

phylogenetic examples

Contents

Comparative example from the main text	1
Reanalysis of Balasubramaniam et al data set	5

This document contains the code to replicate the two analyses of macaque hierarchy steepness (one analysis in the main text, one in the appendix).

```
library(EloSteepness.data)
library(rstan)
library(cmdstanr)
library(ape)
```

Comparative example from the main text

```
# path to raw model code
mod1 <- system.file("extdata/phylogenetic_examples/comparative_steepness_model_1.stan",
                    package = "EloSteepness.data")
file.exists(mod1)
#> [1] TRUE
```

```
# load tree, matrices and key table
data("Primates301_nex")
data("phylo_matrices")
data("phylo_key")
```

```
# prune tree to available species
species <- unique(phylo_key$species_fixed)
tree <- keep.tip(Primates301_nex, species)
# plot(tree)
length(tree$tip.label) == length(species) # sanity check
#> [1] TRUE
```

```
# sort matrices and key so they have the same order
interaction_matrices <- phylo_matrices[sort(names(phylo_matrices))]
key <- phylo_key[order(phylo_key$datasetname), ]
all(names(interaction_matrices) == key$datasetname) # sanity check
#> [1] TRUE
```

```
# summary data for each group
specdata <- data.frame(datasetname = key$datasetname,
                      spec = gsub(" ", "_", key$species_fixed),
                      groupsize = NA,
                      n_interactions = NA,
                      dyads = NA,
```

```

prunks = NA)

for (m in seq_len(nrow(specdata))) {
  mat <- interaction_matrices[[specdata$datasetname[m]]]
  specdata$groupsize[m] <- ncol(mat)
  specdata$n_interactions[m] <- sum(mat)
  specdata$dyads[m] <- ncol(mat) * (ncol(mat) - 1) / 2
  specdata$prunks[m] <- EloRating::prunk(mat)[1]
}

```

For the sake of this vignette we reduce the number of species and also the number of interactions in the networks. We do this just because it speeds up the sampling from the model. With the full data set, this would take several hours (about six on my fairly old laptop). With the reduced data set this goes faster (but still takes a few minutes).

```

# ignore these lines if you want to run the full data set
specdata <- specdata[specdata$groupsize < 6, ]
interaction_matrices <- interaction_matrices[specdata$datasetname]
tree <- keep.tip(tree, specdata$spec)
interaction_matrices <- lapply(interaction_matrices, function(x) ceiling(x / 5))

# prepare the data set as list for sending it to Stan
xdata <- list()
xdata$n_datasets <- length(interaction_matrices)
xdata$n_ids_per_dataset <- as.integer(unlist(lapply(interaction_matrices, ncol)))
xdata$n_total_ids <- sum(xdata$n_ids_per_dataset)
xdata$n_interactions_per_dataset <- as.integer(unlist(lapply(interaction_matrices, sum)))
xdata$n_total_interactions <- sum(xdata$n_interactions_per_dataset)
xdata$interaction_index <- as.integer(seq_len(xdata$n_total_interactions))
xdata$individual_index <- as.integer(seq_len(xdata$n_total_ids))

# generate sequences
set.seed(1)
s <- lapply(interaction_matrices, randomized_sequence_from_matrix)
# global winner and loser indices
xdata$winner <- do.call("rbind", lapply(s, function(x)x$winnermat))
xdata$loser <- do.call("rbind", lapply(s, function(x)x$losermat))

xdata$index_dataset_interactions_start <- cumsum(xdata$n_interactions_per_dataset) -
  xdata$n_interactions_per_dataset + 1
xdata$index_individuals_start <- cumsum(xdata$n_ids_per_dataset) -
  xdata$n_ids_per_dataset + 1

xdata$y <- rep(1, xdata$n_total_interactions)

# phylogeny-related
n_spec <- length(tree$tip.label)
tree_mat <- vcv.phylo(tree, corr = TRUE)
species_index <- integer(nrow(specdata))
for (i in seq_len(nrow(specdata))) {
  species_index[i] <- which(colnames(tree_mat) == specdata$spec[i])
}

# for phylogeny part

```

[illegible]

```

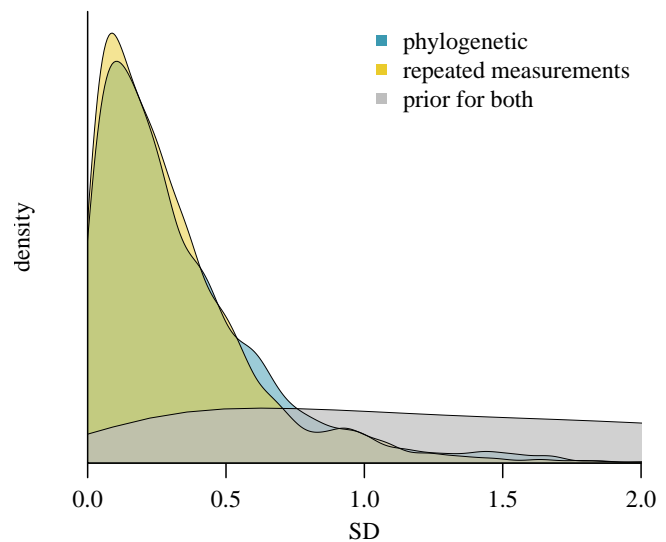
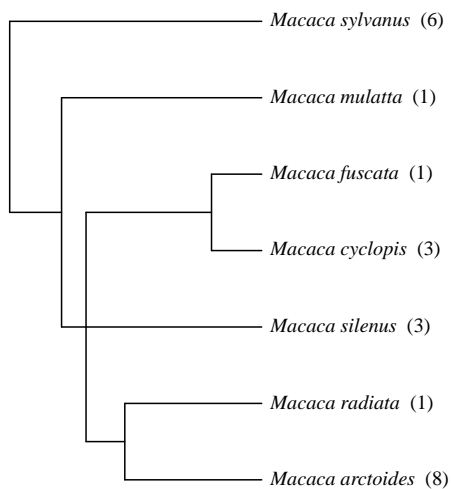
while (any(xvals < 0)) {
  xvals[xvals < 0] <- brms::rstudent_t(sum(xvals < 0), 3, 0, 2.5)
}
p3 <- density(xvals, adjust = 1)

p1 <- density(x, adjust = 1)
p2 <- density(y, adjust = 1)
yr <- c(0, max(p1$y, p2$y) * 1.05)
xr <- c(0, max(p1$x, p2$x) * 1.05)
xr[2] <- 2
plot(0, 0, type = "n", xlim = xr, ylim = yr, xaxs = "i", yaxs = "i", las = 1,
     axes = FALSE, xlab = "SD", ylab = "density")
axis(1)

polygon(c(p1$x, rev(p1$x)), c(rep(0, length(p1$y)), rev(p1$y)),
       border = NA, col = adjustcolor("#3B99B1", 0.5))
polygon(c(p2$x, rev(p2$x)), c(rep(0, length(p2$y)), rev(p2$y)),
       border = NA, col = adjustcolor("#EACB2B", 0.5))
polygon(c(p3$x, rev(p3$x)), c(rep(0, length(p3$y)), rev(p3$y)),
       border = NA, col = adjustcolor("grey", 0.7))
points(p1$x, p1$y, type = "l", lwd = 0.5)
points(p2$x, p2$y, type = "l", lwd = 0.5)
points(p3$x, p3$y, type = "l", lwd = 0.5)
box(bty = "l")

legend("topright", ncol = 1, col = c("#3B99B1", "#EACB2B", "grey"),
      legend = c("phylogenetic", "repeated measurements", "prior for both"),
      pch = 15, bty = "n")

```



Reanalysis of Balasubramaniam et al data set

```
# path to raw model code
mod2 <- system.file("extdata/phylogenetic_examples/comparative_steepness_model_2.stan",
                    package = "EloSteepness.data")
file.exists(mod2)
#> [1] TRUE

# tree data
data("Primates301_nex")

# data from Balasubramaniam et al 2012 (table 2)
sp <- c("Macaca_assamensis", "Macaca_fascicularis", "Macaca_fascicularis",
        "Macaca_fuscata", "Macaca_fuscata", "Macaca_mulatta", "Macaca_mulatta",
        "Macaca_nigra", "Macaca_radiata", "Macaca_sylvanus", "Macaca_thibetana",
        "Macaca_thibetana", "Macaca_tonkeana", "Macaca_tonkeana")
st <- c(0.65, 0.94, 0.79, 0.56, 0.92, 0.65, 0.78,
        0.49, 0.60, 0.45, 0.87, 0.80, 0.22, 0.20)
specdata <- data.frame(spec = sp,
                      steepness = st)

# prune tree to 9 macaque species
species <- unique(specdata$spec)
tree <- keep.tip(Primates301_nex, species)
# plot(tree)
length(tree$tip.label) == length(species) # sanity check
#> [1] TRUE

# prepare data set to be handed over to stan
xdata <- list()
xdata$N <- nrow(specdata)
xdata$the_steepness <- specdata$steepness

# phylogeny-related
tree_mat <- vcv.phylo(tree, corr = TRUE) # covariance matrix
n_spec <- length(tree$tip.label)
# index for species
species_index <- integer(nrow(specdata))
for (i in seq_len(nrow(specdata))) {
  species_index[i] <- which(colnames(tree_mat) == specdata$spec[i])
}

# for phylogeny part
xdata$N_phyl <- n_spec # number of levels
xdata$N_coeff_phyl <- 1 # number of pars: SD for phyl 'intercepts'
xdata$spec_index <- species_index
xdata$chol_mat <- t(chol(tree_mat)) # cholesky factor of phylogenetic correlation matrix
xdata$phyl_predictor <- rep(1, nrow(specdata)) # group-level predictor values

# for repeated measures part
xdata$N_repeated <- n_spec # the same as for the phyl part (number of levels)
# number of pars: SD for repeated measurements (species intercepts)
```

[illegible]

```

pr_phyl <- as.numeric(rstan::extract(stanfit, pars = c("prior_sd_phyl"))$prior_sd_phyl)
pr_rep <- as.numeric(rstan::extract(stanfit, pars = c("prior_sd_repeated_measures"))$prior_sd_repeated_m

# and densities of posteriors
est_phyl <- density(est_phyl, adjust = 1)
est_rep <- density(est_rep, adjust = 1)

# and for priors
pr_phyl <- density(pr_phyl[pr_phyl < 10], adjust = 1) # truncated for smoother display
pr_rep <- density(pr_rep[pr_rep < 10], adjust = 1)

# data ranges for plot
yr <- c(0, max(est_phyl$y, est_rep$y)) * 1.05
xr <- c(0, max(est_phyl$x, est_rep$x)) * 1.05

plot(0, 0, type = "n", xlim = xr, ylim = yr, xaxs = "i", yaxs = "i",
     las = 1, axes = FALSE, xlab = "SD", ylab = "density")
axis(1)

polygon(c(est_phyl$x, rev(est_phyl$x)), c(rep(0, length(est_phyl$y)), rev(est_phyl$y)),
        border = NA, col = adjustcolor("#3B99B1", 0.5))
polygon(c(est_rep$x, rev(est_rep$x)), c(rep(0, length(est_rep$y)), rev(est_rep$y)),
        border = NA, col = adjustcolor("#EACB2B", 0.5))
polygon(c(pr_phyl$x, rev(pr_phyl$x)), c(rep(0, length(pr_phyl$y)), rev(pr_phyl$y)),
        border = NA, col = adjustcolor("grey", 0.4))
polygon(c(pr_rep$x, rev(pr_rep$x)), c(rep(0, length(pr_rep$y)), rev(pr_rep$y)),
        border = NA, col = adjustcolor("grey", 0.4))

points(est_phyl$x, est_phyl$y, type = "l", lwd = 0.5)
points(est_rep$x, est_rep$y, type = "l", lwd = 0.5)
points(pr_phyl$x, pr_phyl$y, type = "l", lwd = 0.5)
points(pr_rep$x, pr_rep$y, type = "l", lwd = 0.5)

box(bty = "l")

legend("topright", ncol = 1, col = c("#3B99B1", "#EACB2B", "grey"),
      legend = c("phylogenetic", "repeated measurements", "priors"),
      pch = 15, bty = "n")

```

