Keep it maximal, be parsimonious: uncertainty dominates the selection of random effects structure in linear mixed-effects models

Giulia Calignano

November 27, 2020

Research fellow @DPSS, Univeristy of Padova

Outline

- Bibliographic resources
- A step back... why LMEMs fit better than ANOVAs?
- Random effects structure
- Maximal vs Parsimonious modeling
- Hands on eye-tracking data
- Take home message

Bibliographic resources

My (limited) readings

- Random effects structure for confirmatory hypothesis testing: Keep it maximal - **Barr et al., 2013**, *J. Mem. Lang.*
- Parsimonious mixed-models - **Bates et al., 2015**, *arXiv*
- Balancing Type I error and power in linear mixed models - **Matuschek et al., 2017**, *J. Mem. Lang.*
- Linear Mixed-Effects Models and the Analysis of Nonindependent Data: A Unified Framework to Analyze Categorical and Continuous Independent Variables that Vary Within-Subjects and/or Within-Items - **Brauer Curtin, 2018**, *Psychol. Method*

A step back... why LMEMs fit better than ANOVAs?

ANOVAs' limits and LMEMs' advantages

- repeated measures are non-independent i.e. clusters. A **violation of the independence assumption** is considered the most serious one for incorrect inferential statistics **using ANOVA**
- **a single LMEMs can replace two separate ANOVAs.**
- **LMEMs better preserves statistical power by dealing with missing data and "unbalanced repeats"** (e.g., different subjects).
- **LMEMs are suited to analyze change over time (or space)** higher-order clustered.
- It is a small step from LMEMs to "generalized linear-mixed effects models,"

Random effects structure in mixed-effects models

Random effects structure (technically speaking)

Fixed effects

- tested against an error term that captures the variability of the effect across individuals.

$$Y_{si} = \beta_0 + \beta_1 X_i + \epsilon_{si},$$

$$\epsilon_{si} \sim N(0, \sigma^2)(1)$$

Random effects

- variance components for each within-subject and -item effect and interaction term (i.e., random slopes)

$$Y_{SI} = \beta_0 + S_{0s} + I_{0i} + (\beta_1 + S_{1s})X_i + \epsilon_{si},$$

$$(\mathsf{S}_{0s}, S_{1s}) \sim N(0, \begin{bmatrix} \tau_{11}^2 & \rho\tau_{00}\tau_{11} \\ \rho\tau_{00}\tau_{11} & \tau_1^2 1 \end{bmatrix}) \qquad I_{0i} \sim N(0, \omega_{00}^2)$$

$$\epsilon_{si} \sim N(0, \sigma^2)(2)$$

Random effects structure (definition for dummies [like me])

Fixed effects

- data are gathered from all the levels of the variable. The values of a fixed variable is the same across studies. **The variables implicated by theoretical predictions tend to be fixed**.

Random effects

- the researchers' interest is in all possible levels, but only a random sample of levels is included in the data.

The major difference between independent data and nonindependent data is the complexity of the error term.

Maximal vs Parsimonious modeling

Maximal modeling' advantages

- If an effect exists, subjects and/or items will vary in the extent to which they show that effect, whether or not that variation is detectable.
- Overfitting the data with a maximal model has only minimal consequences for Type I error and power
- Underfitting the design can incur levels of anticonservativity from minor (best-path model selection) to extremely severe (random-intercepts-only LMEMs) with little real benefit to power.
- random slopes for "critical predictors" is necessary to keep the type-I error rate

A maximal model with no random correlations or even missing within-unit random intercepts is preferable to one missing the critical random slopes.

Parsimonious modeling's advantages

- The complexity of the model increases but the optimization decreases. That is, over-specified random effects

- Non convergence or parameter estimates with degenerate or singular covariance matrices (no variability). (1) The number of parameter is not realistic, estimate of variance and covariance parameters for each random factor

$$\frac{(\prod withinlevels) * (\prod withinlevels + 1)}{2} \tag{3}$$

- *[...]How representative are simulations studies with respect to real data sets? Large correlation parameters are indicative of overparameterization. They do not represent true correlations in the population.[...]* Bates et al., 2015.

And so? ...well, great powers derive from great responsibilities

Bates et al., 2015's proposal:

1. Start with a maximal model;
2. Fit a zero-correlation model (PCA);
3. Remove variance components until the likelihood ratio test or Bayesian hierarchical modeling show no improvement;
4. Add the correlation parameters for the remaining variance components.

And so? ...well, great powers derive from great responsibilities

Matusheck et al.,2017's proposal:

1. The maximal model produce substantial loss of power, if the true value of the variance component is small.
   The best model is neither the maximal nor the minimal
2. The inclusion of random slopes that have a near-zero variance may lead to an unnecessary decrease of power. **I want to know which approach will maximize power!**
3. For small sample sizes, the model selected by the AIC appears anti-conservative whereas **LRT with backward selection heuristic, show stable Type I error rate and benefit in power compared to the maximal model**

And so? ...well, great powers derive from great responsibilities

Brauer and Curtin,2018's proposal:

1. Determine the **maximal random effects structure** while get the LMEM **to converge**
2. What is the next step? **"keep-it-maximal approach" or a "model selection approach"?**
3. Most experts agree that **the final LMEM needs to contain the random slope(s) for the main predictor**, regardless of its/them variance.
4. the presence of **variance components with a near-zero variance in the random effects structure does not affect the goodness-of-fit of the model.**

Hands on eye-tracking data

Saccade Latency

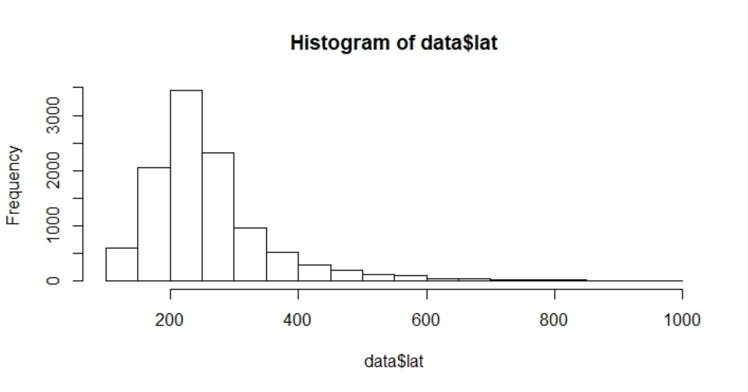


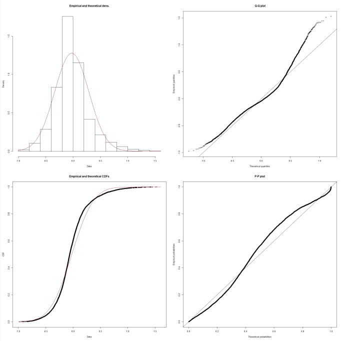
Figure 1: Hist graph for saccade latency frequency

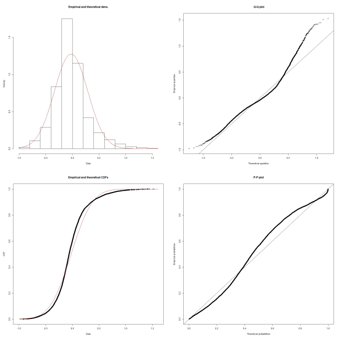
maximal model lat = block * condition +(block*condition—id/item)

mixed between- and within- id design, N = 110 x 10839 observation

Fitted residuals of

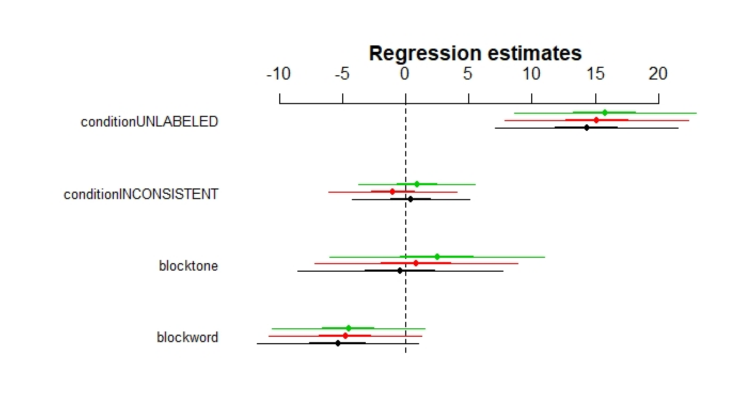
(converged) maximal model vs best selected model





Coefficent' comparison among models

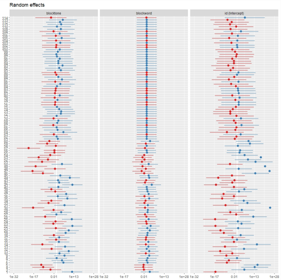
(converged) maximal model vs best selected model vs no-slope model

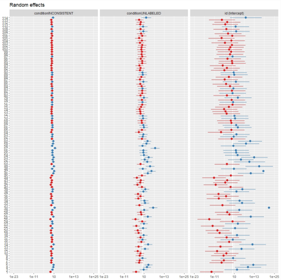


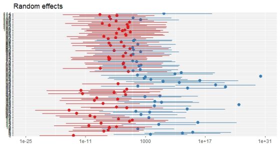
Hands on eye-tracking data

RE plot id

(converged) maximal model vs best selected model vs no-slope model







Take Home message and resources

Take home messages

- An important goal in statistical analysis is the avoidance of overfitting. Any given data-set can tolerate only a limited number of parameters.
- In general, model selection tries to balance the goodness-of-fit of a model with its risk for overfitting the data.
- The effects of interest *can* or *do* exist?
 ...uncertainty dominates

extra-resources

Nice functions summarizing and visualizing regression models, by Jacob Long

calignano.giulia@gmail.com

Thank you,

comments and suggestions are welcome