

- 1 Model selection
- 2 Contrasts and plots of partial fixed effects
- 3 Shrinkage and partial pooling

Advanced LMM Lectures and Exercises

Reinhold Kliegl

2017-08-27 21:22:26

The three lectures and associated exercises on advanced issues of linear mixed models (LMMs) cover (1) model selection, (2) specification and plotting of (partial) fixed effects, and (3) the heuristic value of variance components and correlation parameters as well as their visualization. We will not spend (much) time on statistical theoretical issues, but rather focus on practical issues. Therefore, the lectures draw on publications for which data and R scripts are available on the *Potsdam Mind Research Repository*, **PMR2** (<http://read.psych.uni-potsdam.de/pmr2/>), or on the *Mind Research Repository*, **MRR** (<http://openscience.uni-leipzig.de/index.php/mr2>).

Workshop participants may want to use this material or other articles at the repository for their exercises. We think, however, that benefit is largest if they use their own appropriately complex data.

In this case, we suggest that they upload their data as a zipped archive to the email address `smhp dot potsdam AT gmail dot com`, ideally along with an `Rmd` script that describes the data and potential problems with an LMM analysis. **If you decide to upload your data, please put as subject line: Data from FirstName LastName.**

Obviously, the data can be used as a starting point for the LMM-related analyses in the exercises. Also, if data and script are uploaded, some instructor or tutor may find time to look at the issues. No promises.

1 Model selection

The goal is to learn to determine a parsimonious random-effect structure of an LMM. There are three parts.

1.1 Flexible specification of variance components and correlation parameters

Model selection involves selectively dropping or adding variance components or correlation parameters. There are two useful techniques to achieve this:

- Conversion of factor-based contrasts to numeric covariates; they can be extracted from the model matrix.

- Use of the double-bar syntax to estimate zero-correlation parameter models

We will illustrate these techniques and why they are necessary with script `lme4_doublebar.Rmd`, also available as RPub (<http://www.rpubs.com/Reinhold/22193>).

1.2 Determining parsimonious mixed models

Advance notice: We do not think that there is a single algorithmic way to arrive at parsimonious mixed models. Rather we present/discuss strategies that appear to work according to our experience. It is possible that this strategy may miss a parsimonious LMM. Basically, we illustrate a three-step procedure:

- Is the model overparameterized/degenerate? The main tool we use is the `rePCA()` function in the `RePsychLing` package, available here (<https://github.com/dmbates/RePsychLing>).
- Once an LMM is identified (is not overparameterized) we use likelihood-ratio tests (along with AIC, BIC) to test whether variance components can be eliminated from the LMM without loss of goodness of fit. We always compare nested models.
- For a set of variance components supported by the data, we check whether an LMM extension with correlation parameters increases the goodness of fit.

The primary reference for this section is Bates et al. (2015) (<https://arxiv.org/abs/1506.04967>). We illustrates this part with an updated vignette the dataframes `KWDYZ` and `KKL` available in the `RePsychLing` (<https://github.com/dmbates/RePsychLing>) package. Of course, this part of the lecture touches on the controversy triggered by Barr et al. (2016) (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3881361/>); see also Matuschek et al. (2017) (<http://www.sciencedirect.com/science/article/pii/S0749596X17300013>) and Baayen et al. (2017) (arXiv:1511.03120v2).

1.3 Pruning of fixed-effect terms

Here we distinguish between LMMs of factorial designs and LMMs using a mixture of factors and potentially correlated numeric covariates.

The default for factorial designs is to leave all interaction terms in the LMM; it usually does not make a difference; see, for example, Masson and Kliegl (2013) (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=90:masson-a-kliegl-2013-jeplmc-modulation-of-additive-and-interactive-effects-in-lexical-decision-by-trial-history&catid=11:publications&Itemid=13) or Masson et al. (2017) (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=164:modulation-of-additive-and-interactive-effects-by-trial-history-revisited&catid=11:publications&Itemid=13).

The strategy for pruning designs with (highly) correlated numeric covariates is less straightforward and should be guided by theoretical expectations. It involves decisions about the degree of polynomials of covariates and the degree of interactions to be kept in the model. We illustrate this with an analyses of eye-movements during reading (Hohenstein, Matuschek, & Kliegl, 2017) (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=165:hohenstein-matuschek-a-kliegl-2016-pbr&catid=12:publications&Itemid=11)

1.4 Exercise

The exercise following this lecture will focus on Parts 1 and 2 of this lecture. We will provide data sets for participants, such as as the ones linked above, to try their hands on determining an appropriate LMM. Alternatively, participants are strongly encouraged to apply the techniques to a data set of sufficient complexity of their own.

2 Contrasts and plots of partial fixed effects

This lecture focuses on *fixed effects*.

2.1 Contrasts

LMMs require knowledge about the specification of *a priori* (ideally) planned comparisons. This is most flexibly accomplished with contrasts. We will illustrate these techniques with script

`Contrasts.RK.Rmd` (requires function `mixedDesign.R`); see also material available at PMR2 (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=167:contrast-coding&catid=13:r-playground&Itemid=15).

2.2 Partial effects

Contrasts are usually based on factors. In complex LMMs, these contrasts interact with numeric covariates. Numeric covariates are usually correlated. In this case, partial effects of numeric covariates and their interactions with contrasts or other numeric covariates may look very different from their zero-order relation, indeed, in the case of suppressor constellations, the direction of the effect may be different for partial effect and zero-order effect.

We will illustrate such phenomena with eye-movements in reading such as reported by Yan et al. (2013) (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=132:yan-et-al-2014-cognition-eye-movements-guided-by-morphological-structure-evidence-from-the-uighur-language&catid=12:publications&Itemid=11).

Useful partial-effect plots are also reported in Hohenstein and Kliegl (2014) (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=110:hohenstein-kiegl-2014-jeplmc-semantic-preview-benefit-during-reading&catid=10:publications&Itemid=14), Hohenstein et al. (2017) (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=165:hohenstein-matuschek-a-kiegl-2016-pbr&catid=12:publications&Itemid=11), and Laubrock and Kliegl (2015) (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4585246/>).

Plots of partial effects and zero-order relations can be accomplished with the `remef` package available here on `github` (<https://github.com/hohenstein/remef>); the packages is also available on CRAN .

2.3 Exercise

The exercise covers contrast specifications and partial effect plotting.

For plotting of partial effects and zero-order relations we use the `remef` (<https://github.com/hohenstein/remef>) package.

3 Shrinkage and partial pooling

This lecture focuses on *random effects*, that is on variance components and correlation parameters as well as shrinkage and partial pooling in LMMs. The lecture covers two examples of the heuristic value of these estimates. For an introduction see, for example, Kliegl et al. (2010) (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=5:kliegl-et-al-2010-a-linear-mixed-model-analysis-of-masked-repetition-priming&catid=11:publications&Itemid=13).

3.1 From variance components to fixed effects

An important goal of a research program should be to convert statistically reliable variance components (i.e., between-subject or between-item effects) into fixed effects. An illustration of this technique is available in script `.Rmd`, based on Pan et al. (2012) (<http://openscience.uni-leipzig.de/index.php/mr2/article/view/92>).

3.2 The meaning of correlation parameters

Many users are puzzled by the meaning of correlation parameters. One way to think about them is that they reflect interactions between a post-hoc grouping factor of subjects based on the dependent variable (e.g., slow vs. fast subjects split on median RT or LMM intercept) and an experimental within-subject effect. An illustration of such a case was reported in Kliegl et al. (2011) (http://read.psych.uni-potsdam.de/pmr2/index.php?option=com_content&view=article&id=10:kliegl-type&catid=11:publications&Itemid=13). (Note these are the `KWDYZ` data used in Lecture 1.)

3.3 Exercise

There are (at least) three options. One is to continue the second exercise, because an understanding and application of contrast specification and generation of partial-effects plots likely requires more than 1.5 hours.

The other option is to focus on the content of the third lecture. There is an excellent script by Tristan Mahr at his website (<https://tjmahr.github.io/plotting-partial-pooling-in-mixed-effects-models/>), available also as `Mahr.PartialPooling.Rmd`. We will adapt this script for our own data.

Finally, there is also the option to enhance one's understanding of correlation parameters and how their estimation depends on number of subjects, number of observations per condition per subject, and standard deviations of effect sizes with a shiny shrinkage app (https://pmr2.shinyapps.io/Shrinkage/Shiny_Shrinkage.Rmd).

