STAT347: Generalized Linear Models
Lecture 13

Today's topics: Chapters 9.4, 9.5, 9.7

- Two examples for linear mixed effect models
- GLMM: generalized linear mixed effect model
  - Binomial response: logistic-normal models
  - Poisson GLMM
  - Marginal likelihood MLE for GLMM: Gauss-Hermite Quadrature (Chapters 9.5.1, 9.5.2)
- Example: modeling correlated survey responses

1 Two examples for LMM

1.1 Multilevel model for smoking prevention and cessation study (Chapter 9.2.3)

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:

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

Please see the R Data example 7

1.2 Multi-subject, multi-group example

We try to understand the relationship between a student's GPA on his/her test scores.

- Each student has a GPA x_i
- For $j = 1, 2, \cdots, p$th type of exam, student i has a test score y_{ij}

Here are a few related modeling ideas from different perspectives

- Assume that y_{ij} are i.i.d. across students for each exam j

$$y_{ij} = \beta_{0j} + \beta_{1j} x_i + \epsilon_{ij}$$

- To consider the fact that each student can have different ability/background, that affects scores across all of his/her exams, there are two perspectives

- Each student has a student-specific baseline score:

$$y_{ij} = (\beta_{0j} + u_i) + \beta_{1j} x_i + \epsilon_{ij}$$

which shows that the model has a student-specific intercept (baseline). Here there is both a student indicator and an exam type indicator.

- scores are correlated within each student by sharing a latent variable u_i

$$y_{ij} = \beta_{0j} + \beta_{1j} x_i + u_i + \epsilon_{ij}$$

where $u_i \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$

- The above two ideas are very similar. In the first idea, we can add a prior of u_i to borrow across students, and then we have the same LMM as from the second idea. From the perspective of the first idea, u_i can also be fixed p different parameters.
- Treating u_i fixed we assume less model assumptions while by treating u_i random we obtain more efficient estimate of both μ_i and β.

2 Generalized linear mixed effect models

For LMM, the form is

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

with u_i and ϵ_{is} random. With the typical assumption that $E(u_i) = E(\epsilon_{is}) = 0$, we would also have marginally

$$E(y_{is}) = X_{is}^T \beta$$

However, for GLMM, the model is

$$g[E(y_{is} \mid u_i)] = X_{is}^T \beta + Z_{is}^T u_i$$

when the link function g is non-linear, marginally after integrating out the randomness in μ_i we would have

$$g[E(y_{is})] \neq X_{is}^T \beta$$

In GLMM with non-linear link functions, if u_i does exist but we ignore it, then we will not only have over-dispersion, we will also have a biased estimate of β.

2.1 Binomial response

- Logistic-normal model:

$$\text{logit}[P(y_{is} = 1 \mid u_i)] = X_{is}^T \beta + Z_{is}^T u_i$$

  - Item response models: y_{ij} the yes/no (correct/incorrect) response of subject i on question j

$$\text{logit}[P(y_{ij} \mid u_i)] = \beta_0 + \beta_j + u_i$$

- latent variable threshold model with random effects:

  Remember for binary GLM, we can also write down the link as the form
  $$P(y_{is} = 1) = F(X_{is}^T \beta)$$
  With random effects, we can extend to the assumption:
  $$P(y_{is} = 1 \mid u_i) = F(X_{is}^T \beta + Z_{is}^T u_i)$$
  In other words, from the late variable threshold modeling prospective, we assume there is a latent $y_{is}^\star$ where
  $$y_{is}^\star = X_{is}^T \beta + Z_{is}^T u_i + \epsilon_{is}$$
  where ϵ_{is} are i.i.d. following some distribution (normal, logistic, ...) and we have
  $$y_{is} = \begin{cases} 1 & \text{if } y_{is}^\star >= 0 \\ 0 & \text{else} \end{cases}$$
  Here are some properties:

  - Conditional independence:
    $$P(y_{i1} = a_1, \cdots, y_{id_i} = a_{d_i} \mid u_i = u_\star) = P(y_{i1} = a_1 \mid u_i = u_\star) \cdots P(y_{id_i} = a_{d_i} \mid u_i = u_\star)$$
  - Marginal correlation:
    $$\begin{aligned} \mathrm{cov}(y_{is}, y_{ik}) &= E[\mathrm{cov}(y_{is}, y_{ik} \mid u_i)] + \mathrm{cov}[E(y_{is} \mid u_i), E(y_{ik} \mid u_i)] \\ &= 0 + \mathrm{cov}[F(X_{is}^T \beta + Z_{is}^T u_i), F(X_{ik}^T \beta + Z_{ik}^T u_i)] \end{aligned}$$
    where F is the cdf of $-\epsilon_{is}$. If $Z_{is} = 1$ (the random intercept model), then $\mathrm{cov}(y_{is}, y_{ik}) > 0$.

Marginally,
$$\mathbb{E}(y_{is}) = P(y_{is} = 1) \neq F(X_{is}^T \beta)$$
After some calculations to integrate out the random variable u_i (see page 308), we have

- For the probit link random-intercept model $P[y_{is} = 1 \mid u_i] = \Phi(X_{is}^T \beta + u_i)$,
  $$P(y_{is} = 1) = \int P(y_{is} = 1 \mid u_i = u) f(u) du = \int P(\epsilon_i \leq u + X_{is}\beta) f(u) du$$
  where $\epsilon_i \sim N(0, 1)$ and $f(u)$ is the density of u_i. Since $\epsilon_i - u_i \sim N(0, 1 + \sigma_u^2)$, we have $P(y_{is} = 1) = \Phi(X_{is}\beta / \sqrt{1 + \sigma_u^2})$, so
  $$g(P(y_{is} = 1)) = \frac{X_{is}^T \beta}{\sqrt{1 + \sigma_u^2}}$$
- For the logistic-normal model:
  $$g(P(y_{is} = 1)) \approx \frac{X_{is}^T \beta}{\sqrt{1 + \sigma_u^2 / c^2}}$$
  where $c \approx 1.7$
- Why does the β in the random effect model typically larger than the marginal relationship between x and y? Figure 9.2 (compare with linear regression)

2.2 Poisson GLMM

$$\log[E(y_{is} \mid u_i)] = X_{is}^T\beta + Z_{is}^T u_i$$

Equivalently,

$$E[y_{is} \mid u_i] = e^{Z_{is}^T u_i} e^{X_{is}^T \beta}$$

For the random-intercept model where $Z_{is} = 1$ and $u_i \sim N(0, \sigma_u^2)$, we have

$$E(y_{is}) = e^{X_{is}^T\beta + \sigma_u^2/2}$$

The coefficients β does not change except for the intercept.

2.3 Fitting GLMM with Gauss-Hermite Quadrature methods

Fitting GLMM is more complicated than fitting LMM as the marginal distribution of the observations $\{y_{is}\}$ do not have a closed form. You may learn other methods like MCMC and EM in the future. Here we very briefly discuss how to approximate the marginal likelihood numerically.

The marginal likelihood

$$l(\beta, \Sigma_u; y) = f(y; \beta, \Sigma_u) = \int f(y \mid u, \beta) f(u; \Sigma_u) du$$

This typically do not have a closed form

Gauss-Hermite Quadrature methods: approximate the integral by a weighted sum

$$\int h(u) \exp(-u^2) du \approx \sum_{k=1}^{q} c_k h(s_k)$$

- the tabulated weights $\{c_k\}$ and quadrature points $\{s_k\}$ are the roots of Hermite polynomials.
- The approximation is more more accurate with larger q. For more details, read chapter 9.5.2.
- The approximated likelihood is maximized with optimization algorithms such as Newton's method

Laplace approximation: the marginal density of our data has the form

$$\int e^{l(u)} du \approx \int e^{l(u_0) - \frac{1}{2} l''(u_0)(u-u_0)^2} du = e^{l(u_0)} \sqrt{\frac{2\pi}{l''(u_0)}}$$

Here u_0 is the global maximum of $l(u)$ satisfying $l'(u_0) = 0$. Laplace approximation can be used when u is multi-dimensional.

3 Example: modeling correlated survey responses (Chapter 9.7)

See R Data Example 7.