STAT347: Generalized Linear Models
Lecture 15

Winter, 2023

Jingshu Wang

Today’s topics:

- GLMM: generalized linear mixed effect model
  - Binomial response: logistic-normal models
  - Poisson GLMM
  - Marginal likelihood MLE for GLMM: Gauss-Hermite Quadrature

- Example: modeling correlated survey responses

LMM V.S. GLMM

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

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

If we ignore the random effects but use a regular linear model

- Our estimates for β will still be consistent
- We underestimate the uncertainty in $\hat{\beta}$

LMM V.S. GLMM

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

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

If we ignore the random effects but use a regular GLM model

- Our estimates for β will be biased
- The uncertainty in $\hat{\beta}$ will also be wrongly evaluated (likely under-estimated)

GLMM for binomial response

Logistic-normal model:

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

where $u_i \sim N(0, \Sigma_u)$ and are independent

- Example: item-response models

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

We can view GLMM for binary responses as latent variable threshold model with random effects

We assume that

$$P(y_{is} = 1 \mid u_i) = F(X_{is}^T \beta + Z_{is}^T u_i)$$

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

$$y_{is} = \begin{cases} 1 & \text{if } y_{is}^{\star} >= 0 \\ 0 & \text{else} \end{cases}$$

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

$$\begin{aligned} \mathrm{cov}(y_{is}, y_{ik}) &= E[\mathrm{cov}(y_{is}, y_{ik} \mid u_i)] + \mathrm{cov}[E(y_{is} \mid u_i), E(y_{ik} \mid u_i)] \\ &= 0 + \mathrm{cov}[F(X_{is}^T \beta + Z_{is}^T u_i), F(X_{ik}^T \beta + Z_{ik}^T u_i)] \end{aligned}$$

  - For random intercept Binary GLMM, the correlation between two responses within the same group is still positive (same as LMM)

$$\mathrm{cov}(y_{is}, y_{ik}) > 0$$

Bias in $\hat{\beta}$ is the Binary GLMM is true but we use GLM

- Generally

$$\mathbb{E}(y_{is}) = P(y_{is} = 1) \neq F(X_{is}^T \beta)$$

- For some models, especially, the random intercept Binary GLMM, we can find that the marginal model (ignoring the random effects) is roughly still a GLM, but with true coefficients shrinkage towards 0

Probit link random intercept model

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

- The marginal probability

$$P(y_{is} = 1) = \int P(y_{is} = 1 \mid u_i = u) f(u) du = \int P(\epsilon_i \leq u + X_{is}\beta) f(u) du$$

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

$$g(P(y_{is} = 1)) = \frac{X_{is}^T \beta}{\sqrt{1 + \sigma_u^2}}$$

- This indicates that the marginal probabilities follows a

Probit link random intercept model

$$g(P(y_{is} = 1)) = \frac{X_{is}^T \beta}{\sqrt{1 + \sigma_u^2}}$$

- This indicates that the marginal probabilities still follow a probit link, but with

$$\beta^{\text{marginal}} = \frac{\beta}{\sqrt{1 + \sigma_u^2}}$$

- If we ignore the random effects but fit a probit GLM
- Our estimates for β will be biased by $1/\sqrt{1 + \sigma_u^2}$
- We still underestimate the uncertainty in $\hat{\beta}^{\text{marginal}}$ (as we ignore the fact that samples are correlated)

Marginal GLM for Logistic–normal model

- We have a similar approximation for the logistic-normal model if we only have random intercept

$$g(P(y_{is} = 1)) \approx \frac{X_{is}^T \beta}{\sqrt{1 + \sigma_u^2 / c^2}}$$

where $c \approx 1.7$

Marginal GLM properties

- Why does the β in the random effect model typically larger than the coefficient β^{marginal} in the corresponding marginal GLM?

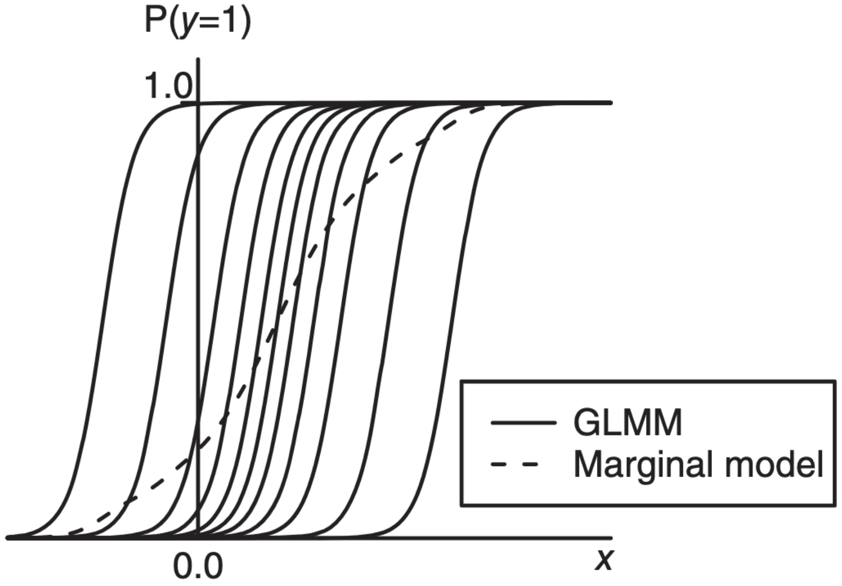

Figure 9.2 Logistic random-intercept GLMM, showing its subject-specific curves and the population-averaged marginal curve obtained at each x by averaging the subject-specific probabilities.

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.

Fitting GLMM

- Fitting GLMM is more challenging than fitting LMM as the marginal distributions of the responses y_{is} typically do not have closed forms

- Typical methods
  - Full Bayes approach MCMC
  - EM algorithm
  - Approximate the marginal likelihood numerically

The marginal likelihood

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

Gauss-Hermite Quadrature

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

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.

Example: modeling correlated survey responses

- Check Example9 R notebook