

Details on Bayesian phylogenetic mixed-effects models

Each population in our dataset can take on any of the following mutually exclusive kinds of social organizations: solitary, pair, stable group, or various kinds of IVSO (see main text). Such a data structure is best modeled using a multinomial distribution, wherein we estimate the probability of being in $k - 1$ of k possible states. Specifically, we here set “solitary” as our reference category, leaving any given observation Y_i to be determined by the probability of pair-living, π_{Pi} , the probability of stable group-living, π_{Gi} , and various kinds of IVSO, here summarized for simplicity by π_{IVSOi} :

$$Y_i \sim \text{Multinomial}(n, \langle \pi_{Pi}, \pi_{Gi}, \pi_{IVSOi} \rangle)$$

Here, we only have a single record, or “trial”, for each unit of observation (i.e. each population), hence $n = 1$, and we therefore have a categorical distribution as a special case of a multinomial - this is analogous to the bernoulli distribution being a special case of the binomial with a single trial, when $k = 2$.

The probabilities of pair-living, stable group-living and various kinds of IVSO then each get their own linear model:

$$\begin{aligned} \text{logit}(\pi_{Pi}) &= \alpha^P + \alpha_{Phylo[i]}^P + \alpha_{Species[i]}^P + \beta_x^P \cdot X_i \\ \text{logit}(\pi_{Gi}) &= \alpha^G + \alpha_{Phylo[i]}^G + \alpha_{Species[i]}^G + \beta_x^G \cdot X_i \\ \text{logit}(\pi_{IVSOi}) &= \alpha^{IVSO} + \alpha_{Phylo[i]}^{IVSO} + \alpha_{Species[i]}^{IVSO} + \beta_x^{IVSO} \cdot X_i \end{aligned}$$

wherein the superscripts P , G and $IVSO$ denote coefficients of the pair-living, group-living, and IVSO linear models, respectively, α is the global intercept, $\alpha_{Phylo[i]}$ is a vector of random intercepts the expected covariance of which is given by the phylogeny (see below), $\alpha_{Species[i]}$ is a vector of additional species-level intercepts accounting for repeated observations (i.e. multiple populations) of some species and capturing any between-species variation that is not captured by phylogeny. Finally, $\beta_x \cdot X_i$ are potential predictors (habitat heterogeneity, number of studies, PC1, PC2) and their associated regression coefficients.

We set weakly regularizing priors on the intercepts and slopes, centering prior probability on 0 and penalizing extreme values, which helps model convergence:

$$\alpha \sim \text{Normal}(0, 10)$$

$$\beta_x \sim \text{Normal}(0, 1)$$

The vector of phylogenetic random effects $\alpha_{Phylo[i]}$ is given by

$$\begin{bmatrix} \alpha_{Phylo[1]} \\ \vdots \\ \alpha_{Phylo[N]} \end{bmatrix} \sim \text{MVNormal}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \sigma_{Phylo} \cdot A\right)$$

wherein A is the phylogenetic variance-covariance matrix, scaled by σ_{Phylo} . In contrast, the additional species-level intercepts $\alpha_{Species}$ are assumed to be independent, with their variance captured by $\sigma_{Species}$:

$$\alpha_{Species} \sim \text{Normal}(0, \sigma_{Species})$$

We put the same weakly regularizing priors on the variance components:

$$\sigma_{Phylo}, \sigma_{Species} \sim \text{HalfCauchy}(0, 2)$$