A brief introduction to mixed models

David Bolin
University of Gothenburg

Gothenburg
April 6, 2017
• An introduction to mixed models based on a few examples:
  • Definition of standard mixed models.
  • Parameter estimation.
  • Confidence intervals for parameters.
  • Models with nested factors.
  • Model comparison.
  • Model diagnostics.

• Show how mixed models can be estimated using \texttt{lme4}.

• The lecture is mainly based on
  2 Douglas M. Bates \texttt{lme4}: Mixed-effects modeling with R
An illustrative example

- Study taken from Belenky et al. (2003). Based on 18 subjects.
- On day 0 the subjects had their normal amount of sleep. They were then restricted to 3 hours of sleep per night.
- Each day the average reaction times in milliseconds (ms) on a series of tests is measured for each subject.
It seems like a linear model would fit the data:

\[ Y_{ij} = \beta_1 + t_{ij} \cdot \beta_2 + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \]

Collecting the measurements from subject \( i \) into a vector \( Y_i \):

\[ Y_i = X_i \beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2 I) \]
library("lme4")
data("sleepstudy")
str(sleepstudy)
'data.frame': 180 obs. of 3 variables:
$ Reaction: num 250 259 251 321 357 ...  
$ Days: num 0 1 2 3 4 5 6 7 8 9 ...  
$ Subject: Factor w/ 18 levels "308","309","310",...: 1 1 1 1 1 1 1...

mod = lm(Reaction ~ Days, sleepstudy)

plot(sleepstudy$Days,sleepstudy$Reaction,
     xlab = "Days", ylab = "Response time (ms)")
lines(sleepstudy$Days,mod$fitted.values)
People need different amounts of sleep

- The data shown separately for each subject.
- An *individual* linear regression for each subject.
A random effect model for the data

• For the linear regression (fixed effect model) we had

\[ Y_i = X_i \beta + \varepsilon_i \]

• For a random effect model we instead have

\[ Y_i = X_i \beta + X_i b_i + \varepsilon_i \]

where \( b_i \sim N(0, \Sigma) \).

• \( \beta \) is now the common average regression coefficients and \( b_i \) are the individual deviations from the mean.

• The matrix \( \Sigma \) determines the variances of the deviations from the mean, as well as the correlations between the different random effects.
Mixed models in R

We formulate the random effect model using the formula specification as

\[ \text{Reaction} \sim \text{Days} + (\text{Days} \mid \text{Subject}) \]

We can fit the model using \textit{lme4} as

\[
\text{fm1 <- lmer(Reaction ~ Days + (Days | Subject), sleepstudy)}
\]

Some results

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Std.Dev.</th>
<th>Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>(Intercept)</td>
<td>24.740</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Days</td>
<td>5.922</td>
<td>0.07</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>25.592</td>
<td></td>
</tr>
</tbody>
</table>

Fixed Effects:

<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>251.41</td>
<td>10.47</td>
</tr>
</tbody>
</table>
Mixed models

- For more complicated problems, we may need to use both fixed effects and random effects.
- A general mixed model can be written as

\[ Y_i = X_i \beta + Z_i b_i + \varepsilon_i \]

where the model matrix \( X_i \) and the random effects matrix \( Z_i \) may be different.
- We could for example have a model with a random intercept but a common slope.
- In this case, \( Z_i \) is only the first column of \( X_i \), and \( b_i \sim N(0, \sigma^2_b) \).
- In the formula specification:
  \[ \text{Reaction} \sim \text{Days} + (1 \mid \text{Subject}) \]
A second example

- The Penicillin data from Davies and Gold-smith (1972), used to assess the variability between samples of penicillin by the B. subtilis method.
- In this test, plates (petri dishes) with bulk-innoculated nutrient agar medium has six small pots at equal distance.
- A few drops of the penicillin solutions is placed in each pot, and the whole plate is placed in an incubator for a given time.
- Penicillin diffuses from the pots into the agar. The diameter of the affected zone is related to the concentration of penicillin.
- We have 24 plates, each with 6 pots, giving in total 144 observations.
- The variation in the diameter is associated with the plates and with the samples (pots).
The general definition
The model for the Penicillin data

- We use two random effects: One for plate that accounts for the between-plate variability, and one for pot that accounts for between-sample variability.
- We can write the model as

$$Y_{ij} = \alpha + b_i + b_j + \varepsilon_{ij}$$

where $i$ denotes plate, $j$ denotes pot, and $b_i \sim N(0, \sigma_{plate}^2)$, $b_i \sim N(0, \sigma_{pot}^2)$, $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- We assume that the two random effects are independent, and can therefore specify the model as

$$f = \text{diameter} \sim 1 + (1|\text{plate}) + (1|\text{sample})$$

- To assume independence between slope and intercept in a linear model, one can use $\|$

$$\text{Reaction} \sim \text{Days} + (\text{Days} \| \text{Subject})$$
Estimating the model in R

```r
f = diameter ~ 1 + (1|plate) + (1|sample)
fm2 <- lmer(f, Penicillin)
```

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>plate</td>
<td>(Intercept)</td>
<td>0.71691</td>
<td>0.84671</td>
</tr>
<tr>
<td>sample</td>
<td>(Intercept)</td>
<td>3.73097</td>
<td>1.93157</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>0.30241</td>
<td>0.54992</td>
</tr>
</tbody>
</table>

Number of obs: 144, groups: plate, 24; sample, 6

Fixed effects:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>22.9722</td>
<td>0.8086</td>
<td>28.41</td>
</tr>
</tbody>
</table>

The sample-to-sample variability has the greatest contribution, then plate-to-plate variability and finally the residual variability that cannot be attributed to either the sample or the plate.
Standard mixed models, such as those in lme4, assumes Gaussian distributions for all components.

The model parameters we need to estimate are:
- fixed effects, $\beta$
- parameters in the covariance matrix of the random effects, $\theta$
- measurement noise variance $\sigma^2$.

The parameters are estimated by numerical maximization of the likelihood for the data observed $y$, $\mathcal{L}(\theta, \beta, \sigma | y)$.

Be default, lme4 does not do ML estimation but Restricted (or residual) ML (REML) estimation.
Restricted ML estimation

REML estimation is done as follows:

- For a fixed value of $\theta$, we can maximize the profiled likelihood $L(\theta, \beta, \sigma | y)$ analytically with respect to $\beta$ and $\sigma$ to obtain profile likelihood estimates $\hat{\beta}_\theta$ and $\hat{\sigma}_\theta$.
- $L_P(\theta | y) = L(\theta, \hat{\beta}_\theta, \hat{\sigma}_\theta | y)$ is called a profiled likelihood.
- $L_P(\theta | y)$ is maximized numerically to obtain the parameter estimates (e.g. using the L-BFGS method).

There are good reasons for doing this:

- It decreases computation time.
- ML estimates of variances are known to be biased. REML estimates tend to have less bias.
- Turn off this by using REML = FALSE in the lmer function.
lme4 uses confidence intervals based on the profile likelihood as follows:

- Let $\mathcal{L} = \arg \max_\theta \mathcal{L}(\theta | y)$.
- Let $\mathcal{L}_p$ denote the best fit with the parameter of interest fixed at value $p$.
- Define the likelihood ratio test statistic on deviance scale: $D(p) = -2(\log(\mathcal{L}_p) - \log(\mathcal{L}))$.
- Define $\xi(p)$ as the signed square root transformation of $D(p)$.
- $\xi(p)$ is compared to quantiles of a standard normal, and for example a 95% profile deviance confidence interval of $p$ is given by the values for which $-1.960 < \xi < 1.960$. 
Confidence intervals in R

```r
fm1 <- lmer(Reaction ~ Days + (Days | Subject), sleepstudy)
pr1 <- profile(fm1)
confint(pr1)

2.5 %   97.5 %
.sig01  14.3815048  37.715996
.sig02  -0.4815007   0.684986
.sig03   3.8011641   8.753383
.sigma  22.8982669  28.857997
(Intercept) 237.6806955 265.129515
Days       7.3586533  13.575919
```
Profile zeta plots: `xyplot(pr1)`
Interpreting the profile zeta plot

- One could think of these curves as showing the distribution of the estimator.
- If the curve is close to a straight line, the likelihood is quadratic close to the maxima and a Gaussian approximation works fine for inference.
- The curve for the fixed effects are sigmoidal, which means that the distribution is over-dispersed compared to a Gaussian (for regression it is a t-distribution).
- The curves for the variances are skewed (for regression it is a $\chi^2$ distribution).
Assessing the random effects

- It is often interesting to look at the posterior distribution \( \pi(b|y) \) of the random effects.
- For example, what is the effect of the sleep deprivation for a specific subject?
- The posterior is Gaussian, and we can use the posterior mean as a point estimate.
- The posterior mean and variance for each subject is computed using the `ranef` function.
- These can be plotted using the `dotplot` function:

```r
rr1 <- ranef(fm1, condVar = TRUE)
dotplot(rr1, scales = list(x = list(relation = 'free')))
```
• In many situations, one have so called nested factors that need to be dealt with in a slightly different way.

• To illustrate this, we use data from Davies and Goldsmith (1972) of deliveries of a chemical paste product contained in casks.
  • As a routine, three casks selected at random from each delivery were sampled.
  • Ten of the delivery batches were sampled at random and two analytical tests carried out on each of the 30 samples.

• The factors are now batch (1-10) and cask (1-3). Thus, it would be tempting to fit a model

\[ \text{strength} \sim 1 + (1 \mid \text{cask}) + (1 \mid \text{batch}) \]

to account for the cask effect and batch effect. This is wrong!
Nested factors

• The problem is that the casks are sampled at random, so “cask 1” in batch 1 has nothing to do with “cask 1” in batch 2.

• Thus, the effect of cask only makes sense within each batch: The factors are nested.

• The model we need to specify is

\[ Y_{ijk} = \alpha + b_i + b_{ij} + \varepsilon_{ijk} \]

where \( b_i \) is batch effect, \( b_{ij} \) is cask effect, and \( \varepsilon_{ijk} \) is measurement error.

• Note that this model requires repeated measurements for each cask (we have 2).
To specify a model with nested factors, we can use

\[ \text{strength} \sim 1 + (1 \mid \text{batch}) + (1 \mid \text{batch}:\text{cask}) \]

which also can be written as

\[ \text{strength} \sim 1 + (1 \mid \text{batch}/\text{cask}) \]

Alternatively, we can pre-compute the nested factor \( b_{ij} \) and append it to the dataset, as

\[ \text{Pastes} <- \text{within(Pastes, sample} < - \text{factor(batch:cask))} \]

and then specify the model as

\[ \text{strength} \sim 1 + (1 \mid \text{batch}) + (1 \mid \text{sample}) \]
f = strength ~ 1 + (1|batch) + (1|batch:cask)
fm3 <- lmer(f, Pastes, REML=FALSE)
fm3

Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: strength ~ 1 + (1 | batch) + (1 | batch:cask)
Data: Pastes

AIC BIC logLik deviance df.resid
255.9945 264.3718 -123.9972 247.9945 56

Random effects:
Groups Name Std.Dev.
batch:cask (Intercept) 2.9041
batch (Intercept) 1.0951
Residual 0.8234

Number of obs: 60, groups: batch:cask, 30; batch, 10

Fixed Effects:
(Intercept) 60.05
A reduced model

- The batch effect is much lower than the sample effect. The profile confidence interval for $\sigma^2$ goes to zero:

  pr3 <- profile(fm3)
  confint(pr3)

  2.5 %  97.5 %
  .sig01  2.1579337  4.053589
  .sig02  0.0000000  2.946591
  .sigma  0.6520234  1.085448
  (Intercept)  58.6636504  61.443016

- Can we get away with a simpler model without the batch effect?

  f = strength ~ 1 + (1|batch:cask)
  fm4 <- lmer(f, Pastes, REML=FALSE)
We want to test $H_0 : \sigma^2 = 0$ against $H_1 : \sigma^2 > 0$.

Alternatively, we can say that we want to test the full model against the reduced model.

We can compare nested models using a likelihood ratio test, using that the LRT statistics is approximately $\chi^2$ distributed.

This test is a classical ANOVA analysis, implemented in the `anova` function of `lme4`.

```r
anova(fm3,fm4)
```

Data: Pastes

Models:

- `fm4`: $\text{strength} \sim 1 + (1 \mid \text{batch:cask})$
- `fm3`: $\text{strength} \sim 1 + (1 \mid \text{batch}) + (1 \mid \text{batch:cask})$

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>Chisq</th>
<th>Chi Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fm4</td>
<td>3</td>
<td>254.40</td>
<td>260.69</td>
<td>-124.2</td>
<td>248.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fm3</td>
<td>4</td>
<td>255.99</td>
<td>264.37</td>
<td>-124.0</td>
<td>247.99</td>
<td>0.4072</td>
<td>1</td>
<td>0.5234</td>
</tr>
</tbody>
</table>

David Bolin
More precise model checks

- One should be a bit careful with LR tests, since they are based on asymptotics.
- The absence of analytical results for null distributions of parameter estimates in complex situations is a long-standing problem in mixed-model inference.
- If a greatly increased computation time is acceptable, one can estimate confidence intervals by parametric bootstrapping.
- This is done by simulation from the model and refitting of the model parameters.
- The function `bootMer` implements this.
As always, we should check that the assumptions of the model hold before using it to draw conclusions.

Some common diagnostic plots are:

```r
plot(fm1, type = c("p", "smooth"))  # residual plot
qqmath(fm1, id = 0.05)  # qq-plot
```
• Longitudinal data analysis.
  • The regression example was the simplest example of a 
    longitudinal model.
  • We often need to deal with temporal correlation using 
    stochastic processes.
• Generalized mixed models.
  • Need if we, for example, want to look at count data.
• Non-Gaussian mixed models.
  • What do we do if we have skewed random effect distributions?
• Bayesian mixed models.
  • In Bayesian models, there is no need to separate random 
    effects from fixed effects, they just have different priors!
  • A popular package for fitting Bayesian mixed models is 
    R-INLA.

+++