Design Document: Observation Weighting for GLMM

Thomas Vlasak

April 18, 2018

Abstract

This design document describes the process of implementing observation weighting funcionality for R's GLMM package. A common assumption of linear models is that each observation is equally informative of the overall process being modeled. This assumption often fails in a real world with messy data sets.

1 Goal

My ojective for the project is to modify the $glmm()$ function from R's GLMM package to take a vector of weights, square the vector, and multiply it by the predictor and response data. The weights vector must also be checked for sanity before it is multiplied.

2 Formulas

Let W be a vector of weights. Y is the response vector, X is a matrix of fixed effects, and Z is a matrix of random effects.

$$\sqrt{W} * Y = Y_w$$
$$\sqrt{W} * X = X_w$$
$$\sqrt{W} * Z = Z_w$$

3 Process

3.1 Step 1: Detect Weights Vector

Observation weighting is optional, and thus we need to establish if the vector is even there.

3.2 Step 2: Check for Sanity

- Check if object being passed is a vector (as opposed to a 1D array).
- Check the length of W aginst the length of Y
- Check if any entries of W are zero or negative
- Check if any entries of W are null
- Check if any entries of W are NA

3.3 Step 3: Save outputs

Store the unweighted X, Y, and Z in order to return them later. This will avoid confusion as to what was actually passed into the function.

3.4 Step 4: Weight Observations

Multiply the weights vector by the response vector, the fixed effects model matrix, and the random effects model matrix. Pass these arguments on to the rest of the function.

4   Tests

- Test error screening by passing function a malformed weights vector (wrong length, null values, negative values, etc.)

- Pass the function a weights = 1. Confirm the output is the same as if not passed a weights vector at all.

- Test output aginst output from $glmer()$ function from the LME4 package for a simple example. LME4 is known to not be reliable for complex models.

- Test output of weights multiplied by fixed and random effects model matrice outside of the function aginst output from the same predictors and weights passed into the function.