STAT347: Generalized Linear Models
Lecture 13

Today's topics: Chapters 9.1-9.3

- Correlated samples /responses in GLM

- Normal linear mixed effect models (LMM)

  - Random intercept and random slope models
  - Hierarchical models for a multi-level design
  - Model estimation: MLE, REML and BLUP

1 Modeling correlated responses

For the responses: $y_1, y_2, \cdots, y_n$, we have assumed independence, but some samples may be correlated. Examples:

- Kids of one mom, longitudinal data for one individual

- Students in the same classroom with many classrooms

- Multiple individuals measured in one day with many different days

Form of the data: there are $i = 1, 2, \cdots, n$ groups (individuals / classrooms / days), and each of them has $s = 1, 2, \cdots, d_i$ samples. The response is denoted as y_{is} with its covariates x_{is}.

Generalized linear mixed model (GLMM):

$$g(\mu_{is}) = X_{is}^T \beta + Z_{is}^T u_i$$

where X_{is} and Z_{is} are observed, and u_i are i.i.d. random variables across i following some unknown distribution F.

- The responses $(y_{i1}, \cdots, y_{id_i})$ within each group i are correlated because they share the same latent random variable u_i

- $Z_{is}^T u_i$ models that the influences of u_i on different samples depend on some covariate Z_{is}

- Two perspectives on GLMM:

  - We use GLMM to model dependence structures among samples
  - We treat u_i as an unknown coefficient of Z_{is}. We add prior on u_i / make u_i random to borrow information across i (so that we only need to estimate unknown parameters in F instead of estimating each u_i).

- The most common model: $Z_{is} = 1$ assuming group members share a common group-level effect.

2 Normal linear mixed models

$$y_{is} = X_{is}^T \beta + Z_{is}^T u_i + \epsilon_{is}$$

- β is a length p vector, and is for fixed effects

- $u_i \overset{i.i.d.}{\sim} N(0, \Sigma_u)$ can be a vector when Z_{is} is a vector. It models the random effects

- $\epsilon_{is} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$ are the individual randomness of each sample

Matrix form for each group i:

$$y_i = X_i \beta + Z_i u_i + \epsilon_i$$

where

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{id_i} \end{pmatrix}, \quad X_i = \begin{pmatrix} X_{i1}^T \\ \vdots \\ X_{id_i}^T \end{pmatrix}, \quad Z_i = \begin{pmatrix} Z_{i1}^T \\ \vdots \\ Z_{id_i}^T \end{pmatrix}, \quad \epsilon_i = \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{id_i} \end{pmatrix}$$

2.1 Random intercept and random slope models

Random intercept model:

$$y_{is} = X_{is}^T \beta + u_i + \epsilon_{ij}$$

- Matrix form for each group i:

$$y_i = X_i \beta + u_i 1 + \epsilon_i \tag{1}$$

and $\mathrm{Var}(y_i) = \sigma_u^2 11^T + \sigma_e^2 I$

- for any $s \neq k$

$$\mathrm{corr}(y_{is}, y_{ik}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \geq 0$$

Linear model with random intercept and random slope:

Example: a clinical study understanding the effect of a drug treating veterans suffering from chronic alcohol dependence.

- Each individual (veteran) is measured at for time points: 4, 26, 52 and 78 weeks

- Total number of veterans: 627

- The response is a financial satisfaction score

- Each individual is randomly assigned to the drug treatment or placebo treatment

- Two covariates: whether the individual takes the drug or not, the time point

- There are in total 726×4 observations: y_{is}

In our model, we want to consider three aspects: 1. the drug may have a different effect at different time points; 2. the four measures for the same individual are correlated; 3. individuals can have a different baseline satisfactory score and time may have a different effect for each individuals. We build the following model:

$$y_{is} = (\beta_0 + u_{i1}) + (\beta_1 + u_{i2})t_s + \beta_2 x_i + \beta_3 t_s x_i + \epsilon_{is}$$

- $t_s = \log(\text{week number} + 1)$, x_i is whether the individual takes the drug or not

- In terms of the general form of the LMM model, here $Z_{is} = (1, t_s)$ and $u_i = (u_{i1}, u_{i2})$

2.2   Hierarchical models for a multi-level design

A lot of social / life science experiments have a complicated design structure. Let's take the smoking prevention and cessation study (Chapter 9.2.3) as an example. 1600 students are collected from 135 classrooms in 28 schools. We want to understand the effect of SC (exposure to a school-based curriculum or not), TV (exposure to a television-based prevention program or not) and previous THK scale on the current THK scale. We have 1600 samples, but some share the same school and some share the same classroom.

The multilevel model:

$$y_{ics} = \beta_0 + \beta_1 \text{PTHK}_{ics} + \beta_2 \text{SC}_{ics} + \beta_3 \text{TV}_{ics} + u_s + v_{cs} + \epsilon_{ics}$$

- School effect: $u_s \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$

- classroom effects: $v_{cs} \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$

- individual randomness: $\epsilon_{ics} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$

- Correlation between students in the same classroom: for any $i \neq i'$

$$\text{corr}(y_{ics}, y_{i'cs}) = \frac{\sigma_u^2 + \sigma_v^2}{\sigma_u^2 + \sigma_v^2 + \sigma_e^2}$$

- Correlation between students in the same school but different classrooms: for any $c \neq c'$, i_1, i_2

$$\text{corr}(y_{i_1 cs}, y_{i_2 c's}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2 + \sigma_e^2}$$

3   Model estimation

3.1   Estimation of β: MLE

Let the total number of individuals be N and total number of unique random effect terms be p_2. In general, we can write down a matrix form of the LMM for the whole dataset:

$$y = X\beta + Zu + \epsilon$$

Here $y \in \mathbb{R}^N$, $u \in \mathbb{R}^{p_2}$ and $\epsilon \in \mathbb{R}^N$ are vectors of random variables, and X and Z are known matrices (Chapter 9.3.1).

For instance, if the data follows the random intercept model (model (1)), then

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, Z = \begin{pmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & Z_n \end{pmatrix}, u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

In LMM, we assume that $u \sim N(0, \boldsymbol{\Sigma}_u)$. If the data follows the random intercept model (model (1)), then $\boldsymbol{\Sigma_u} = \mathrm{diag}(\Sigma_u, \cdots, \Sigma_u)$. Marginally, y follows the distribution that

$$y \sim N(X\beta, Z\boldsymbol{\Sigma_u}Z^T + R_\epsilon)$$

where $R_\epsilon = \mathrm{Cov}(\epsilon) = \sigma_e^2 I$.

Define $V = Z\boldsymbol{\Sigma_u}Z^T + R_\epsilon$, if V is known, then then we have a closed-form MLE solution for β, which is

$$\tilde{\beta} = \tilde{\beta}(V) = (X^T V^{-1} X)^{-1} X^T V^{-1} y$$

In practice, V is unknown, we will plug in an estimate $\hat{V}$ and use the estimate

$$\widehat{\beta} = \tilde{\beta}(\widehat{V})$$

How to find $\widehat{V}$?

3.2   Estimation of V: residual ML (REML)

How can we estimate V without knowing β?

The projection matrix in linear regression: $P_X = X(X^T X)^{-1} X^T$. Remember that the residuals of least square in linear regression is

$$(I - P_X)y = (I - X(X^T X)^{-1} X^T)y$$

Under the LMM model, we have

$$Ly = (I - P_X)y = (I - X(X^T X)^{-1} X^T)y = (I - P_X)(Zu + \epsilon)$$

where we define $L = I - P_X$. We know that

$$Ly \sim N(0, LVL^T)$$

thus the likelihood of Ly does not involve β and we can maximize this likelihood to find the estimate of V.

3.3   Predictions of each u_i: best linear unbiased predictor (BLUP)

Why do we want to predict u_i? We may be interested in finding the groups that has high/low effects.

- We use "prediction" instead of "estimation" as in LMM, u_i are random varibles instead of unknown parameters

- You may wonder why using LMM instead of treating each u_i as a fixed parameter.

- Compared with using just a fixed effect model treating each group as an indicator, in LMM we additionally assume $u_i \sim N(0, \Sigma_u)$

- Benefits of LMM:

    - Reduce the number of parameters
    - Groups can borrow information from each other. Say a group i may have 2 samples, if we treat u_i as a parameter, then there are only two samples to estimate u_i, which can provide very bad estimates.

We predict each u_i by an estimate of its posterior mean:

$$\hat{u}_i = \widehat{E}[u_i \mid y]$$

The joint distribution of y and u is

$$\begin{pmatrix} y \\ u \end{pmatrix} \sim N\left[\begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} Z\boldsymbol{\Sigma_u}Z^T + R_\epsilon & Z\boldsymbol{\Sigma_u} \\ \boldsymbol{\Sigma_u}Z^T & \boldsymbol{\Sigma_u} \end{pmatrix}\right]$$

From above we can get the conditional distribution $u \mid y$ which also follows a Normal distribution, the conditional expectation is

$$E[u \mid y] = \boldsymbol{\Sigma_u}Z^T(Z\boldsymbol{\Sigma_u}Z^T + R_\epsilon)^{-1}(y - X\beta) = \boldsymbol{\Sigma_u}Z^TV^{-1}(y - X\beta)$$

When V is known, our prediction will be

$$\hat{u} = \boldsymbol{\Sigma_u}Z^TV^{-1}[I - X(X^TV^{-1}X)^{-1}X^TV^{-1}]y$$

which is the best linear unbiased predictor (BLUP).

In practice, V is not known, we can plug in the estimate of V (and Σ_u) from REML and get the predictor

$$\hat{u} = \widehat{\boldsymbol{\Sigma_u}}Z^T\widehat{V}^{-1}[I - X(X^T\widehat{V}^{-1}X)^{-1}X^T\widehat{V}^{-1}]y$$

Next time: linear mixed effect model examples