

Mixed effect models with inter-group dependency - The Animal Model

Henry Kirveslahti

11/17/2019

Introduction

This document is a very brief introduction on how to fit animal models with brms. The intended audience is the students in STA410/STA610. As I got lots of enthusiastic feedback for the Inverse-Laplacian from Case Study 1, I thought it would be interesting to say something in connection to that. What I didn't tell in the Case Study 1 presentation is that the idea of using mixed effects models framework to capture spatial correlations was inspired by the animal model framework. As this course provides excellent background for working with these models, I thought it would be interesting to 1) state what animal models are 2) show how one can fit these in brms and 3) point to some references for further reading.

Mixed effects models

The usual setup we have in mixed effect models: Individual i in group j with response Y_{ij} and some covariates X_{ij} . Suppose we want to fit a linear regression on Y explained by X , and that for a subset $Z \subset X$ we want to have group-specific random effects. As seen in the class, we write:

$$Y_{ij} = X_{ij}\beta + Z_{ij}b_j + \epsilon_{ij},$$

where $b_j \sim^{iid} N(0, \Sigma_b) \perp \epsilon_{ij}$.

Punchline:

There is no need to assume b_j are independent (with brms, `lm4` can't quite fit these), and in many cases we know something about the inter-group correlations. If, instead, we write

$b_j \sim MVN(0, \Sigma_b \otimes \Psi)$, with Ψ some sort of similarity matrix, we arrive to what is called an *animal model*. These models are widely used in ecology, such as cattle or race horse breeding. They can answer questions such as which horse will likely produce best offsprings, this is called the *breeding value*.

For simplicity, we assume b_j be one-dimensional.

The animal model - motivation

Oftentimes the ecologists are concerned about variability in some phenotype p and how it passes to progeny. Put simply, the variance V_p of p decomposes into three pieces environmental variance V_E , additive genetic variance V_A and non-additive genetic variance V_{NA} :

$$V_p = V_E + V_A + V_{NA}.$$

The ratio

$$h^2 = \frac{V_A}{V_p}$$

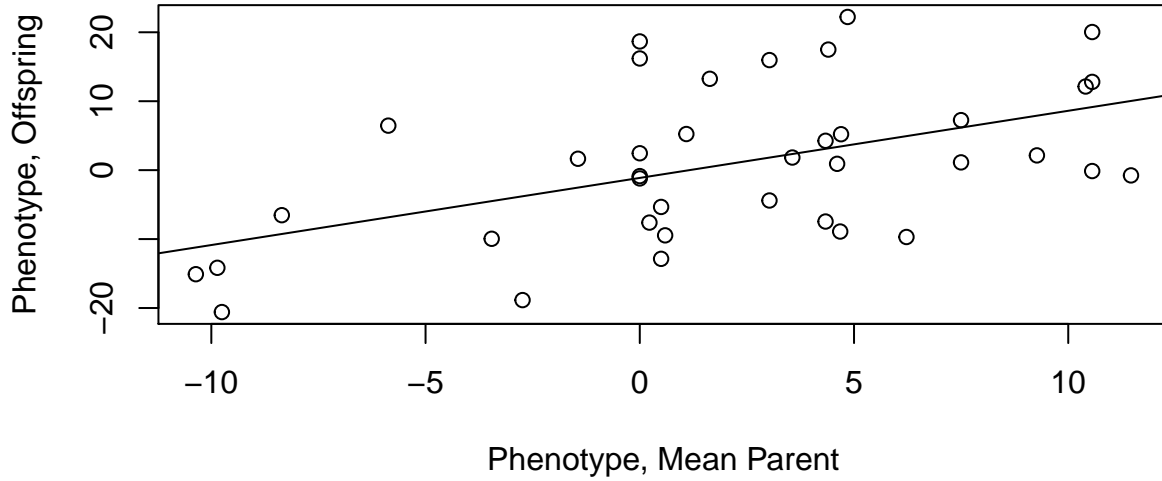
is called the *heritability* of p in population P and the breeding value of individual i is $h^2 \cdot p_i$. Hence, estimating the heritability of phenotype p is of central interest to animal breeders. The idea is that h^2 is the regression slope for parent-offspring regression.

the *breeding value* of individual i is

$$(Y_{ij} - \bar{Y})h^2.$$

- Breeding value tells us the expected value of using subject i as a parent if we are interested in producing offsprings with higher value of Y .

Indeed, parent-offspring regression gives a means to estimate heritability of phenotype p , but in many cases the population is inbred and this regression technique ignores some of the family structure. Also, since the heritability is by definition between 0 and 1, linear regression may give bad estimates especially if there is not much data.



Instead, we can fit a mixed effects model where we set Ψ to be the pedigree matrix.

How to fit animal models with brms

In brms, we can specify the dependency structure by the `coefs_ranev` parameter.

Computing the pedigree matrix

List the animals $a_i = a_1, \dots, a_N$ monotonically by generations. For $j \leq i$, and parents $p_1[i], p_2[i]$.

If both $p_1[i], p_2[i]$ are observed: $\Psi_{i,j} = \Psi_{j,i} = \frac{1}{2}(\Psi_{j,p_1[i]} + \Psi_{j,p_2[i]})$ and $\Psi_{i,i} = 1 + \frac{1}{2}\Psi_{p_1[i],p_2[i]}$.

If only $p_1[i]$ is observed:

$$\Psi_{i,j} = \Psi_{j,i} = \Psi_{j,p_1[i]}, \Psi_{i,i} = 1.$$

Otherwise:

$$\Psi_{i,j} = \Psi_{j,i} = 0 \text{ and } \Psi_{i,i} = 1.$$

R can do this automatically with the MCMCglmm package:

```

library(brms)
PEDA=data[,c(1,2,3)]
##Prepare the relationship matrix
inv.A <- MCMCglmm::inverseA(PEDA)
A <- solve(inv.A$Ainv)
rownames(A) <- rownames(inv.A$Ainv)

m1 <- brm(Phenotype ~ (1|Individual), data = data,
         family = "gaussian", cov_ranef = list(Individual = A), control = list(adapt_delta = 0.95), pr
         brms::prior(cauchy(0, 1), class = sd),
         brms::prior(cauchy(0, 1), class = sigma)),warmup = 6000, iter = 10000)

post<-posterior_samples(m1)
var_A=(post[,2])^2
var_total=(post[,2])^2+(post[,3])^2

```

We can then look at heritability:

```
summary(var_A/var_total)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## 0.0000023 0.9657601 0.9920681 0.9555831 0.9981614 0.9998502
```

And the breeding values:

```

breds=post[,c(4:45)]
names(breds)<-c(1:42)
colMeans(breds)

```

```
##      1      2      3      4      5      6
## 9.7590893 -9.4687399 9.8628273 -9.3315621 14.9667169 2.2950069
##      7      8      9     10     11     12
## 16.8877485 -0.9980951 7.3210010 -0.8246005 1.3472979 18.7446378
##     13     14     15     16     17     18
## -5.7327285 -12.1185194 -18.2213214 2.5970885 -6.8670468 3.8581636
##     19     20     21     22     23     24
## 12.3202152 15.0964414 1.7902903 5.1017883 -6.5726577 -8.8740357
##     25     26     27     28     29     30
## 16.4855029 -0.9930585 -8.7413976 0.9477646 11.8271449 -0.1913946
##     31     32     33     34     35     36
## 1.0997773 -14.2043901 -19.9378174 -9.3927155 20.5485362 -3.9065003
##     37     38     39     40     41     42
## 5.6950056 4.8375454 -7.1486515 -14.7428719 -8.0592928 12.3585750
```

References:

- Code for automatically getting pedigree matrix: <https://github.com/JonBrommer/Multivariate-Mixed-Models-in-R/wiki/brms-examples>
- Lectures on animal model: Guilherme J. M. Rosa: <https://www.biostat.washington.edu/sites/default/files/modules/Seattle-SISG-18-MM-Lecture04.pdf>
- More on parent-offspring regression: <http://content.csbs.utah.edu/~rogers/ant5221/lecture/QTs2.pdf>
- de Villemereuil, Gimenez, Doligez: Comparing parent-offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: A simulation study for Gaussian and binary traits, Methods in Ecology and Evolution 2013