannot define the conditional mean, $\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}$, to be equal to the linear predictor, $\boldsymbol{X\beta} + \boldsymbol{A}'\boldsymbol{P}'\boldsymbol{u}$, which is unbounded.
- We choose an invertible, univariate *link function*, g, such that $\eta = g(\mu)$ is unconstrained. The vector-valued link function, $\boldsymbol{g}$, is defined by applying g component-wise.

$$\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{\mu}) \quad \text{where} \quad \eta_i = g(\mu_i), \quad i = 1, \ldots, n$$

- We require that g be invertible so that $\mu = g^{-1}(\eta)$ is defined for $-\infty < \eta < \infty$ and is in the appropriate range ($0 < \mu < 1$ for the Bernoulli or $0 < \mu$ for the Poisson). The vector-valued inverse link, $\boldsymbol{g}^{-1}$, is defined component-wise.

"Natural" link functions

- There are many choices of invertible scalar link functions, g, that we could use for a given set of constraints.
- For the Bernoulli and Poisson distributions, however, one link function arises naturally from the definition of the probability mass function. (The same is true for a few other, related but less frequently used, distributions, such as the gamma distribution.)
- To derive the natural link, we consider the logarithm of the probability mass function (or, for continuous distributions, the probability density function).
- For distributions in this "exponential" family, the logarithm of the probability mass or density can be written as a sum of terms, some of which depend on the response, y, only and some of which depend on the mean, μ, only. However, only one term depends on **both** y and μ, and this term has the form $y \cdot g(\mu)$, where g is the natural link.

The natural link for the Bernoulli distribution

- The logarithm of the probability mass function is

$$\log(p(y|\mu)) = \log(1-\mu) + y \log\left(\frac{\mu}{1-\mu}\right), \ 0 < \mu < 1, \ y = 0, 1.$$

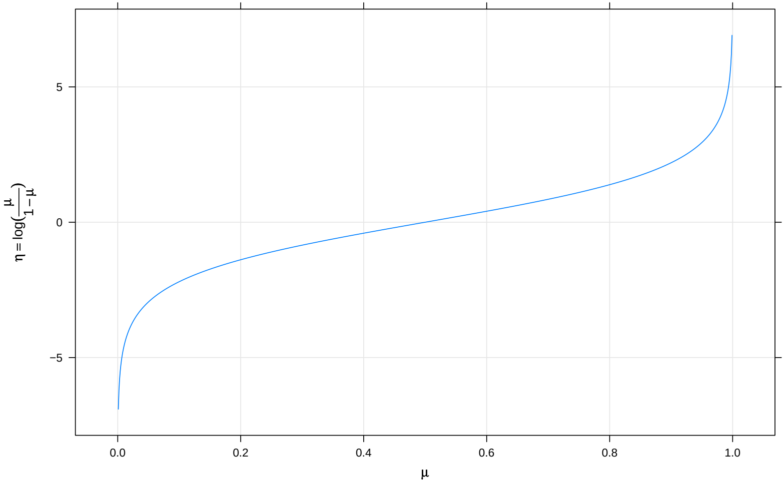
- Thus, the natural link function is the *logit* link

$$\eta = g(\mu) = \log\left(\frac{\mu}{1-\mu}\right).$$

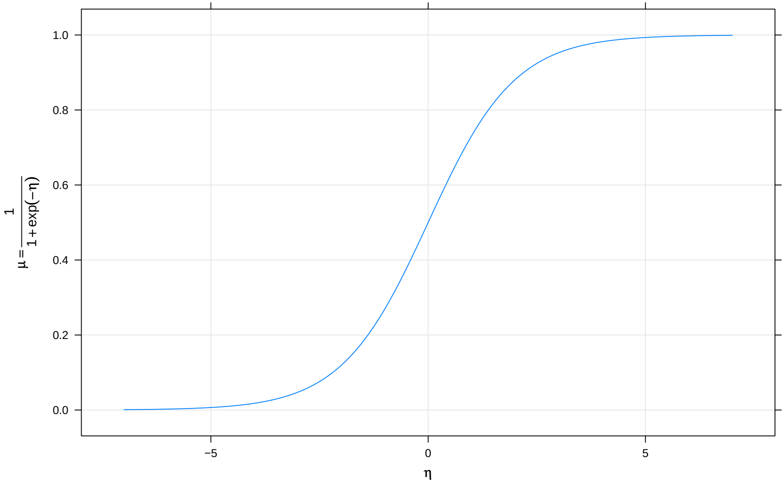
- Because $\mu = P[\mathcal{Y} = 1]$, the quantity $\mu/(1-\mu)$ is the odds ratio (in the range $(0, \infty)$) and g is the logarithm of the odds ratio, sometimes called "log odds".
- The inverse link is

$$\mu = g^{-1}(\eta) = \frac{e^{\eta}}{1+e^{\eta}} = \frac{1}{1+e^{-\eta}}$$

Plot of natural link for the Bernoulli distribution



Plot of inverse natural link for the Bernoulli distribution



The natural link for the Poisson distribution

- The logarithm of the probability mass is

$$\log(p(y|\mu)) = \log(y!) - \mu + y\log(\mu)$$

- Thus, the natural link function for the Poisson is the *log* link

$$\eta = g(\mu) = \log(\mu)$$

- The inverse link is

$$\mu = g^{-1}(\eta) = e^{\eta}$$

The natural link related to the variance

- For the natural link function, the derivative of its inverse is the variance of the response.
- For the Bernoulli, the natural link is the logit and the inverse link is $\mu = g^{-1}(\eta) = 1/(1+e^{-\eta})$. Then

$$\frac{d\mu}{d\eta} = \frac{e^{-\eta}}{(1+e^{-\eta})^2} = \frac{1}{1+e^{-\eta}}\frac{e^{-\eta}}{1+e^{-\eta}} = \mu(1-\mu) = \mathrm{Var}(\mathcal{Y})$$

- For the Poisson, the natural link is the log and the inverse link is $\mu = g^{-1}(\eta) = e^{\eta}$. Then

$$\frac{d\mu}{d\eta} = e^{\eta} = \mu = \mathrm{Var}(\mathcal{Y})$$

The unscaled conditional density of $\mathcal{U}|\mathcal{Y} = \boldsymbol{y}$

- As in LMMs we evaluate the likelihood of the parameters, given the data, as

$$L(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{y}) = \int_{\mathbb{R}^q} [\mathcal{Y}|\mathcal{U}](\boldsymbol{y}|\boldsymbol{u}) \, [\mathcal{U}](\boldsymbol{u}) \, d\boldsymbol{u},$$

- The product $[\mathcal{Y}|\mathcal{U}](\boldsymbol{y}|\boldsymbol{u})[\mathcal{U}](\boldsymbol{u})$ is the unscaled (or *unnormalized*) density of the conditional distribution $\mathcal{U}|\mathcal{Y}$.
- The density $[\mathcal{U}](\boldsymbol{u})$ is a spherical Gaussian density $\frac{1}{(2\pi)^{q/2}} e^{-\|\boldsymbol{u}\|^2/2}$.
- The expression $[\mathcal{Y}|\mathcal{U}](\boldsymbol{y}|\boldsymbol{u})$ is the value of a probability mass function or a probability density function, depending on whether $\mathcal{Y}_i|\mathcal{U}$ is discrete or continuous.
- The linear predictor is $g(\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}(\boldsymbol{\theta})'\boldsymbol{P}'\boldsymbol{u}$. Alternatively, we can write the conditional mean of $\mathcal{Y}$, given $\mathcal{U}$, as

$$\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}(\boldsymbol{u}) = \boldsymbol{g}^{-1}\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}(\boldsymbol{\theta})'\boldsymbol{P}'\boldsymbol{u}\right)$$

The conditional mode of $\mathcal{U}|\mathcal{Y} = \boldsymbol{y}$

- In general the likelihood, $L(\boldsymbol{\theta}, \boldsymbol{\beta}|\boldsymbol{y})$ does not have a closed form. To approximate this value, we first determine the *conditional mode*

$$\tilde{\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \arg\max_{\boldsymbol{u}} [\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{U}}](\boldsymbol{y}|\boldsymbol{u})\,[\boldsymbol{\mathcal{U}}](\boldsymbol{u})$$

  using a quadratic approximation to the logarithm of the unscaled conditional density.
- This optimization problem is (relatively) easy because the quadratic approximation to the logarithm of the unscaled conditional density can be written as a penalized, weighted residual sum of squares,

$$\tilde{\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \arg\min_{\boldsymbol{u}} \left\| \begin{bmatrix} \boldsymbol{W}^{1/2}(\boldsymbol{\mu})\left(\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{U}}}(\boldsymbol{u})\right) \\ -\boldsymbol{u} \end{bmatrix} \right\|^2$$

  where $\boldsymbol{W}(\boldsymbol{\mu})$ is the diagonal weights matrix. The weights are the inverses of the variances of the $\mathcal{Y}_i$.

The PIRLS algorithm

- Parameter estimates for generalized linear models (without random effects) are usually determined by iteratively reweighted least squares (IRLS), an incredibly efficient algorithm. PIRLS is the penalized version. It is iteratively reweighted in the sense that parameter estimates are determined for a fixed weights matrix $\boldsymbol{W}$ then the weights are updated to the current estimates and the process repeated.
- For fixed weights we solve

$$\min_{\boldsymbol{u}} \left\| \begin{bmatrix} \boldsymbol{W}^{1/2} \left(\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{U}}}(\boldsymbol{u}) \right) \\ -\boldsymbol{u} \end{bmatrix} \right\|^2$$

as a nonlinear least squares problem with update, $\boldsymbol{\delta_u}$, given by

$$\boldsymbol{P} \left(\boldsymbol{A M W M A}' + \boldsymbol{I} \right) P' \boldsymbol{\delta_u} = \boldsymbol{P A M W} (\boldsymbol{y} - \boldsymbol{\mu}) - \boldsymbol{u}$$

where $\boldsymbol{M} = d\boldsymbol{\mu}/d\boldsymbol{\eta}$ is the (diagonal) Jacobian matrix. Recall that for the natural link, $\boldsymbol{M} = \mathrm{Var}(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{U}}) = \boldsymbol{W}^{-1}$.

The Laplace approximation to the deviance

- At convergence, the sparse Cholesky factor, $\boldsymbol{L}$, used to evaluate the update is

$$\boldsymbol{L}\boldsymbol{L}' = \boldsymbol{P}\left(\boldsymbol{A}\boldsymbol{M}\boldsymbol{W}\boldsymbol{M}\boldsymbol{A}' + \boldsymbol{I}\right)\boldsymbol{P}'$$

or

$$\boldsymbol{L}\boldsymbol{L}' = \boldsymbol{P}\left(\boldsymbol{A}\boldsymbol{M}\boldsymbol{A}' + \boldsymbol{I}\right)\boldsymbol{P}'$$

if we are using the natural link.

- The integrand of the likelihood is approximately a constant times the density of the $\mathcal{N}(\tilde{\boldsymbol{u}}, \boldsymbol{L}\boldsymbol{L}')$ distribution.
- On the deviance scale (negative twice the log-likelihood) this corresponds to

$$d(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{y}) = d_g(\boldsymbol{y}, \boldsymbol{\mu}(\tilde{\boldsymbol{u}})) + \|\tilde{\boldsymbol{u}}\|^2 + \log(|\boldsymbol{L}|^2)$$

where $d_g(\boldsymbol{y}, \boldsymbol{\mu}(\tilde{\boldsymbol{u}}))$ is the GLM deviance for $\boldsymbol{y}$ and $\boldsymbol{\mu}$.

Modifications to the algorithm

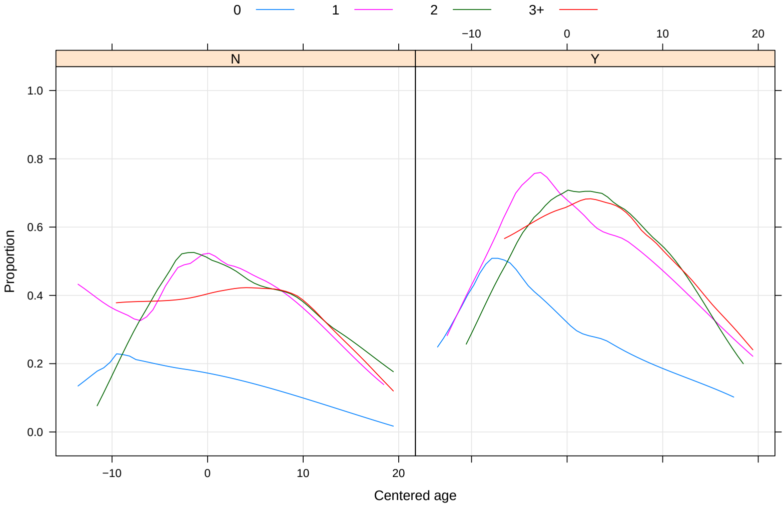
- Notice that this deviance depends on the fixed-effects parameters, $\boldsymbol{\beta}$, as well as the variance-component parameters, $\boldsymbol{\theta}$. This is because $\log(|\boldsymbol{L}|^2)$ depends on $\boldsymbol{\mu}_{\mathcal{Y}|\mathcal{U}}$ and, hence, on $\boldsymbol{\beta}$. For LMMs $\log(|\boldsymbol{L}|^2)$ depends only on $\boldsymbol{\theta}$.
- It is likely that modifying the PIRLS algorithm to optimize simultaneously on $\boldsymbol{u}$ and $\boldsymbol{\beta}$ would result in a value that is very close to the deviance profiled over $\boldsymbol{\beta}$.
- Another approach, which is being implemented as a Google Summer of Code project, is adaptive Gauss-Hermite quadrature (AGQ). This has a similar structure to the Laplace approximation but is based on more evaluations of the unscaled conditional density near the conditional modes. It is only appropriate for models in which the random effects are associated with only one grouping factor

Outline

The Contraception data set

- One of the data sets in the "mlmRev" package, derived from data files available on the multilevel modelling web site, is from a fertility survey of women in Bangladesh.
- We consider a binary response - whether the woman currently uses artificial contraception.
- Covariates included the woman's age (on a centered scale), the number of live children she has, whether she lives in an urban or rural setting, and the district in which she lives.
- These data are quite unbalanced with regard to the covariates (some districts have only 2 observations, some have nearly 120).
- We should bear in mind that the binary responses have low per-observation information content (exactly one bit per observation). Districts with few observations will not contribute strongly to estimates of random effects.
- Within-district plots will be too rough. We can examine the influence of some of the covariates by plotting scatterplot

Contraception use versus age by urban and livch



Comments on the data plot

- On the multilevel modelling web site they compare various forms of multilevel software fitting a model that is linear in `age` to these data. The model is clearly inappropriate.
- The form of the curves suggests at least a quadratic in `age`. Once you see the plot it is obvious why this should be so. (They don't say what age correspond to 0 on this scale but my guess is about 25 years of age.)
- The urban versus rural differences may be additive.
- It appears that the `livch` factor could be dichotomized into "0" versus "1 or more".

Preliminary model fit

```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ age + I(age^2) + urban + livch + (1 | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2389 2433  -1186     2373
Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.22586  0.47524
Number of obs: 1934, groups: district, 60
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0350725  0.1743606  -5.936 2.91e-09
age          0.0035328  0.0092311   0.383    0.702
I(age^2)    -0.0045623  0.0007252  -6.291 3.15e-10
urbanY       0.6972694  0.1198788   5.816 6.01e-09
livch1       0.8150448  0.1621898   5.025 5.03e-07
livch2       0.9165107  0.1850995   4.951 7.37e-07
livch3+      0.9150210  0.1857689   4.926 8.41e-07
```

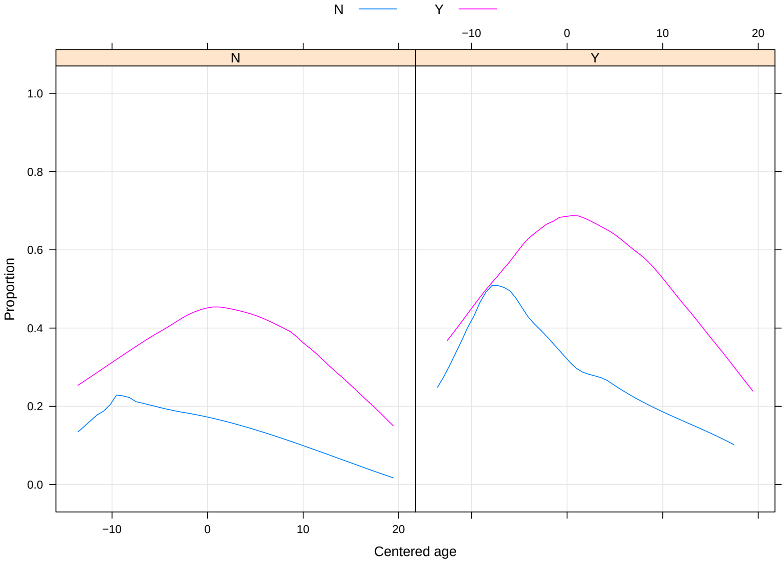
Comments on the model fit

- There is a highly significant quadratic term in `age`.
- The linear term in `age` is not significant but we retain it because the `age` scale has been centered at an arbitrary (and unknown) value.
- The `urban` factor is highly significant (as indicated by the plot).
- Levels of `livch` greater than 0 are significantly different from 0 but may not be different from each other.

Interpreting coefficient estimates

- We are using the logit link, which is the natural link for the Bernoulli.
- For the logit link the coefficients apply to the linear predictor of the log-odds.
- The intercept is the predicted log-odds for a woman with centered age of 0 (we expect this means an age of 25 or so), not in an urban environment and with 0 live children. As an odds ratio this is $e^{-1.035} = 0.355$ and a probability of 0.262.
- For a dichotomous factor like `urban` the coefficient 0.697 is the increase in the log-odds for urban versus rural when other covariates are held fixed.
- This corresponds to multiplication of the odds-ratio by $e^{0.697} = 2.008$. The predicted probability of contraception use for a woman with centered age 0 in an urban environment with no live children is 0.416 (it is the odds-ratio, not the probability, that is multiplied by 2.008)

Consider dichotomizing livch to 0/1+



Reduced model with dichotomized livch

```
> print(cm2 <- glmer(use ~ age + I(age^2) + urban + ch +
+     (1 | district), Contraception, binomial), corr = FALSE)
```

```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ age + I(age^2) + urban + ch + (1 | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2385 2419  -1187     2373
Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.22470  0.47402
Number of obs: 1934, groups: district, 60
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0064262  0.1678949  -5.994 2.04e-09
age          0.0062563  0.0078404   0.798    0.425
I(age^2)    -0.0046354  0.0007163  -6.471 9.73e-11
urbanY       0.6929504  0.1196687   5.791 7.01e-09
chY          0.8603758  0.1473539   5.839 5.26e-09
```

Comparing the model fits

- A likelihood ratio test can be used to compare these nested models.

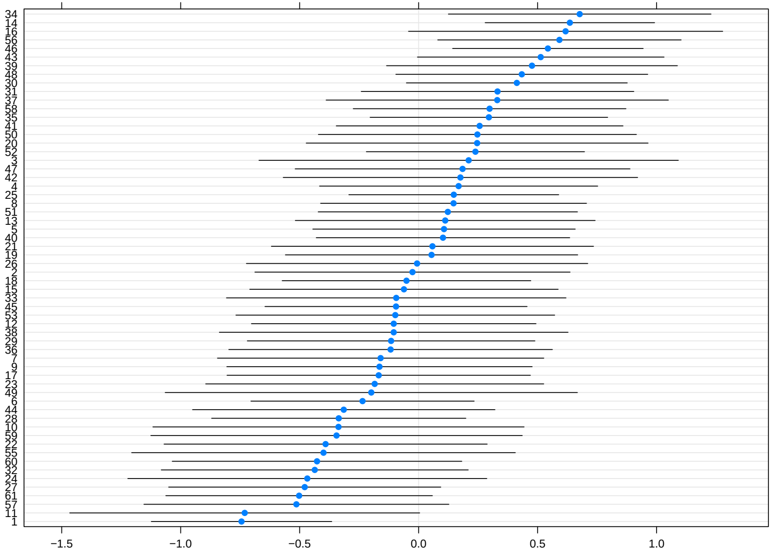
```
> anova(cm2, cm1)

Data: Contraception
Models:
cm2: use ~ age + I(age^2) + urban + ch + (1 | district)
cm1: use ~ age + I(age^2) + urban + livch + (1 | district)
    Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
cm2  6 2385.2 2418.6 -1186.6
cm1  8 2388.7 2433.3 -1186.4 0.4571      2     0.7957
```

- The large p-value indicates that we would not reject `cm2` in favor of `cm1` hence we prefer the more parsimonious `cm2`.
- The plot of the scatterplot smoothers according to live children or none indicates that there may be a difference in the age pattern between these two groups.

Allowing age pattern to vary with ch

```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ age * ch + I(age^2) + urban + (1 | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2379 2418  -1183     2365
Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.22306  0.4723
Number of obs: 1934, groups: district, 60
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3233176  0.2144470  -6.171 6.79e-10
age         -0.0472956  0.0218394  -2.166   0.0303
chY          1.2107858  0.2069938   5.849 4.93e-09
I(age^2)    -0.0057572  0.0008358  -6.888 5.64e-12
urbanY       0.7140326  0.1202579   5.938 2.89e-09
age:chY      0.0683522  0.0254347   2.687   0.0072
```

Prediction intervals on the random effects



Extending the random effects

- We may want to consider allowing a random effect for urban/rural by district. This is complicated by the fact the many districts only have rural women in the study

```
      district
urban  1  2  3  4  5  6  7  8  9 10 11 12 13  14 15 16
    N 54 20  0 19 37 58 18 35 20 13 21 23 16  17 14 18
    Y 63  0  2 11  2  7  0  2  3  0  0  6  8 101  8  2
      district
urban 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
    N 24 33 22 15 10 20 15 14 49 13 39 45 25 45 27 24
```

Including a random effect for urban by district

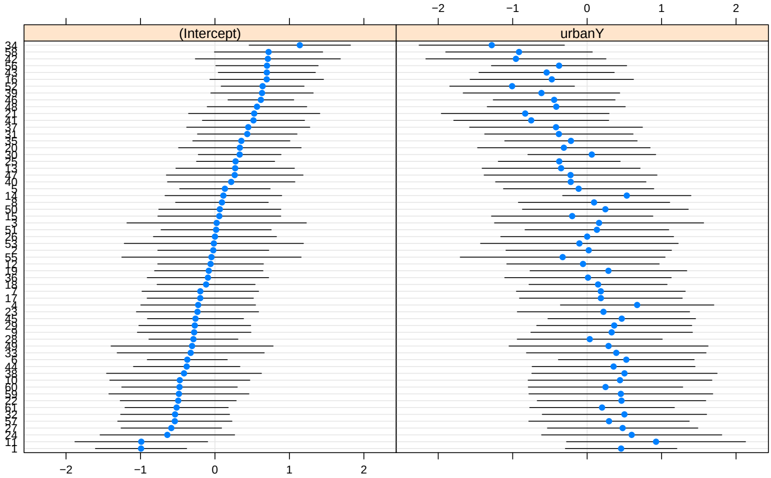
```
Generalized linear mixed model fit by the Laplace approximation
Formula: use ~ age * ch + I(age^2) + urban + (urban | district)
   Data: Contraception
  AIC  BIC logLik deviance
 2372 2422  -1177     2354
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 district (Intercept) 0.37830  0.61506
          urbanY      0.52613  0.72535  -0.793
Number of obs: 1934, groups: district, 60
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3442631  0.2227667  -6.034 1.60e-09
age         -0.0461836  0.0219446  -2.105  0.03533
chY          1.2116527  0.2082373   5.819 5.93e-09
I(age^2)    -0.0056514  0.0008431  -6.703 2.04e-11
urbanY       0.7902095  0.1600484   4.937 7.92e-07
age:chY      0.0664682  0.0255674   2.600  0.00933
```

Significance of the additional random effect

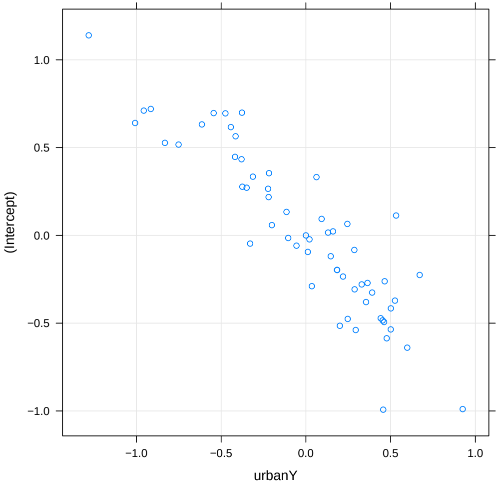
```
> anova(cm4, cm3)

Data: Contraception
Models:
cm3: use ~ age * ch + I(age^2) + urban + (1 | district)
cm4: use ~ age * ch + I(age^2) + urban + (urban | district)
    Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
cm3  7  2379.2  2418.2 -1182.6
cm4  9  2371.5  2421.6 -1176.8 11.651      2   0.002951
```

- The additional random effect is highly significant in this test.
- Most of the prediction intervals still overlap zero.
- A scatterplot of the random effects shows several random effects vectors falling along a straight line. These are the districts with all rural women or all urban women.

Prediction intervals for the bivariate random effects



Scatter plot of the conditional modes



Conclusions from the example

- Again, carefully plotting the data is enormously helpful in formulating the model.
- Observational data tend to be unbalanced and have many more covariates than data from a designed experiment. Formulating a model is often more difficult than in a designed experiment.
- A generalized linear model family, typically `binomial` or `poisson`, is specified as the `family` argument in the call to `glmer`.
- We use likelihood-ratio tests and z-tests in the model building.

Outline

Item Response Models

- Models for binary (or ordered categorical) data that are cross-classified according to subject and item are sometimes called *Item Response* or *IRT* (Item Response Theory) models.
- There is a long history of models for such data with many contributors. Only recently have statisticians become aware of this literature and considered how such models could be framed in the context of GLMMs.
- Even when approaching IRT models as GLMMs they were not expressed as GLMMs with crossed random effects, because of software limitations.
- Because `glmer` can fit GLMMs with crossed random effects, we can approach such models as GLMMs with random effects for subject and item.

Data from a study of verbal aggression

- Results on a study of verbal aggression, used as an example through the book *Expanatory Item Response Models*, edited by De Boeck and Wilson (Springer, 2004) are available as the data set `VerbAgg`, in the "long" format.
- The items correspond to scenarios for which the subject was asked if they would curse, scold or shout.
- The scenarios are classified according to the behavior mode (want versus do) and according to the situation (self-to-blame versus other-to-blame).
- The subjects are classified by sex. Each subject's score on a separately administered anger index (STAXI) is given.
- The response was recorded on a three-level ordinal scale ("no", "perhaps" and "yes"). We will consider a dichotomous version, `r2`.

Structure of VerbAgg data

- We also check that the item-level covariates and the person-level covariates are consistently defined.

```
> str(VerbAgg)
'data.frame':	7584 obs. of  9 variables:
 $ Anger : int  20 11 17 21 17 21 39 21 24 16 ...
 $ Gender: Factor w/ 2 levels "M","F": 2 2 1 1 1 1 1 1 1 1 ...
 $ item  : Factor w/ 24 levels "S1wantcurse",..: 1 1 1 1 1 1 1 1 1 1 ..
 $ resp  : Ord.factor w/ 3 levels "no"<"perhaps"<..: 1 1 2 2 2 3 3 1 1
 $ id    : Factor w/ 316 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 1
 $ btype : Factor w/ 3 levels "curse","scold",..: 1 1 1 1 1 1 1 1 1 1 .
 $ situ  : Factor w/ 2 levels "other","self": 1 1 1 1 1 1 1 1 1 1 ...
 $ mode  : Factor w/ 2 levels "want","do": 1 1 1 1 1 1 1 1 1 1 ...
 $ r2    : Factor w/ 2 levels "N","Y": 1 1 2 2 2 2 2 1 1 2 ...
```

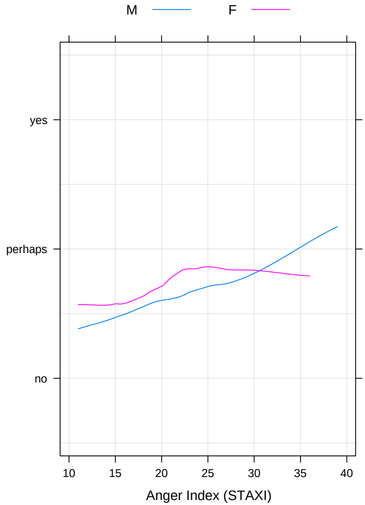
```
> stopifnot(nrow(unique(subset(VerbAgg, select = c(item,
+     btype, situ, mode)))) == 24, nrow(unique(subset(VerbAgg,
+     select = c(id, Anger, Gender)))) == 316)
```

Influence of item-level covariates

- We can check the proportions of responses for combinations of item-level covariates

```
> round(100 * ftable(prop.table(xtabs(~mode + situ +
+     resp, VerbAgg), 1:2)), 1)

           resp   no perhaps  yes
mode situ
want other       37.7    30.0 32.3
     self        55.9    29.1 15.0
do   other       49.8    27.2 23.0
     self        66.2    23.5 10.3
```

Influence of person-level covariates



Initial model fit

```
Generalized linear mixed model fit by the Laplace approximation
Formula: r2 ~ Anger * Gender + situ + btype + mode + (1 | id) + (1 | it
   Data: VerbAgg
  AIC  BIC logLik deviance
 8156 8225  -4068     8136
Random effects:
 Groups Name        Variance Std.Dev.
 id     (Intercept) 1.79337  1.33917
 item   (Intercept) 0.11715  0.34226
Number of obs: 7584, groups: id, 316; item, 24
Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.531954   0.433782   1.226  0.22008
Anger          0.058490   0.019474   3.003  0.00267
GenderF        0.402523   0.782316   0.515  0.60688
situself      -1.054279   0.151192  -6.973 3.10e-12
btypescold    -1.059810   0.184156  -5.755 8.67e-09
btypeshout    -2.103818   0.186515 -11.280  < 2e-16
modedo        -0.707055   0.151005  -4.682 2.84e-06
Anger:GenderF -0.004111   0.038170  -0.108  0.91424
```

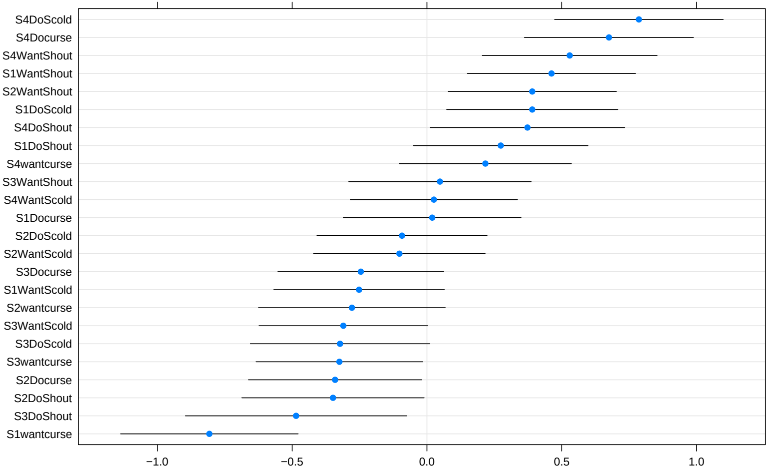
Removing non-significant gender effects

```
Generalized linear mixed model fit by the Laplace approximation
Formula: r2 ~ Anger + situ + btype + mode + (1 | id) + (1 | item)
   Data: VerbAgg
  AIC  BIC logLik deviance
 8155 8210  -4069     8139
Random effects:
 Groups Name        Variance Std.Dev.
 id     (Intercept) 1.81157  1.34595
 item   (Intercept) 0.11720  0.34235
Number of obs: 7584, groups: id, 316; item, 24
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.63927    0.38334   1.668 0.095391
Anger        0.05685    0.01682   3.380 0.000726
situself    -1.05437    0.15122  -6.972 3.12e-12
btypescold  -1.05973    0.18419  -5.753 8.75e-09
btypeshout  -2.10392    0.18655 -11.278  < 2e-16
modedo      -0.70726    0.15104  -4.683 2.83e-06
```

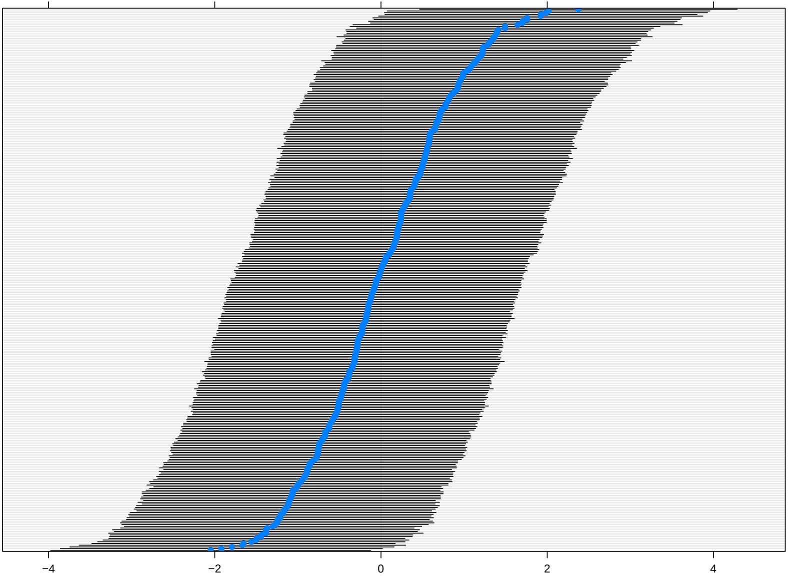
Allowing situational/behavior random effects by person

```
Generalized linear mixed model fit by the Laplace approximation
Formula: r2 ~ Anger + situ + btype + mode + (1 | id:btype) + (1 | id:si
   Data: VerbAgg
  AIC  BIC logLik deviance
 7751 7827  -3865     7729
Random effects:
 Groups   Name        Variance Std.Dev.
 id:btype (Intercept) 1.41069  1.18772
 id:mode  (Intercept) 0.80916  0.89953
 id:situ  (Intercept) 0.61539  0.78447
 id       (Intercept) 1.70950  1.30748
 item     (Intercept) 0.17213  0.41489
Number of obs: 7584, groups: id:btype, 948; id:mode, 632; id:situ, 632;
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.79167    0.48183   1.643 0.100375
Anger        0.07483    0.02107   3.551 0.000384
situself    -1.35742    0.19199  -7.070 1.55e-12
btypescold  -1.36067    0.24118  -5.642 1.68e-08
btypeshout  -2.69333    0.24372 -11.051  < 2e-16
modedo      -0.94095    0.19503  -4.825 1.40e-06
```

Item-specific random effects



Person-specific random effects - Intercept



Correlated random effects by person

```
Generalized linear mixed model fit by the Laplace approximation
Formula: r2 ~ Anger + situ + btype + mode + (1 + situ + btype + mode |
   Data: VerbAgg
  AIC  BIC logLik deviance
 7727 7880  -3842     7683
Random effects:
 Groups Name        Variance Std.Dev. Corr
 id     (Intercept) 4.53799  2.13025
        situself    1.31455  1.14654  -0.521
        btypescold  1.62188  1.27353  -0.085 -0.247
        btypeshout  4.03049  2.00761  -0.374  0.010  0.423
        modedo      1.68341  1.29746  -0.295  0.112  0.102
 item   (Intercept) 0.18222  0.42687
  0.104

Number of obs: 7584, groups: id, 316; item, 24

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.98354    0.47248   2.082  0.03738
Anger        0.06564    0.02041   3.217  0.00130
situself    -1.37646    0.19723  -6.979 2.97e-12
```

Outline