

BIOS 755: Generalized Linear Mixed Models

Alexander McLain

February 23, 2023

Outline

Introduction

Interpretations of coefficients

Estimation

Introduction

- ▶ So far, we have discussed marginal models for longitudinal data and the use of generalized estimating equations to fit these models.
- ▶ To fit marginal models, we made some assumptions about the marginal distribution at each time point, and estimated a matrix of correlation coefficients linking repeated observations of the same subject.
- ▶ In specifying the marginal expectations and variances and the covariance matrices, we were not fully specifying the joint distribution of the repeated measurements.
- ▶ Thus, estimation using GEE is not likelihood-based.

Incorporating Random Effects into GLM

- ▶ The basic premise is that we assume that there is natural heterogeneity across individuals in a subset of the regression coefficients.
- ▶ That is, a subset of the regression coefficients (e.g. intercepts) are assumed to vary across individuals according to some distribution.
- ▶ Then, conditional on the random effects, it is assumed that the responses for a single individual are independent observations from a distribution belonging to the exponential family.

Generalized Linear Mixed Models

- ▶ For non-Normal responses, Y_i , the generalized linear mixed model (GLMM) can also be considered in two steps:

1. Assumes that the conditional distribution of each Y_{ij} , given \mathbf{b}_i , belongs to the exponential family with conditional mean,

$$g\{E(Y_{ij}|\mathbf{b}_i)\} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

where $g(\cdot)$ is a known link function.

2. The \mathbf{b}_i are assumed to vary independently from one individual to another and $\mathbf{b}_i \sim N(0, \mathbf{G})$.
- ▶ There is an additional assumption of 'conditional independence,' i.e., given \mathbf{b}_i , the responses $Y_{i1}, Y_{i2}, \dots, Y_{ip}$ are mutually independent.

Examples of GLMM

- ▶ Binary logistic model with random intercepts:

$$\text{logit}\{E(Y_{ij}|\mathbf{b}_i)\} = \mathbf{X}_i\boldsymbol{\beta} + b_i$$

with $b_i \sim N(0, \sigma^2)$.

- ▶ Random coefficients Poisson regression model:

$$\log\{E(Y_{ij}|\mathbf{b}_i)\} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

with $X_{ij} = Z_{ij} = (1 \ t_{ij})'$ (i.e. random intercepts and random slopes) and $\mathbf{b}_i \sim N(0, \mathbf{D})$.

Recall Linear Mixed (Effects) Models

- ▶ In the mixed model

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i$$

- ▶ We have

$$E(\mathbf{Y}_i|\mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

$$E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}$$

- ▶ Here,

$$E(\mathbf{Y}_i|\mathbf{b}_i) = g^{-1}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i)$$

$$E(\mathbf{Y}_i) \neq g(\mathbf{X}_i\boldsymbol{\beta})$$

Interpretation of GLMM

- For example, for a logistic model

$$E(\mathbf{Y}_i | \mathbf{b}_i) = \frac{\exp(\beta_0 + \beta_1 X_i + b_{i0})}{1 + \exp(\beta_0 + \beta_1 X_i + b_{i0})}$$

but

$$E(\mathbf{Y}_i) \neq \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}.$$

To use the OR interpretation, we have to condition on the random effect.

Interpretation of GLMM

- ▶ Mixed effects models are most useful when the scientific objective is to make inferences about **individuals** rather than the population averages.
- ▶ The interpretation of all β coefficients is **given** the random effects.
 - ▶ Who has the same random effect?
- ▶ Main focus is on the individual and the influence of covariates on the individual.
- ▶ The increase in the probability of a heart attack when aging from 40 to 50 instead of the increase in probability of a heart attack between 40 year olds and 50 year olds

Estimation of GLMM

- ▶ Unlike GEE the GLMM does assume a full joint probability function, and maximum likelihood can be used.
- ▶ The joint probability density function is given by:

$$f(\mathbf{Y}_i, \mathbf{b}_i | \mathbf{X}_i) = f(\mathbf{Y}_i | \mathbf{X}_i, \mathbf{b}_i) f(\mathbf{b}_i)$$

- ▶ From this we get the marginal or integrated density function:

$$f(\mathbf{Y}_i | \mathbf{X}_i) = \int f(\mathbf{Y}_i | \mathbf{X}_i, \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i,$$

and the likelihood $\prod_{i=1}^n f(\mathbf{Y}_i | \mathbf{X}_i)$.

- ▶ ML estimation of β and \mathbf{D} is based on the marginal or integrated likelihood of the data (obtained by averaging over the distribution of the unobserved random effects, \mathbf{b}_i).

Estimation of GLMM

- ▶ Estimation using maximum likelihood involves a two-step procedure:
- 1. For ML estimation of β and D simple analytic solutions are rarely available, and numerical or Monte Carlo integration techniques are required.
- 2. Given estimates of β and D , the random effects can be predicted by something called the **posterior mean**:

$$\hat{b}_i = E(b_i | Y_i; \hat{\beta}, \hat{D})$$

Computational issues

- ▶ **A potential limitation of generalized linear mixed models is their computational burden.** Because, in general, there is no simple closed-form solution for the marginal likelihood, numerical integration techniques are required.
- ▶ Maximum likelihood estimation has been implemented in standard statistical software like `lme` in R or PROC GLIMMIX in SAS.
- ▶ For both, there are some different numerical estimation procedures to consider.