BIO226 Lab Session 8: Generalized Linear Mixed Effects Models (GLMMs)

Professor: Brent Coull

TA: Yoonyoung Park

May 5, 2015

Key Points of GLMM

1. GLMMs extend the approach of linear mixed effects models to categorical data.
2. GLMMs assume heterogeneity across individuals in a subset of regression coefficients (e.g. random intercepts and slopes).
3. While Marginal Models (GEEs) focus on inferences about populations, GLMMs focus on inferences about individuals.
4. Regression parameters from GLMMs have ‘subject specific’ interpretations in terms of changes in the transformed mean response for a specific individual.

Specification of GLMM

The GLMM can be considered in 2 steps:

1. Assume conditional distribution of each Y_{ij}, given individual-specific effects b_i, belongs to exponential family with conditional mean

$$g(E[Y_{ij} \mid b_i]) = X'_{ij}\,\beta + Z'_{ij}b_i,$$

where g(.) is known link function and Z_{ij} is known design vector (a subset of X_{ij}) linking random effects b_i to Y_{ij}.

Specification of GLMM

2. The b_i are assumed to vary independently from one individual to another and $b_i \sim N(0,G)$, where G is covariance matrix for random effects.

   Note: additional assumption of "conditional independence", i.e. given b_i, the responses Y_{i1}, Y_{i2}, ..., Y_{in_i} are assumed to be mutually independent.

GLMM Example

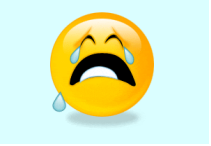
Longitudinal Binary Response to Depression Medication

Cross-classification of responses on depression at 3 times (N=Normal, A=Abnormal)

DX	TRT	NNN	NNA	NAN	NAA	ANN	ANA	AAN	AAA
Mild	Standard	16	13	9	3	14	4	15	6
Mild	New	31	0	6	0	22	2	9	0
Severe	Standard	2	2	8	9	9	15	27	28
Severe	New	7	2	5	2	31	5	32	6

Reprinted by Agresti (2002) with permission from original source (Koch et al., 1977, Biometrics)

The data stored in 'depress.txt' have already been converted into long form and contains the following 6 variables:

ID, Y (0=Abnormal  **, 1=Normal**  **)**

Severe (0=mild, 1=severe), Drug (0=standard, 1=new), Time, and Drug*Time.

SAS Code

```
DATA depress;
INFILE 'depress.txt';
INPUT id y severe drug time dt;
RUN;

DATA depress;
SET depress;
t=time;
\* create categorical time variable*\
RUN;

PROC PRINT DATA=depress(WHERE=(id=65 or
id=101));
RUN;
```

SAS Output

Obs	id	y	severe	drug	time	dt[1]	t
193	65	0	0	0	0	0	0
194	65	0	0	0	1	0	1
195	65	1	0	0	2	0	2
301	101	1	0	1	0	0	0
302	101	1	0	1	1	1	1
303	101	1	0	1	2	2	2

1. drug*time

Marginal Model (GEE) for Depression Data

$$\text{logit}\{Pr(Y_{ij} = 1)\} = \eta_{ij} = \beta_1 + \beta_2 \text{severe}_i + \beta_3 \text{drug}_i + \beta_4 \text{time}_j + \beta_5 \text{drug}_i * \text{time}_j$$

where:

- Y_{ij} = **0** subject i is abnormal in period j; **1** subject i is normal in period j
- $severe_i$ = **0** mild depression, initial diagnosis; **1** severe depression, initial diagnosis
- $drug_i$ = **0** standard; **1** new drug
- $time_j$ = **0** if baseline; **1** if time 1; **2** if time 2

and we assume:

- $Y_{ij} \sim$ Bernoulli $(e^{\eta_{ij}}/(1+e^{\eta_{ij}}))$
- $Var(Y_{ij}) = E(Y_{ij})(1 - E(Y_{ij}))$, note that $Pr(Y_{ij} = 1) = E(Y_{ij})$ because Y_{ij} is binary.
- $\log OR(Y_{ij}, Y_{ik}) = \alpha_{jk}$

SAS Code

```
PROC GENMOD DESCENDING DATA=depress;
CLASS id t;
MODEL y=severe drug time dt / DIST=binomial
  LINK=logit;
REPEATED SUBJECT=id / WITHINSUBJECT=t
  LOGOR=fullclust;
RUN;
```

SAS Output

Log Odds Ratio
Parameter Information

Parameter	Group
Alpha1	(1, 2)
Alpha2	(1, 3)
Alpha3	(2, 3)

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
intercept	-0.0261	0.1722	-0.3635	0.3114	-0.15	0.8797
severe	-1.3018	0.1449	-1.5858	-1.0177	-8.98	<0.0001
drug	-0.0556	0.2267	-0.4999	0.3886	-0.25	0.8062
time	0.4752	0.1188	0.2424	0.7081	4.00	<.0001
dt	1.0125	0.1863	0.6473	1.3777	5.43	<.0001
Alpha1	0.4052	0.2561	-0.0967	0.9071	1.58	0.1136
Alpha2	-0.2366	0.3088	-0.8419	0.3686	-0.77	0.4435
Alpha3	-0.2824	0.2844	-0.8398	0.2750	-0.99	0.3207

GLMM for Depression Data

- Consider the following GLMM:

$$\textbf{logit}\{\Pr(Y_{ij} = 1 | b_{i1})\} = \eta_{ij} = \beta_1 + \beta_2 \text{severe}_i + \beta_3 \text{drug}_i + \beta_4 \text{time}_j + \beta_5 \text{drug}_i * \text{time}_j + b_{i1}$$

  where b_{i1} is a random intercept that allows a different baseline probability of normal (vs abnormal) for each subject.

and we assume:

- $Y_{ij}|b_{i1} \sim$ Bernoulli ($e^{\eta_{ij}}/(1+e^{\eta_{ij}})$) which implies that $\text{Var}(Y_{ij}|b_{i1}) = E(Y_{ij}|b_{i1})(1 - E(Y_{ij}|b_{i1}))$. Note: $E(Y_{ij}|b_{i1}) = \Pr(Y_{ij} = 1|b_{i1})$ because Y_{ij} is binary.
- Given b_{i1}, the responses Y_{i0}, Y_{i1}, Y_{i2}, are mutually independent.
- The b_{i1} are assumed to vary independently from one individual to another and $b_{i1} \sim N(0, \sigma^2_b)$.

Question of interest: Do patient-specific changes in probability of normal differ between the two treatments?

GLIMMIX in SAS

```
proc glimmix data=depress method=quad(qpoints=100);
class id;
model y(descending) = severe drug time dt /
dist=binary link=logit s;
random intercept / subject=id;
estimate 'treatment effect, time 1' drug 1 dt 1;
estimate 'treatment effect, time 2' drug 1 dt 2;
estimate 'time trend standard treatment' time 1;
estimate 'time trend new treatment' time 1 dt 1;
run;
```

Notes:

- Treatment effect, time 1: beta3 + beta5
- Treatment effect, time 2: beta3 + 2*beta5
- Time trend standard treatment: beta4
- Time trend new treatment: beta4 + beta5

GLIMMIX in SAS

- **MODEL** statement: specifies response variable and conditional distribution of response given random effects (e.g. BINARY).
- **RANDOM** effects ~ distribution SUBJECT=variable: defines random effects (RANDOM) and variable that determines clustering of observations within an individual (SUBJECT).
- **method=quad(qpoints=100)**: GLIMMIX approximates the marginal log likelihood with adaptive quadrature. Qpoints determines the number of quadrature points in each dimension of the integral; can significantly increase computational time with increased qpoints.

Estimate Statements

Treatment effect, time 1

$\text{logit}\{Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 severe_i + \beta_3 drug_i + \beta_4 time_j + \beta_5 drug_i * time_j + b_{i1}$

For drug=0 and time=1,

$\text{logit}\{Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 severe_i + \beta_4 + b_{i1}$

For drug=1 and time=1,

$\text{logit}\{Pr(Y_{i',j} = 1 | b_{i',1})\} = \beta_1 + \beta_2 severe_{i'} + \beta_3 + \beta_4 + \beta_5 + b_{i',1}$

Thus, the difference = $\beta_3 + \beta_5$ assuming $b_{i1} = b_{i',1}$ and $severe_i = severe_{i'}$

Estimate Statements

Treatment effect, time 2

$\text{logit}\{Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 \text{severe}_i + \beta_3 \text{drug}_i + \beta_4 \text{time}_j + \beta_5 \text{drug}_i * \text{time}_j + b_{i1}$

For drug=0 and time=2,

$\text{logit}\{Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 \text{severe}_i + 2\beta_4 + b_{i1}$.

For drug=1 and time=2,

$\text{logit}\{Pr(Y_{i',j} = 1 | b_{i',1})\} = \beta_1 + \beta_2 \text{severe}_{i'} + \beta_3 + 2\beta_4 + 2\beta_5 + b_{i1}$.

Thus, the difference = $\beta_3 + 2\beta_5$ assuming $b_{i1} = b_{i',1}$ and $\text{severe}_i = \text{severe}_{i'}$

Estimate Statements

$$\text{logit}\{Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 \text{severe}_i + \beta_3 \text{drug}_i + \beta_4 \text{time}_j + \beta_5 \text{drug}_i * \text{time}_j + b_{i1}$$

Time Trend, Standard Treatment

$$\text{logit}\{Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 \text{severe}_i + \beta_4 \text{time}_{ij} + b_{i1}.$$

Time Trend, New Treatment

$$\text{logit}\{Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 \text{severe}_i + \beta_3 + (\beta_4 + \beta_5) \text{time}_{ij} + b_{i1}.$$

GLIMMIX Output

The GLIMMIX procedure is modeling the probability that y='1'.

Dimensions

G-side Cov. Parameters	1
Columns in X	5
Columns in Z per Subject	1
Subjects (Blocks in V)	340
Max Obs per Subject	3

Optimization Information

Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	6
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Not Profiled
Starting From	GLM estimates
Quadrature Points	100

GLIMMIX Output

Iteration History

Iteration	Restarts	Evaluations	Objective Function	Change
20	1	3	1161.9397523	0.00000000

Iteration History

Max Gradient

1.376E-6

Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics

-2 Log Likelihood	1161.94	
AIC (smaller is better)	1173.94	
AICC (smaller is better)		1174.02
BIC (smaller is better)	1196.91	
CAIC (smaller is better)		1202.91
HQIC (smaller is better)		1183.09

GLIMMIX Output

Covariance Parameter Estimates

Cov Parm	Subject	Standard Estimate	Error
Intercept	id	0.004331	0.1635

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-0.02795	0.1641	337	-0.17	0.8649
severe	-1.3152	0.1546	678	-8.50	<.0001
drug	-0.05970	0.2225	678	-0.27	0.7885
time	0.4828	0.1160	678	4.16	<.0001
dt	1.0184	0.1924	678	5.29	<.0001

$$\text{logit}\{Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 \text{severe}_i + \beta_3 \text{drug}_i + \beta_4 \text{time}_j + \beta_5 \text{drug}_i * \text{time}_j + b_{i1}$$

GLIMMIX Output

Estimates

Label	Estimate	Standard Error	DF	t Value
treatment effect, time 1	0.9587	0.1523	678	6.30
treatment effect, time 2	1.9771	0.2663	678	7.42
time trend standard treatment	0.4828	0.1160	678	4.16
time trend new treatment	1.5013	0.1608	678	9.34

Estimates

Label	Pr > |t|
treatment effect, time 1	<.0001
treatment effect, time 2	<.0001
time trend standard treatment	<.0001
time trend new treatment	<.0001

$$\text{logit}\{\Pr(Y_{ij} = 1 | b_{i1})\} = \beta_1 + \beta_2 \text{severe}_i + \beta_3 \text{drug}_i + \beta_4 \text{time}_j + \beta_5 \text{drug}_i * \text{time}_j + b_{i1}$$

Conclusions

- Research question: are patient-specific changes in probability of normal different between the two treatments over time? This corresponds to a testing

$$\mathbf{H_0: \beta_5 = 0}$$

- $\beta_5 = 1.0184$ (p-value<.0001). Thus, we reject H_0 of no treatment effect and conclude that there are greater patient-specific changes in probability of normal for the new treatment.

- **Treatment effect, time 1 and time 2:**
  - (Time 1) The estimated odds ratio of normal comparing a patient on the new treatment to a patient on the standard treatment with the same random intercept and severity of initial diagnosis is exp(0.9587) = 2.61. 95% CI: (exp(0.9587 – 1.96*0.1523), exp(0.9587 + 1.96*0.1523)) = (1.93, 3.52)
  - (Time 2) 7.22 (4.28, 12.19) [$e^{1.977}(e^{1.453}, e^{2.501})$]

Conclusions, continued

- **Time trend, standard treatment and new treatment:** We estimate that the odds of normal for a subject on standard treatment increases by a factor of 1.62 ($e^{0.483}$) for each time period. We estimate that the odds of normal for a subject on the new treatment increases by a factor of 4.49 ($e^{1.501}$) for each time period.

- The odds of normal of a subject with an initial diagnosis of severe depression are 0.27 ($e^{-1.315}$) times the odds of normal of a subject with mild depression and the same random intercept (i.e., a lower odds of normal).

- Note that, when we interpret the parameter estimates from the mixed model, we interpret them at the *patient level*. When we report odds ratios comparing two patients, we assume that they have the same random intercepts (i.e. the same baseline propensity for normal).