

Multivariate: Linear Mixed Models

Lecture 3: Building & Evaluating Models

Tom Booth

Lecture 3

Good morning

- Dates:

- Set: 29/03 (when LMM is finished)
- Due: 02/05 (5 weeks to complete)
- Marks returned: 24/05 (3 weeks for us to mark)

- Task:

- We will give you some data, and a set of questions requiring short answers.
- You will have one set of questions on SEM and one set on LMM.
- You will submit a document with your answers, and the R-code you used to produce them.

Coursework Announcements

- Task cont.:
 - A core of the questions will be answerable quite directly from course material
 - Some will be a little harder
- Code:
 - We will run your code files.
 - If they fail to run, or do not produce the results in your report you will lose 10%.
 - We will include some guidance on (a) reading in data, and (b) checking your code runs for you to use prior to submitting.
- Format:
 - Page limit is still to be confirmed (so the following are indicative)
 - Approximately 2-4 pages of text, and up to 6 pages of additional tables etc.

From last week...

- We took a more detailed look at estimated LMM with `lmer()`
 - We looked at model code
 - Linking code to equations
 - Basic interpretation
 - Looked at extracting core elements of the results
 - And used prediction equations when we have fixed and random effects to help delve a little further into our models.
- We also discussed ICC and shrinkage

Any questions from last week?

Today

- 1 Model Specification
 - 2 Model Estimation
 - 3 Model Evaluation
 - 4 Model Building
 - 5 Model Comparison
 - 6 Model Assumptions
- Then we will take a look at a new example

Model Specification

This will be short. . . .

- Estimate the model that contains parameters that best represent the effects in your theory.
 - This is about it!
- The in's and out's of how to represent this in R-code we have been discussing, and will continue to.
- Whether the data is good enough for the question is a matter of study design
 - ***This is ALWAYS the case and is often forgotten***
 - Why is this a section?
 - To flag that model specification should be part of the *a priori* analysis plan.

What parameters can I specify?

- Individual level fixed effects:
 - Linear and non-linear effects
 - Interactions
 - Contrasts for categorical variables
- Group level fixed effects:
 - Effect of variables measured at the group level on the outcome.

What parameters can I specify?

- Random effects by grouping variable:
 - Intercept: different outcome average values per group
 - Slopes: different effects of a predictor on the outcome by group
- Random effect covariances
- Cross-level interactions:
 - Different effects of a predictor on an outcome by a level two variable
 - E.g. Intervention group changing rate of change over time.

Model Estimation

- When we specify a model, we have a number of parameters, that have unknown values, we want estimate.
- In the context of LMM these are:
 - All fixed effects
 - Means of random effects
 - Variances of random effects
 - Covariances of random effects
- The process of calculating the best values for these coefficients is broadly referred to as model estimation.

- The values of the residuals for intercepts and slopes for each group (i.e. how much the overall slope and intercept needs to be adjusted for a specific group) are not part of the model estimation (Baayen et al. 2008).
 - These are calculated once the random-effects parameters have been estimated
 - They are referred to as **B**est **L**inear **U**nbiased **P**redictors (BLUPS) (Baayen et al. 2008).

- When we run a standard linear model (regression or ANOVA), we can calculate the parameters using a *closed form solution*
 - In other words, there is one best set of values for the parameters (β 's)
 - We can calculate these algebraically
 - And in this case we do so by minimizing the sum of the squared residuals
- Why not use OLS in LMM?
 - In short, the model is too complicated, and a closed form solution does not exist.
 - In LMM, we *estimate* all the parameters using one of a number of iterative procedures.
- There are two common estimators in LMM, maximum likelihood and restricted maximum likelihood.

Maximum Likelihood

- Maximum likelihood (ML) is an estimation method in which we seek to find the values for the unknown parameters that maximize the likelihood of obtaining the observed data.
- This is done via finding values for the parameters that maximize the (log) likelihood function.

ML: the broad idea

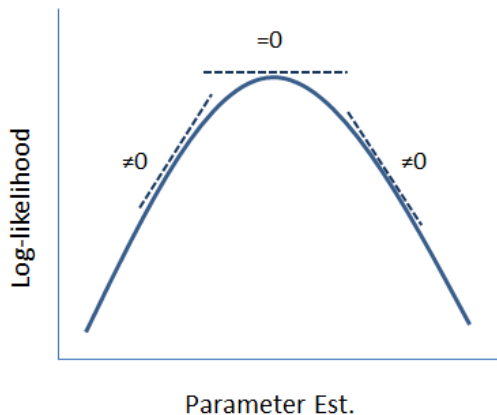


Figure 1:

ML: More than one parameter

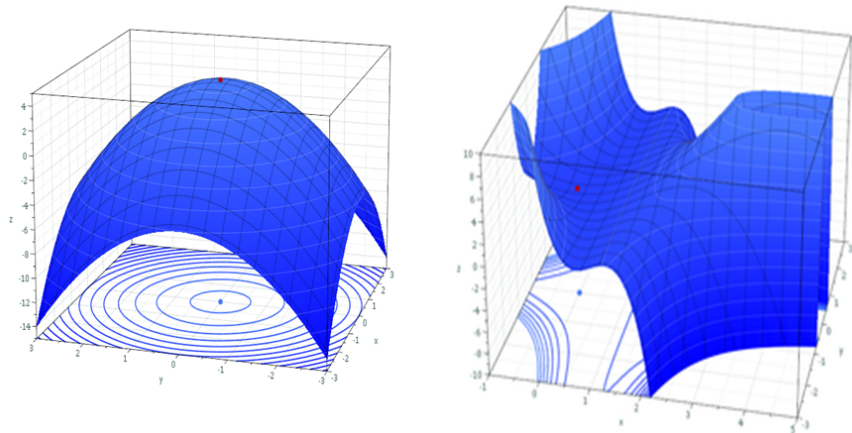


Figure 2:

(log)Likelihood

- Our data is comprised of multiple observations.
- So we can think about selecting parameters that maximize the likelihood of each observation.
- The total sample likelihood is then the combination of these.
- This is where the log-likelihood comes in.
 - From probability, we can combine *i.i.d* by multiplication.
 - This can get a bit mathematically involved.
- In part to overcome this, we use and discuss the natural log of the likelihoods, or the log-likelihood.
 - The primary advantage of the log-likelihood is that instead of taking the product of the individual likelihoods, we can simply sum them.
 - This is because multiplication is addition on a log scale.

- In the context of LMM, the multivariate space for ML includes both the fixed effects and variance components in the calculation of the likelihood.
- The issue with ML in the context of LMM is how it treats the fixed effects when estimating variance components.
 - Short version: ML treats fixed effects as unknown values when it estimates the variance components, but it does not adjust degrees of freedom.
 - This results in biased estimates of the variance components.
 - The bias makes ML estimates of variance components too small.

Restricted maximum likelihood

- An alternative approach in LMM is what is referred to as restricted maximum likelihood (REML).
 - REML is the default estimator in `lmer()`
- REML includes only the variance components in the likelihood, and estimates the fixed effects in a second step.
 - It assumes values for fixed effects are known, and as such no *df* adjustment is needed.
 - This in effect partials the fixed effects out prior to calculating random effects.
- Although REML estimates are not guaranteed to be unbiased, they are usually less biased than ML estimates of the variance components.

- Whether one applies ML or REML, the result is:
 - A value for the (log)likelihood
 - Estimates of the fixed effects parameters (BLUEs)
 - Estimates of the variance components
 - SE estimates for the parameters
 - Model *deviance* which we will discuss later in the context of model comparison.

When should I use ML vs REML?

- As per the use as the default estimator, REML is often our preferred estimator.
- If we want to focus on fixed effects, ML may be preferable.
- If our focus is on random effects, REML may be preferable.
- Given what parameters are included in the likelihood, ML and REML can be used for different sets of model comparisons.

Some comments on estimation

- Estimation methods employ iterative algorithms to search for the best values for parameters.
- As an example, in the case of ML, these work roughly as follows:
 - 1 Begin with some start values (initial guess) for parameter values.
 - 2 Compute the log-likelihood (these two steps are referred to as an iteration)
 - 3 Adjust the guess at parameter values.
 - 4 Recompute the likelihood.
 - 5 Repeat until improvements in likelihood value are negligible across iterations.

- ① Start values
- ② Local minima and maxima: iterations get “trapped” around these points we will not find our MLE
 - Start values can, in principle, be used to place our initial guess in a location where we may have more success of locating the MLE.
 - But this is a bit of a trial and error process.
- ③ Convergence = the point at which the accuracy of the likelihood estimate is within your given tolerance level.
 - I.e. that the difference in likelihood across iterations is very small.
 - MLE has a number of parts; the data, the model and the algorithm used.
 - Lack of convergence could be a result of any or all of these.
 - No MLE

What to do?? (be pragmatic)

- It is **VERY** difficult to make generalized statements on how to deal with such problems.
- Some solutions:
 - Locate and remove problematic variables
 - Locate and rescale/transform problematic variables
 - Simplify the model so you have less parameters to estimate
 - Think about starting with a simple model and building up (see later slides)
- Often the issue is the data!
 - Solution: design a better study to collect more appropriate data!
- Think logically - work step by step!

Model Evaluation & Testing Effects

Our normal approach

- When we run a standard linear regression, we are used to:
 - Evaluating individual coefficients based off of p -values or confidence intervals
 - Looking at the overall model via F -tests
 - Looking at the coefficient of determination (R^2)
 - Then making some interpretation of effects
- In LMM, pretty much all of these steps (except the last) has some complications!

- R^2 in linear models provides an estimate of the variance explained by the predictors.
 - This is calculated based on total and residual variances.
 - The F -test is used to test if this explained variance significantly differs from 0.
- The issue with these metrics in the context of LMM is that we have multiple variance components, not just ϵ (residuals)

- To determine the variance explained by a model, two R^2 statistics have been introduced for (G)LMMs: marginal and conditional R^2 .
 - Marginal R^2 gauges variance explained by fixed effects
 - Conditional R^2 is concerned with variance explained by both fixed and random effects
- Nakagawa & Schielzeth (2013) provide a definition of these measures for LMM and GLMM (generalized LMM) that incorporate random intercepts only.
- Johnson (2014) provides an extension for random slopes models, available through the `r.squaredGLMM` function in the MuMIn package

Proportional Reduction in Variance (PRV)

- An alternative not directly equivalent to R^2 is the PRV.

$$PRV = \frac{var_{M0} - var_{M1}}{var_{M0}}$$

- Where var_{M0} and var_{M1} are level 1 (ϵ_{ij}) or level 2 (v) variances, and subscripts 0 and 1 refer to models with and without an estimate being tested.

p -values Thorny issue!

- `lmer` output provides t -values ($\frac{\beta}{SE}$) but it does not provide p -values
- The issue is that the null distributions for t and for F in the case of model comparisons, are only asymptotically normal.
- They do not follow t and F distributions in finite samples.
 - This is largely due to calculation of appropriate degrees of freedom.
 - Or to put this another way, we do not have an appropriately sampling distribution for the statistics to evaluate p .
- `help("pvalues")` within `lme4` provides discussion and guidance on possible ways to approach this if necessary.
 - e.g. different finite-size-correct p -values.

- An alternative is to compute confidence intervals.
- This can be done for fixed effects using `confint` on an `lmer` model output.
 - This can be done using multiple methods including Wald approximations and parametric bootstraps.

Comparing Models

- Primarily models are compared using the *deviance*

$$\text{deviance} = -2 * \text{loglikelihood}$$

- As a result, it is often denoted $-2LL$ or $-2(\ln)L$
- Deviance is a measure of how well the model fits our data.

Likelihood ratio test

- To compare models, we can look at the difference in deviance between two models.
- The resultant test is called the likelihood ratio test (LRT) or chi-square difference test.

$$LRT = deviance_{M0} - deviance_{M1}$$

- This difference is chi-square distributed with degrees of freedom equal to the difference in the number of model parameters.
- A significant LRT (at a priori α) indicates that the more complex model (M1) has significantly improved our model.
 - Failure to reject the null indicates it does not.

Nested & Non-nested models

- The LRT is only appropriate when models are nested.
- Nested models are when the parameters of one model are a subset of another model.
- Many models we wish to compare will not be formally nested:
 - For example, they may have a different set of fixed effects.

$$AIC = -2\ln(L) + 2k$$

$$BIC = -2\ln(L) + \log(N)k$$

- Where $\ln(L)$ = log-likelihood of the model, and k = no. of estimated parameters.
 - Values of both the AIC and BIC are not informative in their own right, but are informative when comparing alternative models.
- Important observation is that the parsimony penalty for the BIC is bigger.

Model comparison from lmer

- This is somewhat straight-forward to do.
- We can use the `anova()` function on a series of nested `lmer` models.
- For any individual model we can use `logLik()`, `deviance()`, `AIC()` and `BIC()` to get their namesake values.

Model Building

Approaches to model building

- **THEORY** (look back at specification)
- Two broad approaches (as is often the case)
 - Top-down
 - Bottom-up
- There is lots of discussion about best approaches.

- ① Loaded mean structure model where all level 1 covariates (fixed effects) and interactions are included in the model with a view to explaining as much systematic variance as possible.
- ② Add random effects and test based on LRT.
- ③ Once all the effects are added, fit alternative residual structures to find the best.
- ④ Finally reduce the model by removing any fixed effects which are not contributing.

- 1 Estimate the unconditional model and partition variance (as we did to calculate ICC)
- 2 Add level 1 covariate.
- 3 Add the associated level 2 random effect.
- 4 Repeat (2) and (3)
- 5 Add level 2 covariates to explain variation in level 1 units captured in the random effects

Which strategy?

- If you take a bottom up approach, we need to use ML.
 - This is because fixed and random effects are added in pairs.
- So choice may be dependent on what we are testing and how we are building our model.

Model Assumptions

- The key assumptions in LMM are essentially identical to that of LM.
- Here we will briefly look at our two primary assumption checks, namely;
 - Normality of residuals
 - Homoscedasticity
 - Linearity

Differences to LM Assumptions

- The main difference we contend with in LMM is that we do not have a single set of residuals.
- Recall we essentially have multiple linear models when we include random effects, each of which has, at a minimum, an intercept and a residual term.
- So we have checks at multiple levels.

Testing Assumptions

- However, we still use much the same tools as we use in standard linear models.
- `lme4` has a number of built in plots that use the general `plot()` function on an `lmer` object.
- Example code from Bates et al. (2015) *Journal of Statistical Software*, DOI 10.18637/jss.v067.i01

Testing Assumptions - Plots (some different ways in lab)

```
# Where fm1 = model output  
# Residual vs fitted plots  
# (linearity and equal variance along the line)  
plot(fm1, type = c("p", "smooth"))  
  
# Scale location plots (Homoscedasticity)  
plot(fm1, sqrt(abs(resid(.))) ~ fitted(.),  
      type = c("p", "smooth"))  
  
# This can also be looked at by group factor  
plot(fm1, sqrt(abs(resid(.))) ~ fitted(.) | group,  
      type = c("p", "smooth"))  
  
# qq plots (normality of residuals)  
qqmath(fm1, id = 0.05)
```

Worked Examples

- Let's switch over and look at an example

That's all for today