Section 13.2: Generalized Linear Mixed Models

- Observations often occur in related clusters. Phrases like *repeated measures* and *longitudinal data* get at the same thing: there's correlation among observations in a cluster.
- Chapter 12 dealt with an estimation procedure (GEE) that accounted for correlation in estimating population-averaged (marginal) effects.
- This section models cluster correlation explicitly through *random effects*, yielding a GLMM.

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$ be n_i correlated Bernoulli responses in cluster i. Associated with each repeated measure Y_{ij} are fixed (population) effects $\boldsymbol{\beta}$ and cluster-specific random effects $\mathbf{b}_i$. As usual, $\pi_{ij} = E(Y_{ij})$.

Just as in Chapter 9, the linear predictor is augmented to include random effects. Let $\text{logit}(x) = \log\{x/(1-x)\}$. Then

$$\text{logit } \pi_{ij} = \text{logit } P(Y_{ij}=1) = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{b}_i.$$

In vector/matrix terms,

$$\begin{aligned}
\begin{bmatrix} \text{logit } \pi_{i1} \\ \text{logit } \pi_{i2} \\ \vdots \\ \text{logit } \pi_{in_i} \end{bmatrix}
&= \begin{bmatrix} \text{logit } E(Y_{i1}) \\ \text{logit } E(Y_{i2}) \\ \vdots \\ \text{logit } E(Y_{in_i}) \end{bmatrix}
= \begin{bmatrix} \mathbf{x}_{i1}'\boldsymbol{\beta} + \mathbf{z}_{i1}'\mathbf{b}_i \\ \mathbf{x}_{i2}'\boldsymbol{\beta} + \mathbf{z}_{i2}'\mathbf{b}_i \\ \vdots \\ \mathbf{x}_{in_i}'\boldsymbol{\beta} + \mathbf{z}_{in_i}'\mathbf{b}_i \end{bmatrix}
= \begin{bmatrix} \mathbf{x}_{i1}' \\ \mathbf{x}_{i2}' \\ \vdots \\ \mathbf{x}_{in_i}' \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{z}_{i1}' \\ \mathbf{z}_{i2}' \\ \vdots \\ \mathbf{z}_{in_i}' \end{bmatrix} \mathbf{b}_i \\
&= \begin{bmatrix} x_{i11} & \cdots & x_{i1p} \\ x_{i21} & \cdots & x_{i2p} \\ \vdots & \vdots & \vdots \\ x_{in_i1} & \cdots & x_{in_ip} \end{bmatrix}
\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}
+ \begin{bmatrix} z_{i11} & \cdots & z_{i1q} \\ z_{i21} & \cdots & z_{i2q} \\ \vdots & \vdots & \vdots \\ z_{in_i1} & \cdots & z_{in_iq} \end{bmatrix}
\begin{bmatrix} b_{i1} \\ \vdots \\ b_{iq} \end{bmatrix} \\
&= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i
\end{aligned}$$

Note that $\pi_{ij} = \frac{e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+\mathbf{z}_{ij}'\mathbf{b}_i}}{1+e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+\mathbf{z}_{ij}'\mathbf{b}_i}}$ is a *conditional* probability (on $\mathbf{b}_i$).

Example: I ask a random sample of *the same* $n = 30$ graduate students "do you like statistics?" once a month for 4 months.

$Y_{ij} = 1$ if "yes" and $Y_{ij} = 0$ if no. Here, $i = 1, \ldots, 30$ and $j = 1, \ldots, 4$.

Covariates might include m_{ij}, the average mood of the student over the previous month ($m_{ij} = 0$ is bad, $m_{ij} = 1$ is good), the degree being sought ($d_i = 0$ doctoral, $d_i = 1$ masters), the month $t_j = j$, and p_j the number of homework problems assigned in Stat 771 in the previous month.

A GLMM might be

$$\text{logit } P(Y_{ij} = 1) = \underbrace{\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 t_j}_{\mathbf{x}'_{ij}\boldsymbol{\beta}} + \underbrace{b_{i0} + b_{i1} t_j}_{\mathbf{z}'_{ij}\mathbf{b}_i}.$$

This model assumes that log-odds of liking statistics changes linearly in time, holding all else constant.

Alternatively, we might fit a quadratic instead or treat time as categorical.

Here, $\mathbf{b}_i$ represents a student's *a priori* disposition-trend towards statistics.

Does overall mood affect one's disposition toward statistics?

In a given month t_j, for a given population of individuals with the same trend $\mathbf{b}_i$, the difference in log odds for good versus bad moods is

$$(\beta_0+\beta_1(1)+\beta_2 d_i+\beta_3 p_j+\beta_4 t_j+b_{i0}+b_{i1}t_j)-(\beta_0+\beta_1(0)+\beta_2 d_i+\beta_3 p_j+\beta_4 t_j+b_{i0}+b_{i1}t_j)=\beta_1.$$

So e^{β_1} is how the odds of liking statistics changes with mood.

We are conditioning on individual i, or *the subpopulation of all individuals with predisposition* $\mathbf{b}_i$; i.e. everyone "like" individual i in terms of liking statistics (over time) to begin with.

How are e^{β_2}, e^{β_3}, and e^{β_4} interpreted here?

The random effects are assumed to come from (in general) a multivariate normal distribution

$$\mathbf{b}_1, \ldots, \mathbf{b}_n \overset{iid}{\sim} N_q(\mathbf{0}, \mathbf{\Sigma}).$$

The covariance $\text{cov}(\mathbf{b}_i) = \mathbf{\Sigma}$ can have special structure, e.g. exchangeability, AR(1), or be unstructured – usually only unstructured makes sense. The free elements of $\mathbf{\Sigma}$ are estimated along with $\boldsymbol{\beta}$.

The $\mathbf{b}_i$ can account for heterogeneity caused by omitting explanatory variables.

Connection between marginal and conditional models

In the GEE approach, the marginal means are explicitly modeled:

$$\pi_{ij} = E(Y_{ij}) = \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta}}},$$

and correlation among $(Y_{i1}, \ldots, Y_{in_i})$ is accounted for in the estimation procedure.

The conditional approach models the means conditional on the random effects:

$$E(Y_{ij}|\mathbf{b}_i) = \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{b}_i}}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{b}_i}}.$$

The corresponding marginal mean is given by

$$E(Y_{ij}) = \int_{\mathbb{R}^q} \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{b}_i}}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\mathbf{b}_i}} f(\mathbf{b}_i; \boldsymbol{\Sigma}) d\mathbf{b}_i = \boxed{???}$$

- In epi studies, often want to compare disease prevalence across groups. Then it's of interest to compute marginal odds ratios and compare them.
- The more variability that's accounted for in the conditional model, the more we can "focus in" on the conditional effect of covariates. This is true in any situation where we block. This has the effect enlarging $\hat{\beta}$ estimates under a conditional model.
- When correlation is small, independence is approximately achieved, and marginal and conditional modeling yield similar results.
- GLMMs are being increasingly used, in part due to the availability of standard software to fit them!
- Bayesian approach is also natural here.

GLIMMIX

This is a new SAS procedure that fits GLMM's.

- The procedure had been available as a macro for some time.
- GLIMMIX essentially extends the MIXED procedure to GLM's, and in fact iteratively calls MIXED when fitting GLMM's.
- Only normal random effects are allowed.
- GLIMMIX uses an approximation when fitting models. The approximation in effect replaces the intractable integral $\boxed{???}$ with a simple linear Taylor's expansion. It's crude, but works and is fast. See pp. 119–125 in SAS' GLIMMIX documentation for details on "Pseudo-likelihood Estimation Based on Linearization."

- GLIMMIX can also fit marginal models allowing for correlation within a cluster (like GENMOD), but uses a different estimation method than GENMOD with the repeated statement. Then $\mathbf{R}$ has structure, e.g. exchangeability (called compound symmetry here), AR structure, spatial structures, and others found in PROC MIXED.

- The learning curve is steep, although it's nice to be aware of alternative fitting procedures if necessary!

- Let's fit a GLMM to the wheezing (Six Cities) data.

$$\text{logit } \pi_{ij} = \beta_0 + \beta_1 I\{c_i = \text{Kingston}\} + \beta_2 I\{s_{ij} = 0\} + \beta_3 I\{s_{ij} = 1\} + t_j \beta_4 + b_{i0} + b_{i1} t_j.$$

```
proc glimmix data=six method=laplace;
 class case city smoke;
 model wheeze = city smoke age / dist=bin link=logit s;
 random  int age / subject=case type=un;
run;

----------------------------------------------------------------------------------------------------------

                                    The GLIMMIX Procedure

                                      Model Information
                      Data Set                      WORK.SIX
                      Response Variable             wheeze
                      Response Distribution         Binomial
                      Link Function                 Logit
                      Variance Function             Default
                      Variance Matrix Blocked By    case
                      Estimation Technique          Maximum Likelihood
                      Likelihood Approximation      Laplace
                      Degrees of Freedom Method     Containment

                                   Class Level Information
                  Class    Levels    Values

                  case         32    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
                                     19 20 21 22 23 24 25 26 27 28 29 30 31 32
                  city          2    kingston portage
                  smoke         3    0 1 2

                        Number of Observations Read         128
                        Number of Observations Used         100
```

Dimensions

G-side Cov. Parameters	3
Columns in X	7
Columns in Z per Subject	2
Subjects (Blocks in V)	32
Max Obs per Subject	4

Optimization Information

Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	8
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Not Profiled
Starting From	GLM estimates

Iteration History

Iteration	Restarts	Evaluations	Objective Function	Change	Max Gradient
0	0	4	117.81153662	.	104.5189
1	0	8	116.98455729	0.82697933	65.599
2	0	4	115.85922461	1.12533268	3.707623
			...et cetera...		
34	0	3	115.37456051	0.00000076	0.029757
35	0	3	115.37456049	0.00000003	0.000869

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

Fit Statistics

-2 Log Likelihood	115.37
AIC (smaller is better)	131.37
AICC (smaller is better)	132.96
BIC (smaller is better)	143.10
CAIC (smaller is better)	151.10
HQIC (smaller is better)	135.26

Fit Statistics for Conditional Distribution

-2 log L(wheeze | r. effects)	82.07
Pearson Chi-Square	60.59
Pearson Chi-Square / DF	0.61

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	case	46.5032	95.6792
UN(2,1)	case	-4.2813	8.7969
UN(2,2)	case	0.4011	0.8183

Solutions for Fixed Effects

Effect	city	smoke	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			1.5066	3.8219	30	0.39	0.6962
city	kingston		0.3287	0.6789	38	0.48	0.6310
city	portage		0	.	.	.	.
smoke		0	-0.4994	0.8192	38	-0.61	0.5457
smoke		1	-0.9590	0.7697	38	-1.25	0.2204
smoke		2	0	.	.	.	.
age			-0.2095	0.3460	27	-0.61	0.5498

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
city	1	38	0.23	0.6310
smoke	2	38	0.82	0.4475
age	1	27	0.37	0.5498

Nothing is significant...even when weeding out nonsignificant predictors. Just a “bad” data set: no signal, all noise.

Interpretation here is different though...it's *within child*. Marginal approach of last time is a population-averaged approach, i.e. for all kids. Here's the marginal (GEE) output:

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		1.4511	2.1023	-2.6694	5.5715	0.69	0.4901
city	kingston	0.3407	0.4754	-0.5909	1.2724	0.72	0.4735
city	portage	0.0000	0.0000	0.0000	0.0000	.	.
smoke	0	-0.4426	0.5613	-1.5428	0.6575	-0.79	0.4303
smoke	1	-0.3367	0.7138	-1.7356	1.0623	-0.47	0.6372
smoke	2	0.0000	0.0000	0.0000	0.0000	.	.
age		-0.2005	0.2097	-0.6116	0.2105	-0.96	0.3390

The effects are *about the same or smaller*; also the standard errors are all less, some by as much as two thirds smaller. Often this is reversed in terms of standard errors. With only a random intercept, the standard errors from the GLMM decrease.