Exercises for linear mixed effect models, part 2

Timothée Bonnet

February 21, 2019

Contents

1 Uncertainty in random effects

1.1 Warming-up

* Exercise 1 Reading a summary in lme4

1. Load the dataset "thorndata.txt" to fit a linear model of "herbivory" as a function of "thorndensity". What is the estimate for the slope?

2. Add a random effect for site How does the estimate for the slope changes?

3. How much variation is explained by differences among blocks? Is there a measure of uncertainty for this estimate?

1.2 Confidence intervals and Tests

Sometimes random effects are part of the experimental design and are in the models only to control for confounding effects. But sometimes we care about their value or their statistical significance.

* Exercise 2 CI for variance components in lme4

Use the function `confint` to estimate confidence intervals for the variance component for "site".

* Exercise 3 Testing variance components in lme4

Use the function anova to test the statistical significance of the random effect "site" in the thorn dataset.

*** Exercise 4 Does null-hypothesis testing work for random effect?

In principle we want p-values to follow a uniform distribution under the null-hypothesis; in particular the frequency of p-values below 0.05 should be 0.05 when the effect tested for is absent. Is it the case for tests of random intercepts? Simulate some data sets with a random intercept that has a variance of zero, test the significance of this random effect with anova and record the p-values. You can use the following simulation template:

```
RandomVariance <- 0
sampsize <- 500
nbblocks <- 30

pvals <- vector(length = 1000)

for (i in 1:1000)
{
  x <- rnorm(sampsize,mean = 4, sd=0.25)
  block <- sample(x = 1:nbblocks, size = sampsize, replace = TRUE)
  blockvalues <- rnorm(n = nbblocks, mean = 0, sd = sqrt(RandomVariance))
  y <- 8 - x + blockvalues[block] + rnorm(sampsize,0,1)
  dat <- data.frame(response = y, predictor = x, block=block)

  XXX <- lm(XXX)
  XXX <- lmer(XXX)
  XXX <- anova(XXX)
  XXX
}
```

*** Exercise 5 Do kangaroos have personalities?

We want to know whether the distance at which kangaroos run away when we approach is consistent within individuals. Load the dataset "roo.csv" and fit linear mixed models of "EscapeDistance" to find out. What variables to include? Does individual repeatability explain more variance than expected by randomly grouping observations? What is the repeatability of behaviour?

** Exercise 6 Testing variance components in MCMCglmm

lme4 can have difficulties estimating random effect parameters when models get a bit complex. An very powerful alternative, I recommend MCMCglmm. Try and fit one of the kangaroo models. Use summary and plot to discuss the importance of the random effects.

2 Beyond random intercepts

So far we have considered random effects around intercepts only (that is the meaning of the "1" in the lme4 syntax (1| re)). But random effects can be around fixed effects. You may have heard of "random interactions", "random slopes", "random regressions"...

** Exercise 7 Beetles: build a model

Load the dataset "beetles.csv". It contains (fake) data from an (real) experiment on gene-by-environment interactions. The variable of interest is the mass of beetles born

in two different environments, from different parents, and in different cages. Assuming that we can measure genetic varition with parent random effects, we wonder if different genomes respond differently to different environments. **Build the model corresponding to this question in lme4.**

(hints: you could start from a lm() of mass modeled by environment, then add random intercepts, and finally a little something more).

** Exercise 8 Beetles: look at the model

What are the variances related to genetic differences? How are they correlated? Does genetic variation explain a lot of the total variation we observe? Try and draw a representation of genetic variation in the two environments.

*** Exercise 9 Beetles: interpret

Interpret model outputs (use raw numbers and / or graphes) to answer the following:
Is there evidence for genetic variation? Do the two environment differ in their effects on beetles?
Is there evidence for genetic variation in the response to the environment?
Does that mean that genomes good at environment 1 are bad at environment 2?

2.1 Correlated random effects

In all of the above, we have assumed that random effect levels to be perfectly correlated (e.g., observations from the same year) or not at all correlated (e.g., observations from different years). It can be very interesting to allow for intermediate values, in particular for models of spatio-temporal autocorrelation, phylogenetics, quantitative genetics.