Homework 6: GLMMs and GEEs

your name

Due: April 21st at 11:59 PM

Several of these problems will ask you to compare GLMM and GEE fits. Your comparisons should also include differences in interpretations of the βs and the assumptions placed upon the data for each approach. This paper may be useful. Keep in mind that this reference is critical of GLMM which I attribute more to GLMM being overly dominant in the literature (at that time) than to GLMM being a modeling framework.

Problem 1: Do the following:

a) Explain why a GLMM with a Bernoulli response may have biased estimates of regression coefficients.

b) We have the same reasons as with LMMs to view χ^2 testing with some skepticism. Elaborate on these reasons.

c) Construct a bootstrap procedure that estimates the p-value corresponding to the deviance based nested model hypothesis test for the `ctsid` data analysis in the GLMM notes.

Problem 2: Show that a GLM in canonical form can be cast as a GEE. Explain your work in detail.

Problem 3: The National Youth Survey collected a sample of 11–17 year olds, 117 boys and 120 girls, asking questions about marijuana usage. The data is presented in the `potuse` dataset int he `faraway` package. Do the following:

a) Plot the total number of people falling into each usage category as it varies over time separately for each sex.

b) Condense the levels of the response into whether the person did or did not use marijuana that year. Turn the year into a numerical variable. Fit a GLMM for the now binary response with an interaction between sex and year as a predictor using Gauss-Hermite quadrature. Comment on the effect of sex.

c) Fit a reduced model without sex and use it to test for the significance of sex in the larger model.

d) Fit a model with year as a factor. Should this model be preferred to the model with year as just a linear term? Interpret the estimated effects in the year as a factor version of the model.

e) Fit your final model using PQL, Bayesian methods, and MCLA. Compare the results and discuss any inconsistencies if any arise.

f) Fit GEE version of the model and compare it to the analogous GLMM fits.

Problem 4: The `nitrofen` data in `boot` package come from an experiment to measure the reproductive toxicity of the pesticide `nitrofen` on a species of zooplankton called *Ceriodaphnia dubia*. Each animal produced three broods in which the number of live offspring was recorded. Fifty animals in total were used and divided into five batches. Each batch was treated in a solution with a different concentration of the pesticide. Do the following:

a) Plot the total number of live offspring as they vary with concentration and comment. Now plot the numbers for each brood, taking care to distinguish the different broods. Is the trend within each brood the same?

b) Fit a GLMM for the number of live offspring within each brood that varies with concentration and brood number (including an interaction). The model should take account of the dependence between observations from the same animal. Describe what the model says about how the number of live offspring change with concentration for the different broods.

c) Fit an equivalent GEE model and compare it to the GLMM result.