Mixed Models Theory

Bruna Wundervald

August, 2020

Contents

1 Linear Mixed Models

1.1 Model Statement

A LMM for a set of observations $y = (y_1, \ldots, y_n)$ has the general form

$$Y|b \sim N(\mu, \Sigma), \quad \mu = X\beta + Zb, \quad b \sim N(0, \Sigma_b),$$

where X and Z are the $p \times n$ predictor matrices, and $\Sigma = \sigma^2 I$ usually. An example for clustered data is:

$$Y_{ij} \sim N(\mu_{ij}, \Sigma), \quad \mu_{ij} = x_{ij}^T \beta + z_{ij}^T b, \quad b_i \sim N(0, \Sigma_b^*),$$

where x_{ij} now contains the predictor values for the j-th observation in the $i-$th cluster, and z_{ij} is the sub-vector of x_{ij} that exhibits extra between cluster variation in its relationship to Y.

1.1.1 Example from the `lme4` paper

Let us consider now the data from a sleep deprivation study, from the `lme4` package paper. On day 0 the subjects had their normal amount of sleep, and were from that night restricted to 3 hours of sleep per night. The response variable represents the average reaction times in milliseconds (ms) on a series of tests done each day for each subject. In the following figure, we can see a general trend of the reaction time increasing with the passage of days, and the reaction time itself varies quite a lot between the subjects, in both slope (starting reaction time) and intercepts (effect of days in the reaction time). This type of study justifies the use of a LMM, as we can clear see that there is a difference in the response between individuals and days.

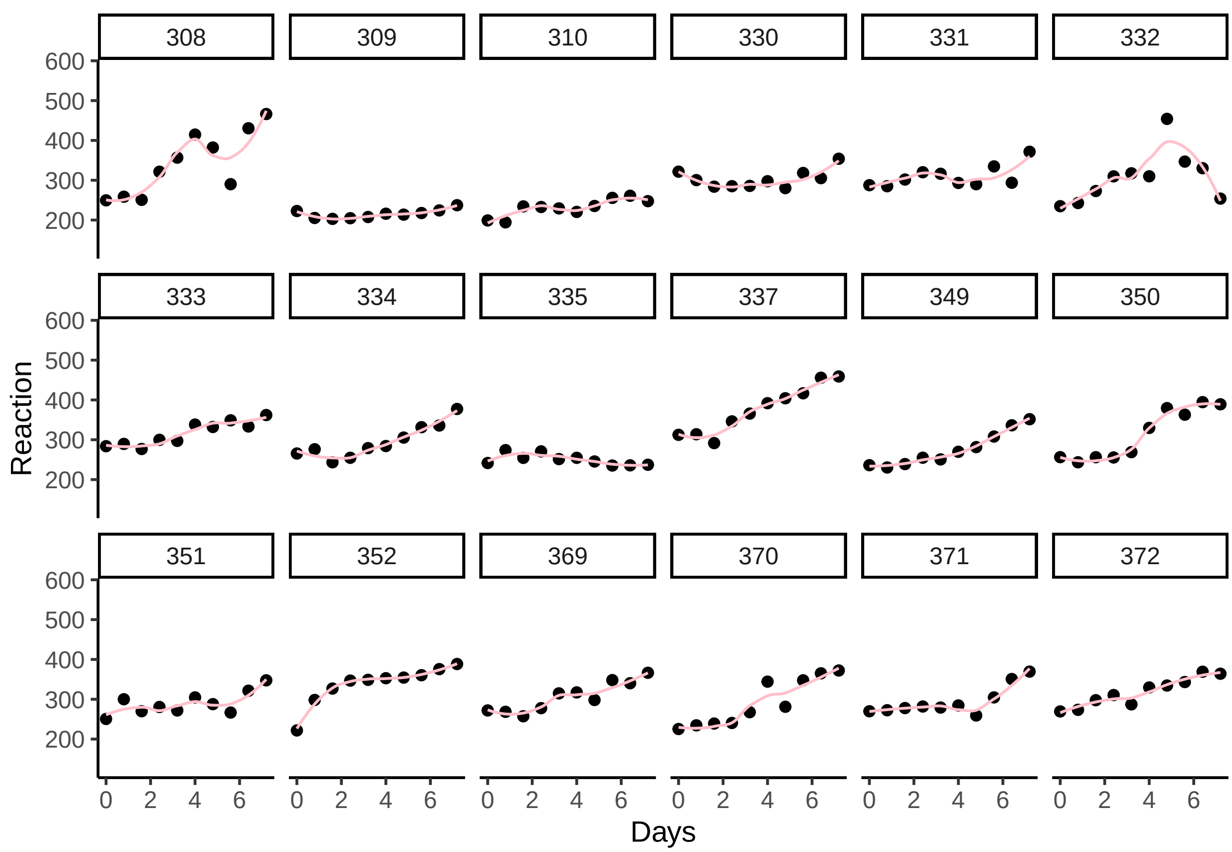

Figure 1: Reaction time per days and subjects in the sleep deprivation study.

The following code starts by creating an LMM model using the subject ID as the random effect, for which the standard error is estimated as 37.12 (26.01, 52.94). The second model now adds the Days variable a a random effect slope, meaning the random intercept and slope are correlated (-0.48, 0.68). In comparison to the previous model, the ICs for the fixed effects got smaller, given that now we have less uncertainty about the population average.

```r
lmm <- lmer(Reaction ~ Days + (1 | Subject), sleepstudy)
summary(lmm)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (1 | Subject)
##    Data: sleepstudy
##
## REML criterion at convergence: 1786.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2257 -0.5529  0.0109  0.5188  4.2506
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept) 1378.2   37.12
##  Residual              960.5   30.99
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 251.4051     9.7467   25.79
## Days         10.4673     0.8042   13.02
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.371
```

```r
round(confint(lmm, oldNames = FALSE), 2)
```

```
##                          2.5 % 97.5 %
## sd_(Intercept)|Subject  26.01  52.94
## sigma                   27.81  34.59
## (Intercept)            231.99 270.82
## Days                     8.89  12.05
```

```r
lmm_days <- lmer(Reaction ~ Days + (Days | Subject), sleepstudy)
summary(lmm_days)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (Days | Subject)
##    Data: sleepstudy
##
## REML criterion at convergence: 1743.6
##
## Scaled residuals:
```

```
##      Min       1Q  Median      3Q     Max
## -3.9536 -0.4634  0.0231  0.4633  5.1793
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 611.90   24.737
##           Days         35.08    5.923   0.07
##  Residual             654.94   25.592
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  251.405      6.824  36.843
## Days          10.467      1.546   6.771
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.138
```

```
round(confint(lmm_days, oldNames = FALSE), 2)
```

```
##                               2.5 % 97.5 %
## sd_(Intercept)|Subject        14.38  37.71
## cor_Days.(Intercept)|Subject  -0.48   0.68
## sd_Days|Subject                3.80   8.75
## sigma                         22.90  28.86
## (Intercept)                  237.68 265.13
## Days                           7.36  13.58
```

1.2 Model Fitting

2 Bayesian Linear Mixed Models

A Bayesian version of the LMM is given by the use of prior distributions for β, Σ and Σ_b. In the Bayesian case, the distiction between random and fixed effects is less clear, as both β and b will have probability distributions $f(\beta)$ and $f(b) = \int f(b|\Sigma_b)f(\Sigma_b)d\Sigma_b$. The most common way of fitting the BLMM is based on a Gibbs sampler.