Chapter 13: Random Effects Models & more

Dipankar Bandyopadhyay

Department of Biostatistics,
Virginia Commonwealth University

BIOS 625: Categorical Data & GLM

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 a generalized estimation equation procedure (GEE) that accounted for correlation in estimating population-averaged (marginal) effects.

- This chapter models cluster correlation explicitly through *random effects*, yielding a generalized linear mixed effects models (GLMM).

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

In a GLMM the linear predictor is augmented to include random effects:

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i.$$

for logistic regression, this is

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

Note that conditional on $\mathbf{u}_i$,

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

Example: A random sample of *the same $n = 30$ graduate students* were asked "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 PubH 7407 in the previous month.
- A GLMM might be

$$\text{logit } P(Y_{ij} = 1) = \beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i.$$

- 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, u_i represents a student's *a priori* disposition towards statistics.

- Let's compare month $j + 1$ to month j for individual i, holding all else (m, d, and p) constant. The difference in log odds is

$$(\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4(j+1) + u_i) - (\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i) = \beta_4.$$

- Not holding everything constant we get

$$(\beta_0 + \beta_1 m_{i,j+1} + \beta_2 d_i + \beta_3 p_{j+1} + \beta_4(j+1) + u_i) - (\beta_0 + \beta_1 m_{ij} + \beta_2 d_i + \beta_3 p_j + \beta_4 j + u_i)$$

$$= \beta_1(m_{i,j+1} - m_{ij}) + \beta_3(p_{j+1} - p_j) + \beta_4.$$

- Either way, we are conditioning on individual i, or *the subpopulation of all individuals with predisposition u_i*; i.e. everyone "like" individual i to begin with.
- How are e^{β_1}, 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

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

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

- The $\mathbf{u}_i$ can account for heterogeneity caused by omitting explanatory variables.
- They can also explicitly model overdispersion, e.g.

$$Y_i \sim \text{Pois}(\lambda_i), \ \log \lambda_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i, \ u_i \overset{iid}{\sim} N(0, \sigma^2).$$

Logit model for binary matched pairs

Example: PV data.

	2008 Election		
2004 Election	Democrat $Y = 1$	Republican $Y = 2$	Total
Democrat $X = 1$	$n_{11} = 175$	$n_{12} = 16$	$n_{+1} = 191$
Republican $X = 2$	$n_{21} = 54$	$n_{22} = 188$	$n_{+2} = 242$
Total	$n_{1+} = 229$	$n_{2+} = 204$	$n_{++} = 433$

Recall $j = 1, 2$ denotes a binary covariate; for the PV data it's time.

$$\text{logit } P(Y_{ij} = 1) = \alpha + u_i + \beta I\{j = 2008\}.$$

Here, e^{β} is a cluster-specific odds ratio. We further assume $u_i \stackrel{iid}{\sim} N(0, \sigma^2)$.

- Recall that when fitting this type of data using marginal model in GENMOD, $\hat{\beta} = log[(n_{+1}/n_{+2})/(n_{1+}/n_{2+})]$
  $= log[(229/204)/(191/242)] = 0.352$ and so $e^{\hat{\beta}} = 1.42$.
- Recall that when fitting a conditional logistic regression, a close form estimate of β exists $\hat{\beta} = \log(n_{21}/n_{12}) = 1.22$ and and so $e^{\hat{\beta}} = 3.38$, $\hat{\sigma}(\hat{\beta}) = \sqrt{1/n_{21} + 1/n_{12}} = 0.28$.
- When the sample log odds ratio $log(\frac{n_{11}n_{22}}{n_{21}n_{12}}) \geq 0$, $\hat{\sigma} > 0$, $\hat{\beta} = \log(n_{21}/n_{12})$.
- When the sample log odds ratio $log(\frac{n_{11}n_{22}}{n_{21}n_{12}}) < 0$, $\hat{\sigma} = 0$, $\hat{\beta} = log[(n_{+1}/n_{+2})/(n_{1+}/n_{2+})]$.
- Although explicit forms exist, we'll fit this in SAS using two different data structures for illustrative purposes.

In the following code, first is conditional logistic approach from Chapter 11, second is marginal GEE logistic approach from Chapter 12.

```
data Data1;
 do ID=1 to 175; dem=1; time=0; output; dem=1; time=1; output; end;
 do ID=176 to 191; dem=1; time=0; output; dem=0; time=1; output; end;
 do ID=192 to 245; dem=0; time=0; output; dem=1; time=1; output; end;
 do ID=246 to 433; dem=0; time=0; output; dem=0; time=1; output; end;
* conditional logistic regression ;
proc logistic data=Data1; strata ID;
 model dem(event='1')=time;
* marginal inference , appropriately accounting for within-subject correlation ;
proc genmod data=Data1 descending; class ID;
 model dem=time / link=logit dist=bin;
 repeated subject=ID / corr=exch corrw; run;
```

Here is the GLMM approach of Chapter 13 with $u_i \overset{iid}{\sim} N(0, \sigma^2)$:

```
proc nlmixed data=Data1 maxiter=100 method=GAUSS qpoints=100;
 parms beta0=-1.0 beta1=1.0 sigma=5.0;
 eta = beta0+beta1*time+u;
 pi = exp(eta)/(1+exp(eta));
 model dem ~ binary(pi);
 random u ~ normal(0,sigma*sigma) subject=ID out=empBayesUA;
 * OUT requests an output data set containing empirical Bayes estimates of
 * the random effects and their approximate standard errors of prediction;
 estimate 'subject-specific OR of 08/04' exp(beta1);
data matched; input case occasion response count @@; datalines;
1 0 1 175    1 1 1 175    2 0 1 16    2 1 0 16
3 0 0 54     3 1 1 54     4 0 0 188   4 1 0 188
;
proc nlmixed data=matched maxiter=100 method=GAUSS qpoints=100;
 eta = beta0 + beta1*occasion + u;
 p = exp(eta)/(1 + exp(eta));
 model response ~ binary(p);
 random u ~ normal(0, sigma*sigma) subject = case out=empBayesUB;
 replicate count;run;
```

Output from the first fit:

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	−0.8170	0.3427	432	−2.38	0.0176	0.05	−1.4906	−0.1434	0.00018
beta1	1.2164	0.2846	432	4.27	<.0001	0.05	0.6570	1.7759	0.000308
sigma	5.2169	0.6976	432	7.48	<.0001	0.05	3.8458	6.5881	0.000054

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower
subject−specific OR of 08/04	3.3751	0.9607	432	3.51	0.0005	0.05	1.4869

Read through **13.1.5**: random effects versus conditional approach.

NLMIXED option METHOD=value specifies the method for approximating the integral of the likelihood over the random effects. Valid values are as follows:

- FIRO specifies the first-order method of Beal and Sheiner (1982). When using METHOD=FIRO, you must specify the NORMAL distribution in the MODEL statement and you must also specify a RANDOM statement.
- GAUSS specifies adaptive Gauss-Hermite quadrature (Pinheiro and Bates 1995). You can prevent the adaptation with the NOAD option or prevent adaptive scaling with the NOADSCALE option. This is the default integration method.
- HARDY specifies Hardy quadrature based on an adaptive trapezoidal rule. This method is available only for one-dimensional integrals; that is, you must specify only one random effect.
- ISAMP specifies adaptive importance sampling (Pinheiro and Bates 1995). You can prevent the adaptation with the NOAD option or prevent adaptive scaling with the NOADSCALE option. You can use the SEED= option to specify a starting seed.

Alternative coding using GLIMMIX with an adaptive Gauss-Hermite quadrature approximation to marginal integrated likelihood:

```
proc glimmix data=Data1 METHOD=QUAD (qpoints=100); class ID;
 model dem(event='1') = time /s link=logit  dist =bin;
 random INTERCEPT/subject=ID;
run;
```

Output from GLIMMIX:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	ID	27.2157	7.2788

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	−0.8170	0.3427	432	−2.38	0.0176
time	1.2164	0.2846	432	4.27	<.0001

NOTE:

- Default METHOD = option for GLIMMIX is RSPL.
- METHOD=QUAD (LAPLACE, RSPL, MSPL, RMPL, MMPL). Estimation methods ending in "PL" are pseudo-likelihood techniques. The first letter of the METHOD= identifier determines whether estimation is based on a residual likelihood ("R") or a maximum likelihood ("M"). The second letter identifies the expansion locus for the underlying approximation. Pseudo-likelihood methods for generalized linear mixed models can be cast in terms of Taylor series expansions (linearizations) of the GLMM. The expansion locus of the expansion is either the vector of random effects solutions ("S") or the mean of the random effects ("M"). The expansions are also referred to as the "S"ubject-specific and "M"arginal expansions. The abbreviation "PL" identifies the method as a pseudo-likelihood technique.

A special, often-used case of the GLMM.
The logistic normal model is given by:

$$\text{logit } P(Y_{ij} = 1|u_i) = \mathbf{x}'_{ij}\boldsymbol{\beta} + u_i, \quad u_i \overset{iid}{\sim} N(0, \sigma^2).$$

When $\sigma = 0$ we get the standard logistic regression model, when $\sigma > 0$ we account for extra heterogeneity in clustered responses (each i is a cluster with it's own random u_i).

Connection between marginal and conditional models

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

$$\mu_{ij} = E(Y_{ij}) = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta}),$$

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

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

$$E(Y_{ij}|\mathbf{u}_i) = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i).$$

The corresponding marginal mean is given by

$$E(Y_{ij}) = \int_{\mathbb{R}^q} g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i)f(\mathbf{u}_i; \boldsymbol{\Sigma})d\mathbf{u}_i.$$

- In general, this is a complicated function of $\boldsymbol{\beta}$.

- However for the logistic-normal model when σ is "small," we obtain (not obvious)

$$E(Y_{ij}) \approx \exp(c\mathbf{x}'_{ij}\boldsymbol{\beta})/[1 + \exp(c\mathbf{x}'_{ij}\boldsymbol{\beta})],$$

where $c = 1/\sqrt{1 + 0.346\sigma^2}$ (See Zeger, Liang and Albert 1988). The *marginal odds* change by approximately $e^{c\beta_s}$ when x_{ijs} is increased by unity.

- Because $c < 1$, the marginal effect is smaller than the conditional effect, reflecting that we are averaging with respect to the population. Note that the larger σ is, the more subject-to-subject variability there is, and the *smaller* the averaged effect $\hat{c}\hat{\beta}_s$ becomes. See Page 496, **Figure 13.1**.
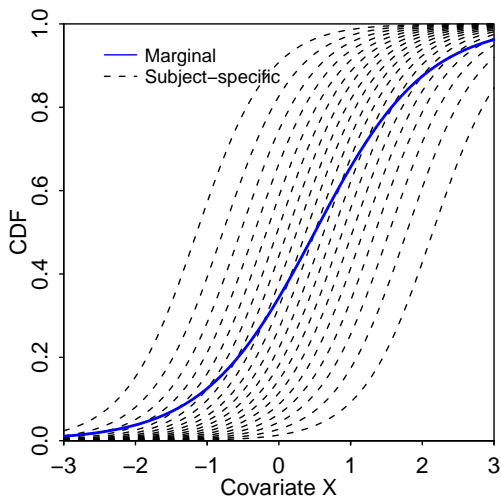
Figure : Logistic Random-intercept Model: Subject-specific vs Marginal curves

- From problem 13.25: The GLMM for binary data using probit link function is $\Phi^{-1}[P(Y_{ij} = 1|\mathbf{u}_i)] = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_i$, one can easily prove that $\Phi^{-1}[P(Y_{ij} = 1)] = \mathbf{x}'_{ij}\boldsymbol{\beta}(1 + \mathbf{z}'_{ij}\boldsymbol{\Sigma}\mathbf{z}_{ij})^{-1/2}$. In the univariate random intercept case, it means that the marginal effect equals to that from the GLMM divided by $\sqrt{1 + \sigma^2}$.

PV data, a final look. Here, $\hat{c} = 1/\sqrt{1 + 0.346(5.22)^2} = 0.31$. Then $e^{1.22(0.31)} = 1.46$. Recall that the GEE approach yields $e^{0.35} = 1.42$; a very good approximation! Also recall that the conditional approach yielded $e^{1.22} = 3.38$. Annotated output:

The LOGISTIC Procedure
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
time	1	1.2164	0.2846	18.2627	<.0001

The GENMOD Procedure

Exchangeable Working Correlation

Correlation   0.6910151517

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	−0.2367	0.0968	−0.4264	−0.0470	−2.45	0.0145
time	0.3523	0.0761	0.2031	0.5014	4.63	<.0001

The NLMIXED Procedure
Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta0	−0.8170	0.3427	432	−2.38	0.0176	0.05	−1.4907	−0.1434	−0.00012
beta1	1.2164	0.2846	432	4.27	<.0001	0.05	0.6569	1.7758	−0.00001
sigma	5.2169	0.6976	432	7.48	<.0001	0.05	3.8457	6.5880	−0.00005

Text comments:

- In epi studies, often want to compare disease prevalence across groups. Then it's of interest to compute marginal odds ratios and compare them.
- We did not discuss MLE approach to marginal models; uses a huge multinomial distribution; can be unstable. See text.
- Direction and significance of effects usually the same across marginal/conditional models (e.g. PV data).

- 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 $\hat{\beta}_s$ 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.