Multivariate Miscellany
Lecture 11.3 Multivariate Correlation Structure

Lauren Sullivan

Module: Multivariate Models

Readings

Required for class:

- ► NA

Optional:

- ► Latin American Macroevolution Workshop (2017) Phylogenetic generalized least squares regression.

- ► Pinheiro, J. C. and Bates, D. M. (2000) Mixed-Effects Models in S and S-Plus - Correlation

# More Complicated Correlation Structure

Sometimes when runinng linear models (and generalized linear models) you want to account for a correlation structure that is more complicated than simple autocorrelation in one dimension (e.g. space/time).

- ▶ You may want to include an entire correlation matrix (or distance matrix).
  - ▶ In essence, you are using this matrix as a random effect of sorts.

- ▶ For example, you may want to control for the phylogenetic relationship of all of your species or individuals to each other.

These models tend to be a lot more "custom-build", and deviate quickly from the basic ways to structure your correlation argument in `lme()` as we did with spatial/temporal autocorrelation, but we will go over some basic examples.

Autocorrelation Reminder

Looking back at our Mystery Snail example, we remember that there were 6 measurements of biomass consumption through time per individual. It may be more likely that the measurements closer in time are more likely to be more similar to each other than those farther in time.
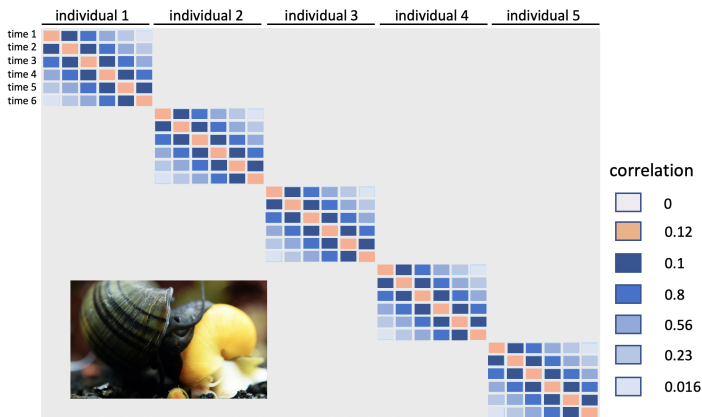
Autocorrelation Reminder

```
mod_corr <- lme(biomass ~ sex + month,
                data = snails,
                random = ~1|individual,
                correlation = corAR1(form = ~month))
summary(mod_corr)
```

```
## Linear mixed-effects model fit by REML
##  Data: snails
##        AIC      BIC    logLik
##   4178.27 4208.795 -2083.135
##
## Random effects:
##  Formula: ~1 | individual
##         (Intercept) Residual
## StdDev:    1.012856  1.36648
##
## Correlation Structure: AR(1)
##  Formula: ~month | individual
##  Parameter estimate(s):
##        Phi
## 0.4183476
## Fixed effects:  biomass ~ sex + month
##                  Value Std.Error  DF  t-value p-value
## (Intercept) 12.795193 0.15535686 999 82.36001  0e+00
## sexmale     -0.729291 0.18122413 198 -4.02425  1e-04
## month        0.535718 0.02639684 999 20.29477  0e+00
##  Correlation:
##         (Intr) sexmal
## sexmale -0.554
## month   -0.595  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.95734092 -0.58804004 -0.01897382  0.61183122  2.76106338
##
## Number of Observations: 1200
```

# Correlation Structure with a Matrix

Phylogenetic Data Example

We will be looking at an example of Asian Barbets, used in this macroevolution workshop. These scientists were interested in understanding the factors associated with the evolution of song in these birds.

```
barbet[1:10,1:4]
```

```
##                                          wing    Lnalt    patch   colour
## Calorhamphus_fuliginosus_fuliginosus 4.388257 5.298317 2.000000 1.666667
## Calorhamphus_fuliginosus_hayi        4.427239 5.298317 2.000000 1.666667
## M_armillaris                         4.532599 7.170120 6.333333 4.000000
## M_asiatica_asiatica                  4.611152 6.802395 7.333333 5.000000
## M_asiatica_davisoni                  4.605170 7.003065 6.666667 3.333333
## M_australis_duvaucelli               4.282206 6.620073 9.000000 4.000000
## M_chrysopogon                        4.829113 6.620073 9.000000 5.333333
## M_corvina                            4.824306 7.313220 2.000000 2.000000
## M_eximia                             4.359270 6.620073 7.666667 5.000000
## M_faiostricta                        4.693181 5.298317 5.000000 5.000000
```

Phylogenetic Data Example

We need to take into account the possible correlations among traits in closely related species.
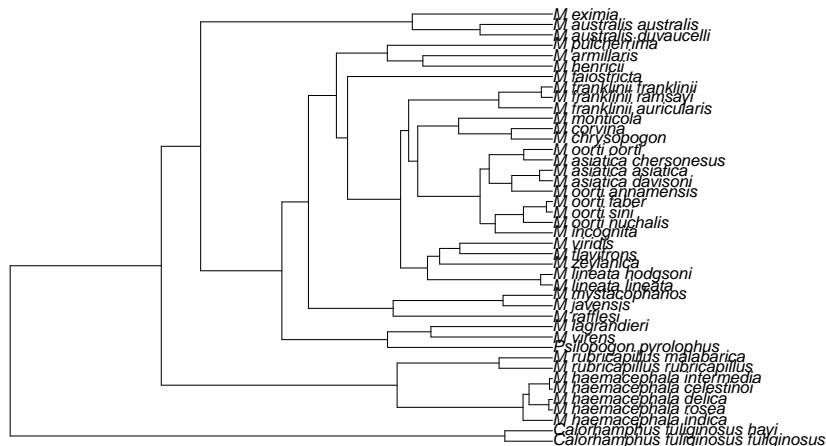
Data Needs to Match

You need to make sure you have the same species and spellings across your tree and your dataset. You can use `name.check()` for this.

```
obj<-name.check(barbet_tree, barbet)
obj
```

```
## $tree_not_data
## [1] "M_asiatica_chersonesus"  "M_australis_australis"
## [3] "M_haemacephala_celestinoi" "M_haemacephala_delica"
## [5] "M_haemacephala_intermedia" "M_haemacephala_rosea"
## [7] "M_lineata_lineata"         "M_oorti_faber"
## [9] "M_oorti_sini"
##
## $data_not_tree
## character(0)
```

You see there are 9 species in the tree that are not in the dataset, so we can drop them using `drop.tip()`, and then check the names again.

```
barbet_cuttree <- drop.tip(barbet_tree, obj$tree_not_data)
name.check(barbet_cuttree, barbet)
```

```
## [1] "OK"
```

Phylogenetic Least Squares Model

This is equivalent to a linear model with the correlation structure that matches the phylogeny assuming Brownian motion.

First, let's create the correlation matrix.

```
bm<-corBrownian(1, barbet_cuttree)
bm
```

```
## Uninitialized correlation structure of class corBrownian
```

Phylogenetic Least Squares Model

Now run the model with `gls()` assuming Brownian motion. This is a general least squares model.

```
mod1 <- gls(Lnote ~ Lnalt, data = barbet, correlation = bm)
summary(mod1)
```

```
## Generalized least squares fit by REML
##   Model: Lnote ~ Lnalt
##   Data: barbet
##         AIC       BIC    logLik
##   -14.84531 -10.54335 10.42265
##
## Correlation Structure: corBrownian
##  Formula: ~1
##  Parameter estimate(s):
## numeric(0)
##
## Coefficients:
##                   Value   Std.Error    t-value p-value
## (Intercept)  0.16821373 0.21394057  0.7862638  0.4377
## Lnalt       -0.00761511 0.02034964 -0.3742136  0.7108
##
##  Correlation:
##       (Intr)
## Lnalt -0.643
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.30893712 -0.27804583 -0.19007275  0.08217655  1.31021476
##
## Residual standard error: 0.2967954
## Degrees of freedom: 33 total; 31 residual
```

Phylogenetic Least Squares Model

What about if we assume Pagel's λ model? This method stretches tip branches relative to internal branches.

```
mod2 <- gls(Lnote ~ Lnalt, data = barbet,
            correlation = corPagel(1, barbet_cuttree))
summary(mod2)
```

```
## Generalized least squares fit by REML
##   Model: Lnote ~ Lnalt
##   Data: barbet
##         AIC       BIC    logLik
##   -32.81041 -27.07446 20.40521
##
## Correlation Structure: corPagel
##  Formula: ~1
##  Parameter estimate(s):
##      lambda
## -0.06003187
##
## Coefficients:
##                   Value Std.Error    t-value p-value
## (Intercept) -0.12977761 0.17841865 -0.7273769  0.4725
## Lnalt        0.03723651 0.02741941  1.3580345  0.1843
##
##  Correlation:
##       (Intr)
## Lnalt -1
##
## Standardized residuals:
##         Min         Q1        Med         Q3        Max
## -1.02117658 -0.61438211 -0.24990317  0.07528683  3.18190947
##
## Residual standard error: 0.4108274
```