Exercises for linear mixed effect models, part 3

Timothée Bonnet

March 29, 2019

Contents

1 Random interactions

1.1 Random slopes

* Exercise 1 Random slopes and unbalanced data

Load the dataset `hares.csv`. It contains (fake) measurements of snowshoe hare color (darkness) and their detectability against the background where they live. Measurements were taken in 50 different locations. We want to know whether darkness has an effect on detectability.

1. Fit a simple linear model of detectability on darkness. What is the effect?
2. Add a random intercept to the previous model. Does it change the result quantitatively?
3. Add a random slope. What do you see now?

** Exercise 2 Visualize random slopes

Visualize the effect from the random slope model, that is, plot the relationship detectability over darkness for every location. Add the overall relationship (for instance extracted from a simple linear regression). You can use the functions `ranef()` and `fixef()`. Why did the fixed effect of darkness changed so much when you added the random slope?

** Exercise 3 Does natural selection vary?

Load the dataset `AllM.txt`. It contains true data from the long term monitoring of a wild animal population. We are interested in quantifying natural selection on Weight. To simplify let's assume natural selection is the slope of `fitnessR` on `Weight`.

1. Fit a linear regression of fitnessR on Weight. Include `Age` as a predictor Is there evidence for selection?
2. Change your model to a mixed model with year as a random intercept. Do you think fitnessR varies a lot among years?
3. Now add a random slope on weight. How much variation is there in selection?
4. Make a graph to visualize selection on different years (the function `ranef()` extract random effects) (you can make the graph for adults, for juveniles, or both).
5. Looking at the estimated variance for the intercept and for Weight, which one looks more important? Is that your impression graphically? Why?
6. Bonus: test for the statistical significance of the variation in selection (you can use `anova()` to compare two models).

1.2 Random factor interaction

* Exercise 4 Random interaction with a factor

Load the file `interfactor.csv` and fit two random interaction modelw of y as a function of treat, using both the reaction norm and the character state approach. How do the estimate differ? Use the functions `AIC()`, `fitted()` and `resid()` to compare the fit of the two models? What can you conclude?

** Exercise 5 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 6 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 7 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 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.

2.1 Quantitative genetics

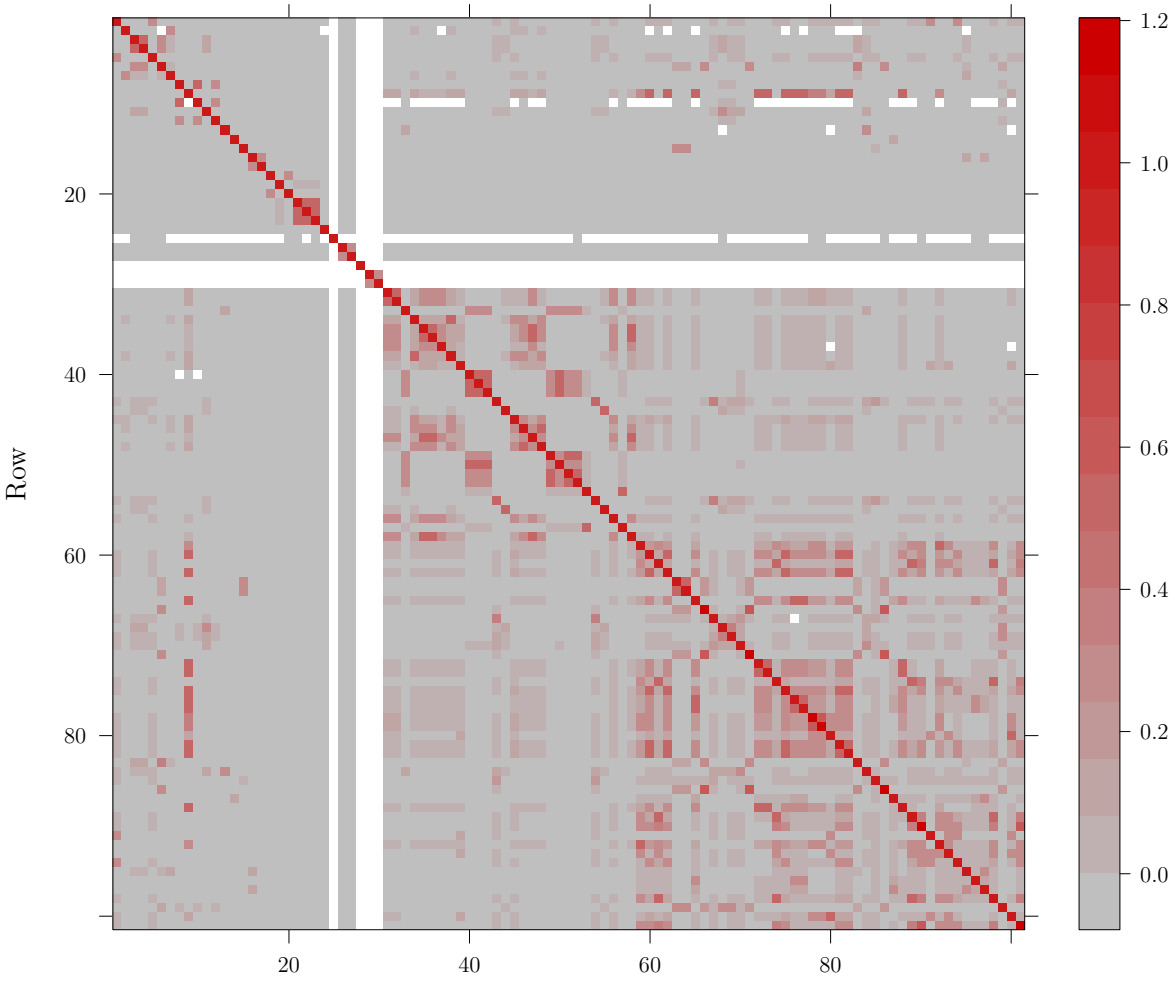
Genes are transmitted from parents to offspring in a very predictable way: their genes have a correlation of 50%. Therefore it is possible to estimate the genetic variance

without any DNA sequencing when a dataset contains parents and offspring. Going back to the long term monitoring of the wild animal population, let's load the population pedigree (ped.txt) and calculate this relatedness matrix:

```
library(MCMCglmm)

## Loading required package:  coda
## Loading required package:  ape

ped <- read.table("ped.txt", header = TRUE)
ainv <- inverseA(ped)$Ainv #the inverse of relatedness matrix
image(solve(ainv)[500:600,500:600], useRaster=T) #the relatedness matrix
```



Dimensions: 101 x 101

Is there genetic variation in weight? Here is a demonstration of fitting a quantitative genetic model.

```
allm <- read.table("AllM.txt", header = TRUE)

mweight <- MCMCglmm(Weight ~ 1 + Age*Sex, random=~Year + id,
                    ginverse = list(id=ainv), data=allm)
summary(mweight)
```

The effect "id" has a large variance attached to it, suggesting the presence of a lot of genetic variation.

* Exercise 8 Is it really genetic?

However that was a bit cheating, because individuals had several observations, so the "genetic" random effect may be just repeated measurements. Let's add another random effect for individual, but not connected to the relatedness matrix (you can use the variable "animal" which is a duplicate of "id").

2.2 Phylogenetic model

In a phylogenetic tree some species have a longer common evolutionary history than others. What those species look like may be influenced by the common evolutionary history. We can model that by considering phylogenetic correlations between two lineage as the time of common evolution relative to their outgroup.

* Exercise 9 Demo of Phylogeny and correlated phenotypes

Load a phylogeny of bird families and some bird phenoypes:

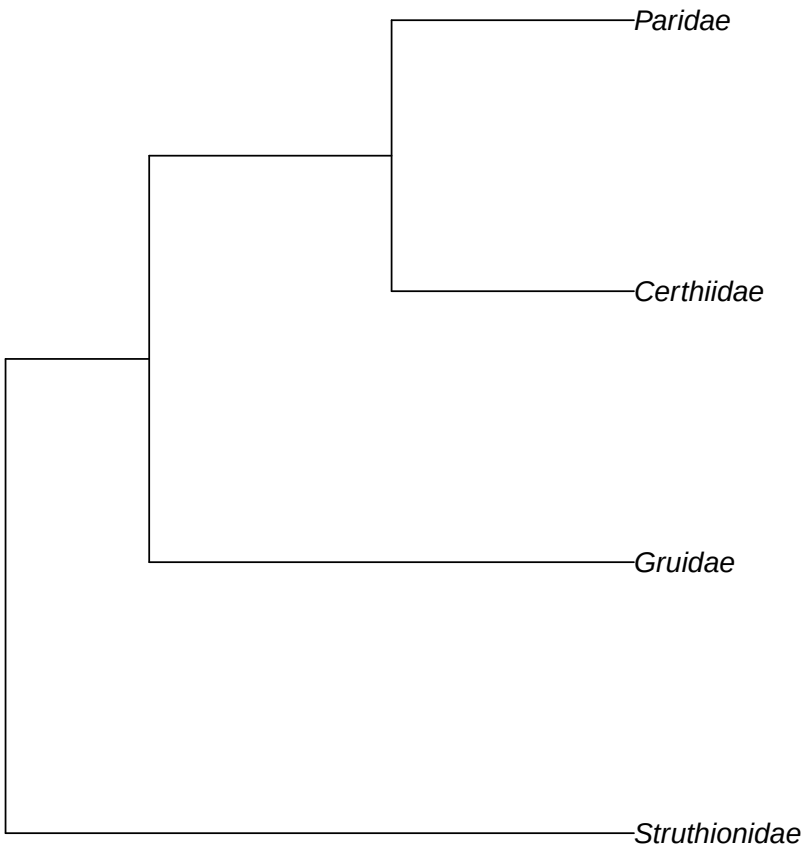
```
load("birdfamilies") #phylogeny. Loading creates the object bird.families
birds <- read.csv("birdpheno.csv") #family phenotypes
```

To start, we subset the phylogeny to a few families

```
bird.families <- makeNodeLabel(bird.families)
some.families <- c("Certhiidae", "Paridae", "Gruidae",
"Struthionidae")
Nphylo <- drop.tip(bird.families, setdiff(bird.families$tip.label,
some.families))
```

So we can easily visualize it:

```
plot(Nphylo)
```



Which families will tend to have more similar phenotypes do you think?

We can calculate the variance covariance matrix of that tree as:

```r
library(MCMCglmm)
INphylo <- inverseA(Nphylo)
sA <- as.matrix(solve(INphylo$Ainv))
colnames(sA) <- rownames(sA) <- rownames(INphylo$Ainv)
sA
```

```
##               Node58 Node122 Struthionidae Gruidae Certhiidae Paridae
## Node58        0.2286  0.2286             0  0.2286     0.2286  0.2286
## Node122       0.2286  0.6143             0  0.2286     0.6143  0.6143
## Struthionidae 0.0000  0.0000             1  0.0000     0.0000  0.0000
## Gruidae       0.2286  0.2286             0  1.0000     0.2286  0.2286
## Certhiidae    0.2286  0.6143             0  0.2286     1.0000  0.6143
## Paridae       0.2286  0.6143             0  0.2286     0.6143  1.0000
```

What are the zero?

Coming back to the full dataset, we can fit a phylogenetic model as:

```r
INphylofull <- inverseA(bird.families) # Object contains inverse relatedness matrix,
prior0 <- list(G=list(G1=list(V=1, nu=1, alpha.mu=0, alpha.V=100)),
               R=list(V=1, nu=0.002))
m1 <- MCMCglmm(y ~ 1, random = ~id, ginverse = list(id=INphylofull$Ainv),
               data = birds, prior = prior0)
summary(m1)
```