

PMM (phylogenetic mixed model)

As I understand it:

$$y = \alpha + \beta x + a + e \quad (1)$$

$$a \sim \text{normal}(0, \sigma_P^2 \Sigma) \quad (2)$$

$$e \sim \text{normal}(0, \sigma_R^2 I) \quad (3)$$

... 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.

What I 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):

$$y = \alpha + \beta x + e \quad (4)$$

$$\beta \sim \text{MVN}(\mu, \sigma_\beta^2 \Sigma) \quad (5)$$

$$e \sim \text{normal}(0, \sigma_R^2 I) \quad (6)$$

But I am not sure if my math above is correct (should it be $\beta \sim \text{MVN}(0, \sigma_\beta^2 \Sigma)$ since it's a Gaussian Process?), or how to code it in Stan.

The version I sent before was this:

$$y = \beta x + e \quad (7)$$

$$\beta \sim \text{MVN}(0, \Phi \Sigma) \quad (8)$$

$$e \sim \text{normal}(0, \sigma_y I) \quad (9)$$

$$(10)$$

In the code, β was:

```
b_force ~ multi_normal(rep_vector(0, n_sp),  
diag_matrix(rep_vector(null_interceptsb, n_sp)) + lam_interceptsb * Vphy;
```

To make this clearer (maybe), let α be `null_interceptsb` and λ be `lam_interceptsb`.

When `Vphy` is set to have 1s down the diagonal (e.g., `Vphy=vcv(phylo, corr=TRUE)`) the above simplifies to this:

$\Phi \Sigma =$

$$\begin{bmatrix} \alpha + \lambda & \lambda & \lambda \\ \lambda & \alpha + \lambda & \lambda \\ \lambda & \lambda & \alpha + \lambda \end{bmatrix} \quad (11)$$

When V_{phy} is set to have 0s down the diagonal the above simplifies to this:

$$\Phi\Sigma = \begin{bmatrix} \alpha & \lambda & \lambda \\ \lambda & \alpha & \lambda \\ \lambda & \lambda & \alpha \end{bmatrix} \quad (12)$$

And that separation seemed better to me, so that's why I did it even though I agree it seems wrong.

Perhaps something similar to what happens in PGLS (as best I understand, next page) would be better? This would be something like:

$$\beta \sim MVN(\mu, \Phi\Sigma) \quad (13)$$

$$\Phi\Sigma = \begin{bmatrix} \sigma_\beta & \lambda\sigma_\beta & \lambda\sigma_\beta \\ \lambda\sigma_\beta & \sigma_\beta & \lambda\sigma_\beta \\ \lambda\sigma_\beta & \lambda\sigma_\beta & \sigma_\beta \end{bmatrix} \quad (14)$$

where μ is a a single coefficient estimate.

PGLS, as I understand it

PGLS, allowing λ to vary ...

$$y \sim MVN(\mu, S) \tag{15}$$

$$\mu = \alpha + \beta * x \tag{16}$$

$$\tag{17}$$

where μ is a usual linear model, y is response 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:

$$\begin{bmatrix} \sigma^2 & \lambda\sigma^2 & \lambda\sigma^2 \\ \lambda\sigma^2 & \sigma^2 & \lambda\sigma^2 \\ \lambda\sigma^2 & \lambda\sigma^2 & \sigma^2 \end{bmatrix} \tag{18}$$