Multilevel Bayesian Regression model

Axel Arango

2022-11-03

R Markdown

This is the code to perform the Multilevel Baysian Regression Model(s) to test the prevalence of the relationship between dispersal ability evolutionary rate and speciation rate across the Emberizoidea superfamily:

First, we will load the necessary packages:

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.4.4      v tibble    3.2.1
## v lubridate 1.9.3      v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(brms)
```

```
## Loading required package: Rcpp
## Loading 'brms' package (version 2.20.4). Useful instructions
## can be found by typing help('brms'). A more detailed introduction
## to the package is available through vignette('brms_overview').
##
## Attaching package: 'brms'
##
## The following object is masked from 'package:stats':
##
##      ar
```

```r
library(cmdstanr)
```

```
## This is cmdstanr version 0.7.1
## - CmdStanR documentation and vignettes: mc-stan.org/cmdstanr
## - CmdStan path: /Users/axelarango/.cmdstan/cmdstan-2.33.1
## - CmdStan version: 2.33.1
##
## A newer version of CmdStan is available. See ?install_cmdstan() to install it.
## To disable this check set option or environment variable CMDSTANR_NO_VER_CHECK=TRUE.
```

Then, the necessary data.

In this case we are only going to use the branch lengths obtained using the Cor-STRATES framework, and separated them by families.

```
dat<-read_csv("edgmeasures.csv")
```

```
## Rows: 1452 Columns: 4
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## chr (2): family, Type
## dbl (2): edge_lambda, edge_beta
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
dat2<-data.frame(family=dat$family,loglambda=log(dat$edge_lambda),logbeta=log(dat$edge_beta))
head(dat2)# this file contains the log transformed values for the speciation and phenotypic evolutionar
```

```
##         family loglambda  logbeta
## 1 Calcariidae -1.388366 1.449821
## 2 Calcariidae -1.760391 1.294828
## 3 Calcariidae -1.749554 1.344179
## 4 Calcariidae -1.650152 1.221144
## 5 Calcariidae -1.650152 1.219401
## 6 Calcariidae -1.349397 1.460212
```

Once we have loaded the data we can start to fit the Bayesian multilevel models.

First we are going to set up the initial conditions.

```
iters = 5000 #number of iterations
burns = iters / 4 # 25% as burnings
nCores = 2 #number of cores for our computer to use
nChains = 4 #number of Markov chains
```

With the initial values set up, we can fit our Bayesian model. In this model we are going to evaluate the variation of the speciation within each family.

```
fit_All_within<- brm(logbeta~loglambda  + (1 + loglambda| family),
                      data = dat2,
                      chains = nChains,
                      iter = iters,
                      cores = nCores,
                      backend = "cmdstanr")
summary(fit_All_within)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: logbeta ~ loglambda + (1 + loglambda | family)
##    Data: dat2 (Number of observations: 1452)
##   Draws: 4 chains, each with iter = 5000; warmup = 2500; thin = 1;
```

```
##          total post-warmup draws = 10000
##
## Group-Level Effects:
## ~family (Number of levels: 15)
##                          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## sd(Intercept)                0.69      0.17     0.44     1.09 1.00     3033
## sd(loglambda)                0.38      0.11     0.23     0.63 1.00     3155
## cor(Intercept,loglambda)     0.84      0.10     0.57     0.96 1.00     4045
##                          Tail_ESS
## sd(Intercept)                4303
## sd(loglambda)                4165
## cor(Intercept,loglambda)     4947
##
## Population-Level Effects:
##            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept      1.44      0.23     1.00     1.91 1.00     3043     3858
## loglambda      0.15      0.14    -0.13     0.45 1.00     3004     3781
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.28      0.01     0.27     0.29 1.00    11506     7446
##
## Draws were sampled using sample(hmc). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Here we found that the estimated standard deviation (sd) of the relationship between dispersal ability evolutionary rate and speciation rate varied substantially (Mean sd = 0.69 [0.45:1.08; 95% credible intervals])

Let's explore further to estimate the effect of the association between the rates of trait evolution and speciation for each family

First, we will extract the intercepts of species varying within family using several credible intervals

```
fam_ALL_within <- ranef(fit_All_within, probs = c(0.025, 0.05, 0.11, 0.89, 0.95, 0.975))$family
```

Then we calculate the prevalence of such family intercepts, to facilitate visualization we are going to load first a custom function and establish the colors for the intercept deviations

```
load("theme_nice.R")
barCOLS <- c(scales::alpha("red", 0.7),
             scales::alpha("darkblue", 0.7),
             scales::alpha("darkgray", 0.7))
```

Now we evaluate the effect on the association per family

```
p_family <- fam_ALL_within[, , "Intercept"] %>%
  data.frame() %>%
  rownames_to_column() %>%
  rename(family = rowname) %>%
  as_tibble() %>%
  mutate(direction = ifelse(Q11 > 0, "Positive",
                     ifelse(Q89 < 0, "Negative", "Null"))) %>%
  arrange(family) %>%
```

```r
ggplot(aes(x = family, y = Estimate, color = direction)) +
geom_pointrange(aes(ymin = Q2.5, ymax = Q97.5),
                fatten = 7, linewidth = 1.5, alpha = 0.4) +
geom_pointrange(aes(ymin = Q11, ymax = Q89),
                fatten = 10, linewidth = 2.5, alpha = 0.7) +
#coord_cartesian(ylim = c(-1, 1.5)) +
geom_hline(yintercept = 0, linewidth = 1, colour = "darkgray", linetype = "dashed") +
scale_colour_manual(values = c("Positive" = barCOLS[2],
                               "Negative" = barCOLS[1],
                               "Null" = barCOLS[3])
) +
coord_flip() +
labs(x = NULL,
     y = "Family-level effect - SD") +
theme_nice() +
theme(
  legend.position = "none",
  axis.text.x = element_text(vjust = 0.5, hjust = 0.7),
  axis.text.y = element_text(face = "italic"),
  axis.title = element_text(size = 18),
  axis.text = element_text(size = 15, colour = "black")
)

plot(p_family)
```
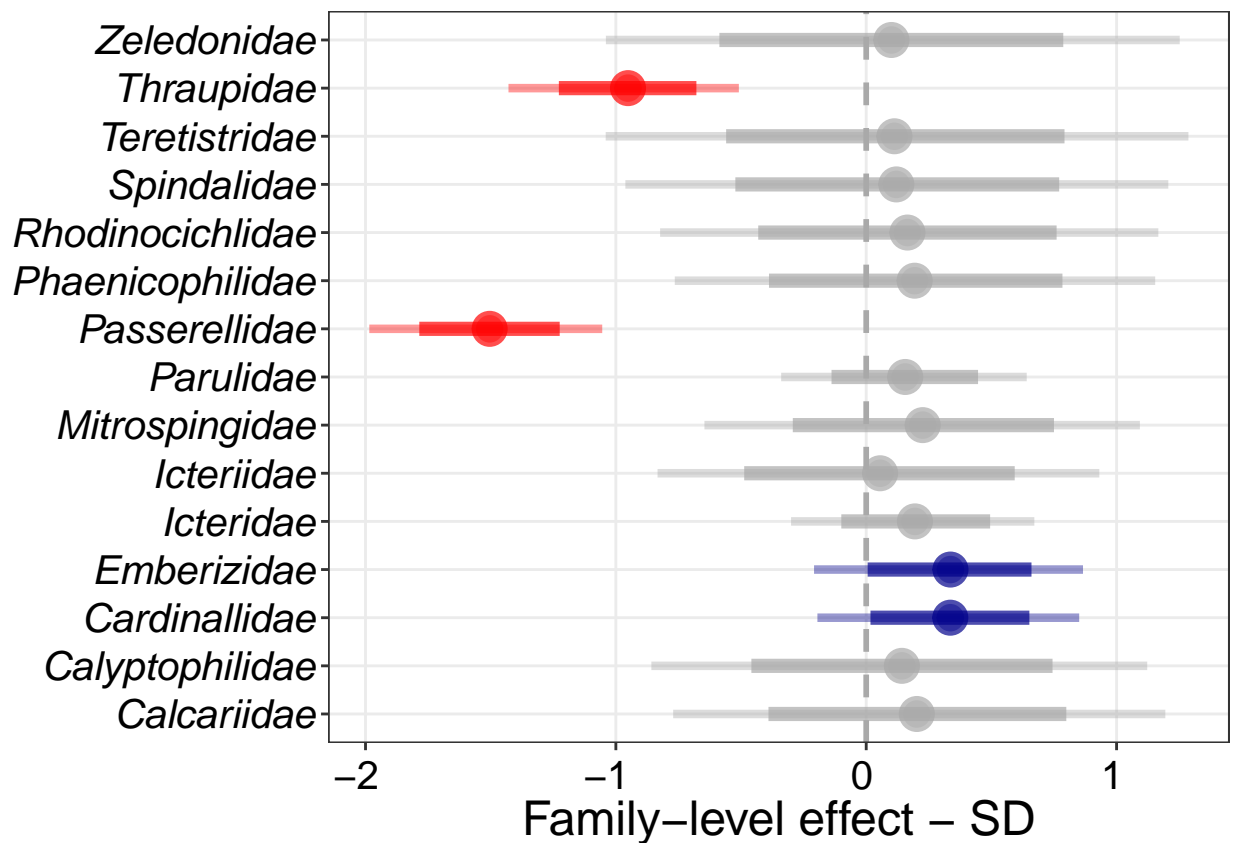

Results show that the families Thraupidae and Passerellidae have a strong negative effect on the association

between the rates of trait evolution and speciation. In contrast, Emberizidae and Cardinallidae show a positive effect—although the 95% credible intervals overlap with zero