STAT6494: Adv Topics Bayesian Statistics

Inyoung Kim

Outline

Generalized Linear Mixed Models

- The generalized linear mixed model (GLMM) is the GLM generalization of the normal linear random effects model.
- It is defined as follows.
- Let y_{ij} denote the jth measurment on the ith subject.
- Suppose the sampling distribution of y_{ij}, $j = 1, ..., n_i$, $i = 1, ..., n$ is from an exponential family. so that

$$p(y_{ij}|\theta_{ij}, \phi) = exp\{\phi^{-1}(y_{ij}\theta_{ij} - a(\theta_{ij})) + c(y_{ij}, \phi)\}$$

- Without loss of generality, let us assume that $\phi = 1$ as in the logistic and Poisson regression models.

Generalized Linear Mixed Models

- Thus we assume the y_{ij}'s are independent, each y_{ij} has canonical parameter θ_{ij} and each y_{ij} comes from an exponential family. As before, we have

$$\begin{aligned} \mu_{ij} &= E(y_{ij}|\theta_{ij}) = a'(\theta_{ij}) = \frac{da(\theta_{ij})}{d\theta_{ij}} \\ v_{ij} &= Var(y_{ij}|\theta_{ij}) = a''(\theta_{ij}) = \frac{d^2a(\theta_{ij})}{d\theta_{ij}^2} \end{aligned}$$

- In the GLMM, the canonical parameter θ_{ij} is related to the covariates by

$$\theta(\theta_{ij}) = \eta_{ij} = x_{ij}'\beta + z_{ij}'b_i$$

where $\theta(\theta_{ij})$ is a monotonic function of θ_{ij}, x_{ij}' is a $1 \times p$ vector denoting the jth row of X_i and z_{ij}' is a $1 \times q$ vector denoting the jth row of Z_i.

- Thus $X_i = \begin{pmatrix} x'_{i1} \\ x'_{i2} \\ \vdots \\ x'_{in_i} \end{pmatrix}_{n_i \times p}$, $Z_i = \begin{pmatrix} z'_{i1} \\ z'_{i2} \\ \vdots \\ z'_{in_i} \end{pmatrix}_{n_i \times q}$
- Again, the link function $\theta(\theta_{ij})$ is referred to as the θ-link and $\eta_{ij} = x'_{ij}\beta + z'_{ij}b_i$ is the **linear predictor**.
- Assuming $\phi = 1$, we can write the GLMM as

$$p(y_{ij}|\beta, b_i) = exp\{y_{ij}\theta(x'_{ij}\beta + z'_{ij}b_i) - a(\theta(x'_{ij}\beta + z'_{ij}b_i) + c(y_{ij}))\}$$

- Thus, **conditional** on the **random effects** b_i, the observations on subject i are independent.
- The likelihood function for all n subjects is thus given by

$$p(y|\beta, b) = \prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\beta, b_i)$$

where $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$

Generalized Linear Mixed Models

- Each b_i is $q \times 1$ and β is $p \times 1$.
- When $\theta(\theta_{ij}) = \theta_{ij} = \eta_{ij}$, then the link is a **canonical link**.
- For example, in the GLMM logistic regression model, we have

$$p(y_{ij}|\beta, b_i) = exp\{y_{ij}(x_{ij}'\beta + z_{ij}'b_i) - log(1 + exp(x_{ij}'\beta + z_{ij}'b_i))\}$$

Frequentist Likelihood Based Inference for GLMM's

- Likelihood based inference is based on the **marginal likelihood** with the random effects being integrated out.
- As usual we assume

$$b_i \sim N_q(0, D)$$

so that the full likelihood for subject i is

$$p(y_{ij}|\beta, b_i)\pi(b_i)$$

which leads to

$$p(y, b|\beta) = \prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij|\beta,b_i})\pi(b_i) \tag{0}$$

where

$$p(y_{ij}|\beta, b_i) = exp\{y_{ij}\theta(x_{ij}'\beta + z_{ij}'b_i) - a(\theta(x_{ij}'\beta + z_{ij}'b_i)) + c(y_{ij})\}$$

and

$$\pi(b_i) = (2\pi)^{-\frac{q}{2}}|D|^{-\frac{1}{2}} exp\{-\frac{1}{2}b_i'D^{-1}b_i\}$$

Frequentist Likelihood Based Inference for GLMM's

- The **marginal likelihood of** β is given by

$$p(y|\beta) = \int_{R^{nq}} p(y,b|\beta)db \qquad (1)$$

  where $p(y,b|\beta)$ is given by (1)

- We see that

$$\begin{aligned} p(y|\beta) &= \int_{R^{nq}} p(y,b|\beta)db \\ &= \int_{R^q} \cdots \int_{R^q} \{\prod_i^n \prod_j^{n_i} p(y_{ij}|\beta, b_i)\pi(b_i)db_i\} \\ &= \prod_i^n [\int_{R^q} \prod_{j=1}^{n_i} p(y_{ij}|\beta, b_i)\pi(b_i)db_i] \qquad (2) \end{aligned}$$

- Thus, by integrating out b_i, we induce a correlation structure **within subjects**, that is, (y)

Frequentist Likelihood Based Inference for GLMM's

- Thus the marginal likelihood involves evaluating *nq* dimensional integrals.
- That is, (3) involves *nq* dimensional integrals over R^q.
- For the general class of GLMM's, these integrals do not have a closed form, and are very difficult to evaluate.
- Thus frequentist likelihood based inference from the GLMM is **essentially impossible**, since the marginal likelihood cannot be computed or even well approximated.
- This problem led to the development of non-likelihood based methods
- In particular, methods based on **Generalized Estimating Equations** (GEE).
- These methods are quite common among frequentists, and have been made popular by Zeger& Liang (1986, Biometrics) and Liang and Zeger (1986, Biometics).
- See Diggle, Liang, and Zeger (1994, Oxford University Press) for a goof discussion of GEE methods.

Frequentist Likelihood Based Inference for GLMM's

- GEE's can be viewed as multivariate analogs of quasi-likelihood (Wedderburn, 1974, Biometrika)
- The main idea of GEE's is to develop a set of **score equations** or **estimating equations** to solve for (β, D).
- The estimating equations take the form

$$S_\beta = \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta}\right)' Var(y_i)^{-1}(y_i - \mu_i) = 0$$

- Since $E(y_i|\beta)$ and $Var(y_i|\beta)$ depend on D, we replace D by a consistent estimator, $\hat{D}$, and solve $S_\beta(\beta, \hat{D}) = 0$

Frequentist Likelihood Based Inference for GLMM's

- The covariance matrix D may be estimated by simultaneously solving $S_\beta(\beta, D) = 0$ and

$$S_D(\beta, D) = \sum_{i=1}^{n} (\frac{\partial \tilde{\mu}_i}{\partial D})' H_i^{-1} (w_i - \tilde{\mu}_i) = 0$$

  where $w_i = (y_{i1}y_{i2}, y_{i1}y_{i3}, \ldots, y_{in_{i-1}}y_{in_i}, y_{in_i}^2, \ldots, y_{n_i}^2)'$ is the set of all products of pairs of response and squared responses, and $\tilde{\mu}_i = E(w_i|\beta)$.
- The choice of the weight matrix, H_i, depends on the type of response.
- For binary response, the last n_i components of w_i can be ignored since the variance of a binary response is determined by its mean
- In this case $\begin{pmatrix} Var(y_{i1}y_{i2}|\beta) & 0 & 0 \\ & Var(y_{i1}y_{i3}|\beta) & 0 \\ & \ddots & \\ 0 & & Var(y_{in_{i-1}}y_{in_i}|\beta) \end{pmatrix}$

See Diggle, Liang, and Zeger for a detailed account of GEE methods.