BIOS 755: Generalized Linear Mixed Models

Alexander McLain

Outline

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 $\boldsymbol{b}_i$, belongs to the exponential family with conditional mean,

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

     where $g(\cdot)$ is a known link function.
  2. The $\boldsymbol{b}_i$ are assumed to vary independently from one individual to another and $\boldsymbol{b}_i \sim N(0, \boldsymbol{G})$.
- There is an additional assumption of 'conditional independence,' i.e., given $\boldsymbol{b}_i$, the responses $Y_{i1}, Y_{i2}, \ldots, Y_{ip}$ are mutually independent.

Examples of GLMM

- Binary logistic model with random intercepts:

$$\text{logit}\{E(Y_{ij}|\boldsymbol{b}_i)\} = \beta_0 + \beta_1 t_{ij} + \beta_2 Sex_i + b_i$$

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

- Random coefficients Poisson regression model:

$$\log\{E(Y_{ij}|\boldsymbol{b}_i)\} = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij},$$

  i.e. random intercepts and random slopes, and $\boldsymbol{b}_i = (b_{i0} \ b_{i1})' \sim N(0, \boldsymbol{D})$.

Recall Linear Mixed (Effects) Models

- In the mixed model
$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$$
- We have
$$E(\boldsymbol{Y}_i|\boldsymbol{b}_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$$
$$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$
- Here,
$$E(\boldsymbol{Y}_i|\boldsymbol{b}_i) = g^{-1}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i)$$
$$E(\boldsymbol{Y}_i) \neq g(\boldsymbol{X}_i\boldsymbol{\beta})$$

Interpretation of GLMM

- For example, for a logistic model

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

but

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

To use the OR interpretation, be 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:

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

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

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

  and the likelihood $\prod_{i=1}^{n} f(\boldsymbol{Y}_i|\boldsymbol{X}_i)$.
- ML estimation of $\boldsymbol{\beta}$ and $\boldsymbol{D}$ is based on the marginal or integrated likelihood of the data (obtained by averaging over the distribution of the unobserved random effects, $\boldsymbol{b}_i$).

Estimation of GLMM

- Estimation using maximum likelihood involves a two-step procedure:

1. For ML estimation of $\boldsymbol{\beta}$ and $\boldsymbol{D}$ simple analytic solutions are rarely available, and numerical or Monte Carlo integration techniques are required.
2. Given estimates of $\boldsymbol{\beta}$ and $\boldsymbol{D}$, the random effects can be predicted my something called the **posterior mean**:

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

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 `glmer` in `R` or `PROC GLIMMIX` in `1SAS`.
- For both, there are some different numerical estimation procedures to consider.
- In `R`, there is a new package `GLMMadaptive`, which is made by some of the top people in computational statistics.