Lesson 10:
The animal model

The limits of GLS

- We have learnt how to run linear models with complex autocorrelation structures (GLS)
- Unfortunately, the method has limits:

1. It assumes the response data is normally distributed
2. It cannot handle replication of individuals in species

So how can we deal with repetition and response data with non-normal distributions?

The animal model

- The problems described in the previous slide can be dealt with using an extension of mixed models and thus extending the covariance matrix **V**

- Recall, the general likelihood solution of the coefficient estimates in linear models is:

$$\widehat{\boldsymbol{\beta}} = \ (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$

- In the case of ordinary linear models,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \ \boldsymbol{\varepsilon} \quad \text{where} \quad \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}\sigma_R^2), \ \text{so}\ \boldsymbol{V} = \boldsymbol{I}\sigma_R^2$$

- Hence:

$$\widehat{\boldsymbol{\beta}} = \ (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \qquad (\because \boldsymbol{V}^{-1} = \boldsymbol{I}^{-1} = \boldsymbol{I})$$

The animal model

- In mixed models with grouping random effects, $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{Z}\sigma_B^2), \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}\sigma_R^2)$,
- as random effects are simply variance terms,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}\sigma_B^2 + \boldsymbol{I}\sigma_R^2, \qquad \text{so } \boldsymbol{V} \sim \boldsymbol{Z}\sigma_B^2 + \boldsymbol{I}\sigma_R^2$$

- Phylo GLS is a special case where the residual errors are distributed according to the phylogeny covariance matrix A

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad \text{where} \quad \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{A}\sigma_R^2), \text{ so } \boldsymbol{V} \sim \boldsymbol{A}\sigma_R^2$$

- It is easy to see that phylogenetic covariance can instead be included as a separate random term,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}\sigma_B^2 + \boldsymbol{A}\sigma_a^2 + \boldsymbol{I}\sigma_R^2, \qquad \text{so: } \boldsymbol{V} \sim \boldsymbol{Z}\sigma_B^2 + \boldsymbol{A}\sigma_a^2 + \boldsymbol{I}\sigma_R^2$$

The animal model

- The expression:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}\sigma_B^2 + \boldsymbol{A}\sigma_a^2 + \boldsymbol{I}\sigma_R^2$$

- is known as the animal model because it was first developed in quantitative genetics research for animal breeding, where ancestry was an important consideration for breed traits. Phylogenies are similar to ancestries, so model applicable in comparative biology

- The elegance of using the phylogenetic data as an additive term in the model means that we can now simply use the mixed model framework to combine multiple individuals per species and non-normal response data with phylogenetic information in GLMM

A crisis of code

- The only package I am aware of that runs the animal model using the frequentist statistical framework is *asreml*. Unfortunately we have to pay to use this software!!!
- By contrast there are several packages for handling the animal model using the Bayesian statistical framework. Therefore we will run the animal model using Bayesian packages in R. I am aware of two packages that do this well: *MCMCglmm* and *brms*. We will use *brms* here.
- The linear function in brms has been set up to mimic glmer() so is relatively easy to use

Frequentism vs Bayesianism

- Before going any further, its important that we say something very briefly about the main differences between frequentist versus bayesian approaches

- It is necessary to understand this to understand the basics of the code

Frequentism

- Frequentists use only the data collected in the present experiment to draw inferences.

- Thus the likelihood function for a normally distributed response variable uses only the present information

$$L(\boldsymbol{y}, \boldsymbol{X}, \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}\sigma^n}\, exp\left(-\frac{1}{2\sigma^2}\,(y - \quad)^{\mathbf{2}}\right)$$

- **μ** is representative for the betas here

Bayesianism

- Bayesianists assume that there is pre-existing data describing the parameter/s (**μ**,**σ**) that can be combined with the present information to draw inferences. This information is known as the prior information

- Thus the likelihood function uses only present information

$$L(y, X, \boxed{?}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, exp\left(-\frac{1}{2\sigma^2}(y - \boxed{?})^2\right)$$

- The Prior is earlier information describing the parameters.

$$P(\mu, \tau^2) = \frac{1}{\sqrt{2\pi\tau^2}}\, exp\left(-\frac{1}{2\tau^2}(\mu - M)^2\right)$$

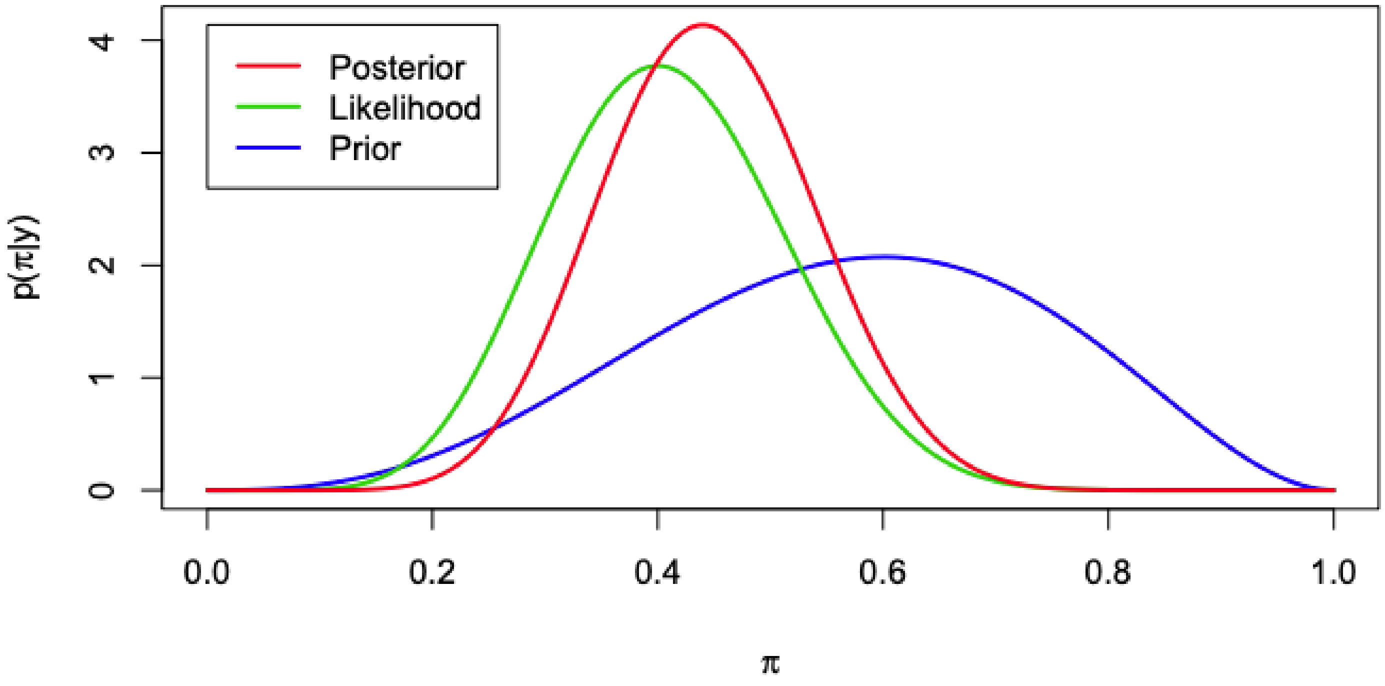
- Where M and **τ** are estimates of **μ** and **σ** from earlier experiments/knowledge.

Bayesianism

- Posterior = Likelihood × Prior
- $P(\mu, \sigma^2) \propto \frac{1}{\sqrt{\tau^2 \sigma^2}}\ exp\left(\frac{-(\mu - M)^2}{2\tau^2} + \frac{-(y - XB)^2}{2\sigma^2}\right)$

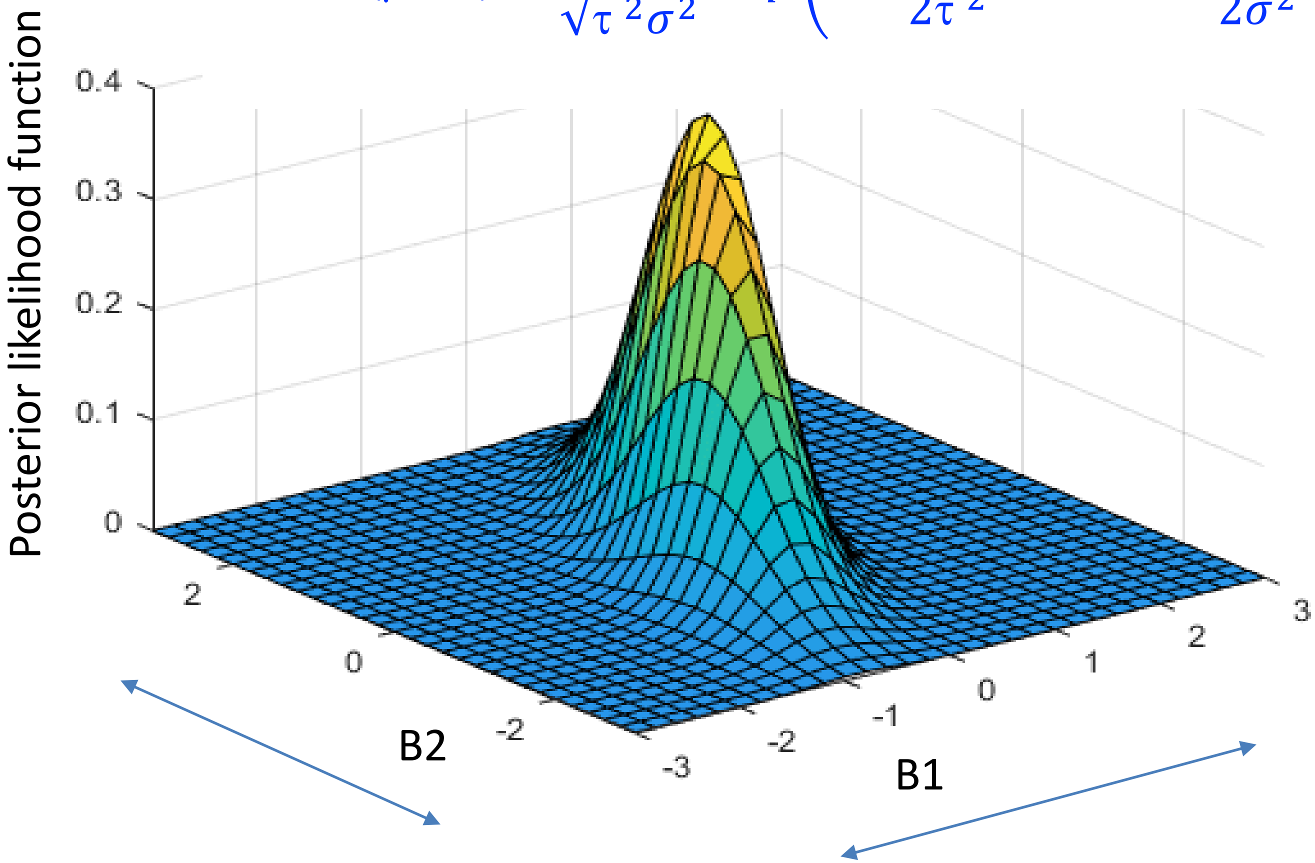
- Then takes logs and solve posterior likelihood as before
- -log Posterior = -log Likelihood + -log Prior

Frequentism vs Bayesianism



Parameter drawing in Bayesian models

$$P(\mu, \sigma^2) \propto \frac{1}{\sqrt{\tau^2 \sigma^2}} exp\left(\frac{-(\mu - M)^2}{2\tau^2} + \frac{-(y - XB)^2}{2\sigma^2}\right)$$



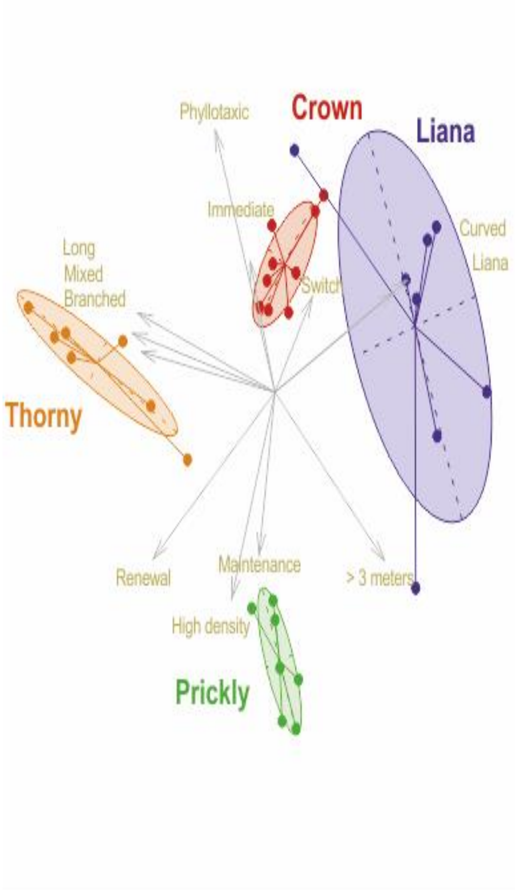
- Sample the predictor space and draw the distributions

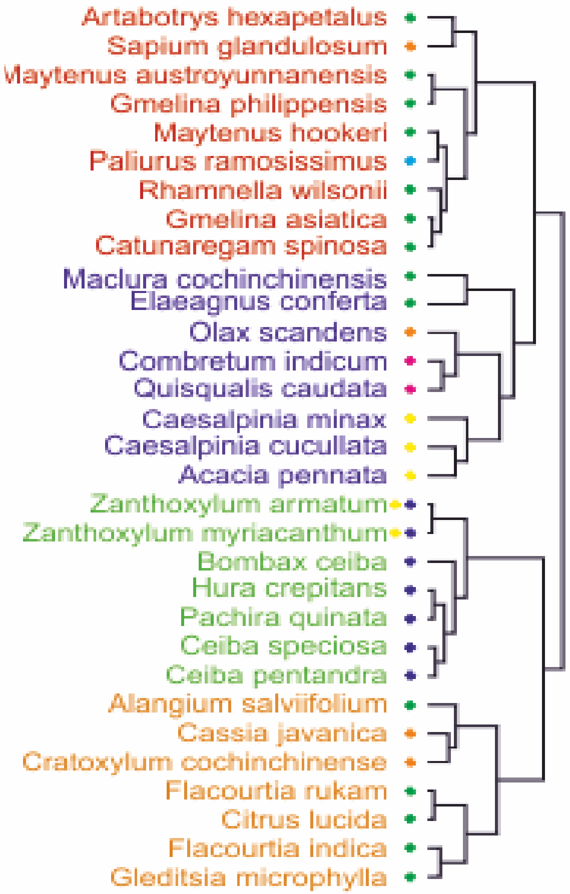
The brm() glmm function

```
•  mod <- brm(Dlog~x1+(1|phy)+(1|Sp), cov_ranef =
   list(phy = phylo_cor),data = bark2, family =
   gaussian(), sample_prior = TRUE,
         iter = 20000,warmup = 10000, chains = 4, cores =
4, thin = 10, save_all_pars = TRUE,seed=T,
control=list(adapt_delta=0.99))
```

- Priors can be “informative” or “non-informative:”
- These will be discussed in the next lecture
- For now we will accept the defaults in the brm() function.

Spiny trunk data





- Different species with spiny trunks cluster out.
- Do they have different nutritional value and defence efficiency?

Code 10.1
Linear model with species repeats