Lecture 10: GLMM for Clustered Data

1 Data and Model

This example uses the depression data in Section 12.1.1 of the Agresti textbook. This is the same data used in Lecture 9.

```
setwd("H:/GWU/Fall 2018/Lecture")
rm(list=ls())
data = read.csv("depression.csv", header = T)
names(data)
```

```
## [1] "case"     "severity" "treat"    "time"     "outcome"
```

```
dim(data)
```

```
## [1] 1020    5
```

```
data[1:9,]
```

```
##   case severity treat time outcome
## 1    1        0     0    0       1
## 2    1        0     0    1       1
## 3    1        0     0    2       1
## 4    2        0     0    0       1
## 5    2        0     0    1       1
## 6    2        0     0    2       1
## 7    3        0     0    0       1
## 8    3        0     0    1       1
## 9    3        0     0    2       1
```

In Lecture 9 we fit the following logistic model from Agresti using GEE:

$$logit\left[P(outcome_{it}=1)\right] = \beta_0 + \beta_1 * severity_{it} + \beta_2 * treat_{it} + \beta_3 * time_{it} + \beta_4 * treat_{it} * time_{it}$$

where subscript i denoted the subject and subscript t denotes the observations time. Now we will fit the same data using a GLMM assuming a logit link for the binomial distribution with a random intercept:

$$logit\left[P(outcome_{it}=1)\right] = \beta_0 + \beta_1 * severity_{it} + \beta_2 * treat_{it} + \beta_3 * time_{it} + \beta_4 * treat_{it} * time_{it} + u_i.$$

where u_i is the subject-specific random effect for each subject i. That is, it is a shared effect for all observations taken on a subject.

2 Logistic-Normal Random Intercept GLMM

We will fit the logistic-normal random intercept model using the *glmer* command in the *lme4* package. The *glmer* command looks a lot like the *glm* command, but requires specifying the form of the random effect in the linear predictor.

We use the syntax *(1|case)* to indicate inclusion of a random effect associated with the intercept (i.e. the first column of the design matrix - a column of ones). To include a random slope associated with time, for example, you would specify *(1+time|case)*.

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.4.4
```

```
## Loading required package: Matrix
```

```
fit.glmm = glmer(outcome ~ severity + treat + time + treat*time + (1|case), family=binomial, data=data)
summary(fit.glmm)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: outcome ~ severity + treat + time + treat * time + (1 | case)
##    Data: data
##
##      AIC      BIC   logLik deviance df.resid
##   1173.9   1203.5   -581.0   1161.9     1014
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.2849 -0.8268  0.2326  0.7964  2.0181
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  case   (Intercept) 0.003231 0.05684
## Number of obs: 1020, groups:  case, 340
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.02797    0.16406  -0.170    0.865
## severity    -1.31488    0.15261  -8.616  < 2e-16 ***
## treat       -0.05967    0.22239  -0.268    0.788
## time         0.48274    0.11566   4.174 3.00e-05 ***
## treat:time   1.01817    0.19150   5.317 1.06e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) sevrty treat  time
## severity   -0.389
## treat      -0.614 -0.005
## time       -0.673 -0.123  0.524
## treat:time  0.462 -0.121 -0.742 -0.562
```

The first line of the summary of the model indicates that *glmer* used the Laplace approximation for approximating the marginal loglikelihood. This is the default. There are MANY options for fitting a GLMM. You can change how the marginal loglikelihood is approximated, as well as the algorithm for maximizing the loglikelihood. For example, you can obtain a better approximation by using adaptive Gaussian quadrature by specifying the number of adaptive quadrature points ($nAGQ > 1$). However a more precise quadrature is implemented at the cost of longer run time.

Model fit measures such as AIC, BIC and deviance are automatically calculated and can be used for model comparisons in the usual way.

The results for the random effects indicate that the estimate of the variance for the random intercept is $\hat{\sigma}^2 = .0032$, equivalently $\hat{\sigma} = .0568$. That is, it is estimated that,

$$u_i \sim N(0, .0032).$$

This variance component is close to zero, providing evidence that there is not a cluster effect in this data. This is consistent with the GEE fit of this model in Lecture 9.

The coefficient estimates for the fixed effects β are interpreted as usual in a logistic regression model. The reported z-statistics can be treated in the usual way to carry out hypothesis tests and find confidence intervals.