

Example

Michael Li and Ben Bolker

We will reproduce the examples in [chapter 11](#) of Garamszegi (2014) using phylogenetic GLMMs based on `lme4` and `glmmTMB`.

To fit we need a random effect in the formula that includes a `(...|phylo)` term, to build the basic random effect structure multiplied by the `phyloZ` matrix (see vignette for more details on the `phyloZ` matrix).

```
library(ape)
library(Matrix)
library(lme4)
library(MASS)
library(glmmTMB)
library(coda)
library(lattice)
library(broom)
library(dplyr)
```

```
source("lme4_phylo_setup.R")
source("glmmTMB_phylo_setup.R")
```

Get data

From [chapter 11](#) of Garamszegi (2014): data are [here](#)

```
if (!file.exists("data/phylo.nex")) {
  dir.create("data")
  download.file("http://mpcm-evolution.com/OPM/Chapter11_OPM/data.zip",
               dest="data/OPM_ch11_data.zip")
  setwd("data")
  untar("OPM_ch11_data.zip")
  setwd("..")
}
phylo <- ape::read.nexus("data/phylo.nex")
```

Compute appropriate Z matrix up front, to measure speed (also reusable in a few places below):

```
system.time(phyloZ <- phylo.to.Z(phylo))
```

```
##      user  system elapsed
##    1.226    0.021    1.256
```

Result comparison with Gaussian example in chapter 11

```
datG <- read.table("data/data_simple.txt",header=TRUE)
## create observation-level grouping variable
datG$obs <- factor(seq(nrow(datG)))
datG$sp <- datG$phylo
phylo_lmm_fit <- phylo_lmm(phen~cofactor+(1|sp)
  , data=datG
  , phylonm = "sp")
```

```
, phylo = phylo
, phyloZ=phyloZ
, REML = TRUE
, control=lmerControl(check.nobs.vs.nlev="ignore",check.nobs.vs.nRE="ignore")
)
```

```
## Warning: 'rBind' is deprecated.
## Since R version 3.2.0, base's rbind() should work fine with S4 objects
```

```
print(summary(phylo_lmm_fit))
```

```
## Linear mixed model fit by REML ['lmerMod']
##
## REML criterion at convergence: 1550.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.4487 -0.5124 -0.0311  0.5663  2.2279
##
## Random effects:
##  Groups   Name                Variance Std.Dev.
##  sp      (Intercept) 207.03    14.388
##  Residual                83.74     9.151
## Number of obs: 200, groups:  sp, 200
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   39.821      6.999    5.69
## cofactor      5.175      0.136   38.06
##
## Correlation of Fixed Effects:
##              (Intr)
## cofactor -0.186
```

Similarly, fitting using glmmTMB:

```
glmmTMB_fit <- glmmTMBphylo(phen~cofactor+(1|sp)
, data=datG
, phyloZ=phyloZ
, phylonm = "sp"
, doFit=TRUE
, dispformula = ~1
, REML = FALSE
)
print(summary(glmmTMB_fit))
```

```
## Family: gaussian ( identity )
## Formula:          phen ~ cofactor + (1 | sp)
## Data: datG
##
##      AIC      BIC    logLik deviance df.resid
##  1562.0   1575.2   -777.0   1554.0     196
##
## Random effects:
##
## Conditional model:
```

```
## Groups      Name          Variance Std.Dev.
## sp          (Intercept) 198.2    14.077
## Residual                84.1     9.171
## Number of obs: 200, groups:  sp, 200
##
## Dispersion estimate for gaussian family (sigma^2): 84.1
##
## Conditional model:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  39.801      6.857   5.80 6.45e-09 ***
## cofactor      5.176      0.136  38.05 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Result comparison with Gaussian with repeated measures example in chapter 11

```
datR <- read.table("data/data_repeat.txt",header=TRUE)
datR$obs <- factor(seq(nrow(datR)))
datR <- (datR
  %>% mutate(sp = species
    , animals = phylo
    )
)
datR$spec_mean_cf <- sapply(split(datR$cofactor,datR$phylo),mean)[datR$phylo]
datR$within_spec_cf <- datR$cofactor-datR$spec_mean_cf
phylo_lmm_fit <- phylo_lmm(phen~spec_mean_cf+within_spec_cf+(1|sp) + (1|animals)
  , data=datR
  , phylonm = "sp"
  , phylo = phylo
  , phyloZ=phyloZ
  , REML = FALSE
  , control=lmerControl(check.nobs.vs.nlev="ignore",check.nobs.vs.nRE="ignore")
)

print(summary(phylo_lmm_fit))

## Linear mixed model fit by maximum likelihood ['lmerMod']
##
##      AIC      BIC    logLik deviance df.resid
## 7425.8 7455.2 -3706.9 7413.8      994
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.0237 -0.6058 -0.0005  0.5853  2.5184
##
## Random effects:
## Groups      Name          Variance Std.Dev.
## sp          (Intercept) 257.11    16.03
## animals     (Intercept) 25.30     5.03
## Residual                65.45     8.09
## Number of obs: 1000, groups:  sp, 200; animals, 200
##
```

```

## Fixed effects:
##           Estimate Std. Error t value
## (Intercept)  38.24868    7.68938   4.974
## spec_mean_cf   5.09606    0.10227  49.831
## within_spec_cf -0.05911    0.18646  -0.317
##
## Correlation of Fixed Effects:
##           (Intr) spc_m_
## spec_men_cf -0.128
## wthn_spc_cf  0.000  0.000

glmmTMB_fit2 <- glmmTMBphylo(phen~spec_mean_cf+within_spec_cf+(1|sp) + (1|animals)
  , data=datR
  , phyloZ=phyloZ
  , phylonm = "sp"
  , doFit=TRUE
  , dispformula = ~1
  , REML = FALSE
)

## Warning in condReStruc[i] <- `*vtmp*`: number of items to replace is not a
## multiple of replacement length

print(summary(glmmTMB_fit2))

## Family: gaussian ( identity )
## Formula:
## phen ~ spec_mean_cf + within_spec_cf + (1 | sp) + (1 | animals)
## Data: datR
##
##           AIC          BIC   logLik deviance df.resid
##    7425.8    7455.2  -3706.9   7413.8     994
##
## Random effects:
##
## Conditional model:
## Groups   Name                Variance Std.Dev.
## sp       (Intercept)  257.11    16.03
## animals  (Intercept)   25.30     5.03
## Residual                    65.45     8.09
## Number of obs: 1000, groups: sp, 200; animals, 200
##
## Dispersion estimate for gaussian family (sigma^2): 65.5
##
## Conditional model:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  38.24895    7.68946   4.97 6.55e-07 ***
## spec_mean_cf   5.09606    0.10227  49.83 < 2e-16 ***
## within_spec_cf -0.05911    0.18646  -0.32   0.751
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Result comparison with non-Gaussian example in chapter 11

```
dat <- read.table("data/data_pois.txt",header=TRUE)
dat$obs <- factor(seq(nrow(dat)))

dat <- dat %>% mutate(sp=phylo)
phylo_glmm_fit <- phylo_glmm(phen_pois~cofactor+(1|sp)+(1|obs)
, data=dat
, phylonm = "sp"
, family = poisson
, phylo = phylo
, phyloZ=phyloZ
, control=lmerControl(check.nobs.vs.nlev="ignore",check.nobs.vs.nRE="ignore")
)

summary(phylo_glmm_fit)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
##
##      AIC      BIC    logLik deviance df.resid
##    699.8    713.0   -345.9   691.8      196
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.0615 -0.5704 -0.3418  0.4268  5.0133
##
## Random effects:
##  Groups Name            Variance Std.Dev.
##  sp      (Intercept) 0.01224  0.1106
##  obs     (Intercept) 0.04108  0.2027
## Number of obs: 200, groups:  sp, 200; obs, 200
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.06628    0.18658  -11.07  <2e-16 ***
## cofactor     0.25022    0.01119   22.36  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr)
## cofactor -0.926
```

```
glmmTMB_fit3 <- glmmTMBphylo(phen_pois~cofactor+(1|sp)+(1|obs)
, data=dat
, family = poisson
, phyloZ=phyloZ
, phylonm = "sp"
, doFit=TRUE
, dispformula = ~1
, REML = FALSE
```

```

)

## Warning in condReStruc[i] <- `*vtmp*`: number of items to replace is not a
## multiple of replacement length

print(summary(glmTMB_fit3))

## Family: poisson ( log )
## Formula:          phen_pois ~ cofactor + (1 | sp) + (1 | obs)
## Data: dat
##
##           AIC          BIC    logLik deviance df.resid
##        699.8         713.0    -345.9    691.8      196
##
## Random effects:
##
## Conditional model:
##   Groups Name          Variance Std.Dev.
##   sp      (Intercept) 0.01224  0.1106
##   obs      (Intercept) 0.04111  0.2028
## Number of obs: 200, groups:  sp, 200; obs, 200
##
## Conditional model:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.06651    0.18748  -11.02  <2e-16 ***
## cofactor     0.25023    0.01125   22.25  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

References

Garamszegi, László Zsolt. 2014. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*. Springer.