STAT347: Generalized Linear Models
Lecture 13

Winter, 2024
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
  - Generalized estimating equations
- 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

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

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 binary response: 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

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

Example: probit model with random intercept

- Latent continuous variable follow LMM:

$$y_{is}^{*} = X_{is}^{T}\beta + u_i + \epsilon_{is}, \qquad \epsilon_{is} \sim N(0,1), u_i \sim N(0, \sigma_u^2)$$

- Conditional mean model for the observed y_{is}

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

- Marginal mean model for the observed y_{is}

$$P(y_{is} = 1) = P\left(u_i + \epsilon_{is} \leq X_{is}^{T}\beta\right) = \Phi\left(\frac{X_{is}^{T}\beta}{\sqrt{1 + \sigma_u^2}}\right)$$

Example: probit model with random intercept

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

- 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)

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

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

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 for binary GLMM

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

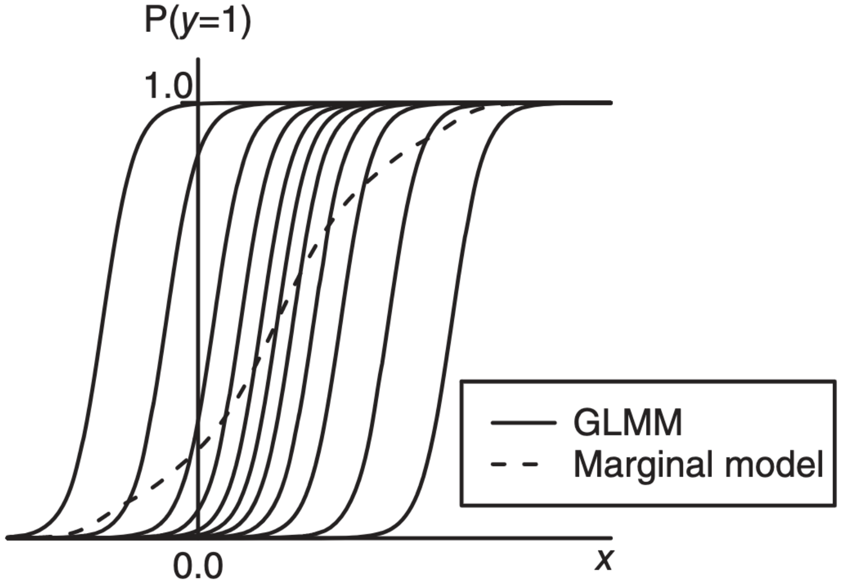

Figure 9.2 Logistic random-intercept GLMM, showing its subject-specific curves and the population-averaged marginal curve obtained at each x by averaging the subject-specific probabilities.

Some properties

- Conditional independence

$$P(y_{i1} = a_1, \cdots, y_{id_i} = a_{d_i} \mid u_i = u_\star) = P(y_{i1} = a_1 \mid u_i = u_\star) \cdots P(y_{id_i} = a_{d_i} \mid u_i = u_\star)$$

  - Latent class model

- Marginal correlation

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

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

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

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

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

- For the marginal model, the link function is still log-linear
- The coefficient $\beta^{\text{marginal}} = \beta$ except for the intercept
- Marginal GLM is not longer a Poisson GLM → over-dispersion due to the random effect term (Agresti book Chapter 9.4.2)

$$var(y_{is}) = E(y_{is}) + \left(E(y_{is})\right)^2 \left(e^{\sigma_u^2} - 1\right)$$

Matrix form of the GLMM model

- Similar to LMM, denote the model for the whole dataset
$$g(E[y|u]) = X\beta + Zu$$

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

- Number of groups is n

- y_i, X_i, Z_i, u_i are the response, covariates and random effects for group i
- Can also allow multiple grouping structures (hierarchical or not)

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 (not easy)
  - Approximate the marginal likelihood numerically
  - Generalized estimating equations (GEE): fitting the marginal model

The marginal likelihood

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

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.

- $l(u) = \log[f(y|u,\beta)] + \log[f(u,\Sigma_u)]$ which is the log density of the joint likelihood of y and u
- For canonical link, $\dot{l}(u) = \frac{Z^T(y - E[y|u])}{a(\phi)} - \Sigma_u^{-1} u$

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

Generalized estimating equations (GEE)

- A way to estimate the marginal model under dependence across observations

- For group i, the response is $y_i = (y_{i1}, \cdots, y_{in_i})$
- Denote the marginal means as $\mu_i = E(y_i)$, marginal GLM:
$$g(\mu_{is}) = X_{is}^T \beta$$

- Elements in y_i are correlated due to shared random effects, we just model a working covariance matrix (may not be true):
$$\mathrm{var}(y_i) = V_i(\alpha) = v(\mu_i)$$

- Responses across groups are independent

Generalized estimating equations (GEE)

- Generalized estimating equation for β

$$\sum_{i=1}^{n} (\partial \mu_i / \partial \boldsymbol{\beta})^{\mathrm{T}} v(\mu_i)^{-1} (y_i - \mu_i) = \mathbf{0},$$

- Compare with estimating equation for β for independent responses

$$\varphi_{1j}(\beta, \phi) = \frac{\partial L}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i) x_{ij}}{a(\mu_i, \phi)} \frac{1}{g'(\mu_i)} = 0$$

- We also need a generalized estimating equation for scale parameters α
    - We can use moment equations as before

$$\sum_{i=1}^{n} (y_i - \mu_i)^T V_i(\alpha)^{-1} (y_i - \mu_i) \quad = N - p$$

- Typically, we assume the correlation matrix is shared across groups
- Can use Sandwich estimator to robustly estimate the variance of $\hat{\beta}$

Example: modeling correlated survey responses

- Check Example9 R notebook