TMA4315: Project 3

jototlan@stud.ntnu.no (10018), martigtu@stud.ntnu.no (10037)

Load data:

```
long <- read.csv("https://www.math.ntnu.no/emner/TMA4315/2020h/eliteserie.csv", colClasses = c("factor"
head(long)
```

```
##                attack            defence home goals
## 1               Molde Sandefjord_Fotball  yes     5
## 2 Sandefjord_Fotball              Molde   no     0
## 3      Stroemsgodset            Stabaek  yes     2
## 4            Stabaek      Stroemsgodset   no     2
## 5                Odd          Haugesund  yes     1
## 6          Haugesund                Odd   no     2
```

a)

We consider the model

```
library(glmmTMB)
mod <- glmmTMB(goals ~ home + (1|attack) + (1|defence),  poisson, data=long, REML=TRUE)
```

The distributional assumption on the i'th response (number of goals) is $y_i \sim \text{Poisson}(\lambda_i)$. The mean is connected to the covariates:

$$\lambda_i = \exp\left(\beta_h x_i + \gamma_{j(i)}^{\text{attack}} + \gamma_{k(i)}^{\text{defence}} + \varepsilon_i\right).$$

Here, β_h is the effect of playing home, $\gamma_{j(i)}^{\text{attack}}$ is the effect of team $j(i)$ attacking, $\gamma_{k(i)}^{\text{defence}}$ is the effect of team $k(i)$ defending, and ε_i is the error term. The distributional assumption is reasonable, since the number of goals is discrete, and one could argue that the time between goals is independent (exponentially distributed). One could, however, argue that this is not the case, for example because a team is more likely to score right after having conceded a goal.

b)

```
summary(mod)
```

```
##  Family: poisson  ( log )
## Formula:          goals ~ home + (1 | attack) + (1 | defence)
## Data: long
##
##      AIC      BIC   logLik deviance df.resid
##   1147.2   1163.1   -569.6   1139.2      382
##
## Random effects:
```

```
##
## Conditional model:
##  Groups  Name        Variance Std.Dev.
##  attack  (Intercept) 0.007478 0.08647
##  defence (Intercept) 0.016383 0.12800
## Number of obs: 384, groups:  attack, 16; defence, 16
##
## Conditional model:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.12421    0.07809   1.591    0.112
## homeyes      0.40716    0.08745   4.656 3.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ranef(mod)
```

```
## $attack
##                    (Intercept)
## BodoeGlimt         -0.036781062
## Brann               0.012026209
## Haugesund           0.011223106
## Kristiansund       -0.011367328
## Lillestroem        -0.049915996
## Molde               0.078390643
## Odd                 0.003654179
## Ranheim_TF          0.023375599
## Rosenborg           0.050622609
## Sandefjord_Fotball -0.058333079
## Sarpsborg08         0.026946364
## Stabaek            -0.026801293
## Start              -0.060500163
## Stroemsgodset       0.024556017
## Tromsoe             0.005756700
## Vaalerenga          0.007147494
##
## $defence
##                    (Intercept)
## BodoeGlimt         -0.042616090
## Brann              -0.123934761
## Haugesund          -0.061931278
## Kristiansund        0.008112432
## Lillestroem         0.030699257
## Molde              -0.036630979
## Odd                -0.052013600
## Ranheim_TF          0.062209734
## Rosenborg          -0.152631173
## Sandefjord_Fotball  0.133164228
## Sarpsborg08         0.006574064
## Stabaek             0.085376126
## Start               0.081958112
## Stroemsgodset       0.040486666
## Tromsoe            -0.009852817
## Vaalerenga          0.031030079
```

Marginal variance and intraclass covariance probit model via pmvnorm

```
#install.packages("mvtnorm")
library(mvtnorm) # to use pmvnorm()
```

Power of correct mixed vs misspecified fixed effect model vs pseudoreplication

Numerical computation of the critical value for LRT test of random slope