Update on working PMM model

By Lizzie
29 June 2021

Let's start with the basic **PMM (phylogenetic mixed model)**

As I understand it:

$$y = \alpha + \beta x + a + e \tag{1}$$
$$a \sim normal(0, \sigma_P^2 \Sigma) \tag{2}$$
$$e \sim normal(0, \sigma_R^2 I) \tag{3}$$

... where α and β, respectively, are the intercept and the slope for the co-factor x, a is the phylogenetic random effect, and e is the residual error. Σ is a phylogenetic correlation matrix, I stands for the relevant identity matrix. (The above from Chapter 11: General Quantitative Genetic Methods for Comparative Biology, by Villemereuil & Nakagawa, though seems similar to Housworth.)

What we would like to do is have the phylogenetic effect structure the species-level slopes (while also allowing partial pooling on the species given uneven data, that is, for some species we might have 3 observations, each at a different x value, and for other species, we have 30 observations). We think we have this running in `ubermini_2.R` (simulation code) and `ubermini_2.stan`.

In the Stan code, β is:

```
b_force~ multi_normal(rep_vector(b_z,n_sp),
lambda_vcv(Vphy, lam_interceptsb, sigma_interceptsb)
```

So the `b_z` represents the root trait value, and all the action happens in the second part of the function, which calls `lambda_vcv`, which is defined at the top of the Stan code. This function sets up two matrices (both the size of the VCV in your data): (1) a matrix of your VCV with λ on the off-diagonals only (`local_vcv`) and (2) a matrix with σ on the diagonal (`sigma_mat`) and then returns the product `sigma_mat` * `lambda_mat` * `sigma_mat`. So that, when you put the β code above with the `lambda_vcv` function you get:

$$\beta \sim MVN(\mu, \Phi\Sigma)$$
$$\Phi\Sigma = \begin{bmatrix} \sigma_\beta^2 & \lambda\sigma_\beta^2 & \lambda\sigma_\beta^2 \\ \lambda\sigma_\beta^2 & \sigma_\beta^2 & \lambda\sigma_\beta^2 \\ \lambda\sigma_\beta^2 & \lambda\sigma_\beta^2 & \sigma_\beta^2 \end{bmatrix}$$

What we tried...
From May to October 2020, Lizzie worked on this model (getting help from others, with help on the matrix work especially Will Pearse). We tried a couple forms of the $MVN(\mu, \Phi\Sigma)$ (see `notes/phylabtalk_oct2020` for an overview) to get to this current one.

It looks like it does not matter in the current test code whether or not the VCV goes in as a correlation matrix or not. This seems odd to me.

We have this code running on simulations with just slope (`ubermini_2.R`) as well as with an intercept and up to three slopes (phylogeny on all slopes and intercept), see the simsmore folder.

PGLS, as I understand it (in case you want to compare...)

PGLS, allowing λ to vary ...

$$y \sim MVN(\mu, S) \tag{4}$$
$$\mu = \alpha + \beta * x \tag{5}$$
$$\tag{6}$$

where μ is a usual linear model, y is reponse data (one per species), and S is a covariance matrix with as many rows and columns as species. In PGLS (allowing λ to vary), Σ, the phylogenetic covariance matrix, is used to construct an S that should look like this:

$$\begin{bmatrix} \sigma^2 & \lambda\sigma^2 & \lambda\sigma^2 \\ \lambda\sigma^2 & \sigma^2 & \lambda\sigma^2 \\ \lambda\sigma^2 & \lambda\sigma^2 & \sigma^2 \end{bmatrix} \tag{7}$$