Lecture On Generalized Linear Mixed Models

Outline

The topics to be covered include:

- Generalized Mixed Effects Models (GLMMs)
  - Random Effects
  - GLMM models
  - Likelihood Inference for Mixed Effects models
- (Alternative) Marginal Modeling Approach and GEEs method

1 Random Effects and GLMM models

Random effects

Let us introduce the idea of *random effects* modeling through an example.

- (Motivation example on Auto quality control tests by an auto manufacture) Suppose n cars are randomly selected from new car pool and each car is tested m_i times for fuel efficiency. The measurements y_{ij}, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m_i$ are the (average) miles achieved on a gallon of gas. Since cars are different (gas consumption) from one to the other, we often to use the following model to analysis the data:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad \text{where} \quad \epsilon_{ij} \sim N(0, \sigma^2). \tag{1}$$

  Here, the α_i is a (constant) parameter used to model the effect (different behavior) associated with the ith car.

  - All y_{ij}'s are independent with each other.

- If only these n cars are of our interest, model (1) is a correct model to use, in which the fixed unknown parameter α_i is used to model the (fixed) effect of the ith car. This is the traditional models of fixed effects. Note, the α_i's are usually called *fixed effect* parameters. They are fixed and non-random.

- However, the auto manufacture is interested in the performance of all the cars produced (or to be produced) in the same production lines. These particular n cars are not the targets but random represent Ives of all cars. In another words, the ith car in the quality control test is just a random sample from the (much larger) population pool. Thus, α_i of the ith car should be considered as a random quantity that is sampled from the pool of the α's of all cars produced (or to be produced) in the production lines. To emphasize that the α_i's are random, we replace it by b_i in equation (1). This leads to the so called *random effect* model,

$$y_{ij} = \mu + b_i + \epsilon_{ij} \text{ where } b_i \sim [\text{some distribution}] \text{ and } \epsilon_{ij} \sim N(0, \sigma^2). \tag{2}$$

  The b_i in the model is called a *random effect.*

- A model with random effects coefficients is called *a random effects model.* A model with both random effects coefficient(s) and the traditional non-random fixed effects coefficient(s) is also called as *mixed effects model.*

- Note, under model (2) the observations within the same car (i.e., y_{i1}, y_{i1}, y_{i3} etc.) are correlated and the observations between the cars (i.e., y_{ij} and $y_{i'k}$ for $i \neq i'$) are independent.
- (Model (1) versus model (2)) In the motivation example of auto quality control tests by an auto manufacture, we think model (2) is better, although the estimates of the average gas miles $\hat{\mu} = \bar{y}$ = average of all y_{ij}'s are the same in both models. The difference is that the standard errors of the estimators $\hat{\mu}$ are not the same. We are in favor of model (2), because (1) the model mimic the random sampling nature of the practice; (2) making inference based on model (2) protects us against wrong conclusions by some events of random chance when we have a set of biased samples (e.g., it could happen that all/majority cars sampled in the test are from one side of population with poor gas miles performance).
- The random effects can also be used to model observations that are batch correlated. This includes the so called longitudinal data
  - Longitudinal data: comprise repeated observations over time on each of many individuals. For example, clinical trial, each patients/participants may have multiple measurements.

Generalized Linear Mixed Effects Model

- (Notations/input data) To reflect the batch correlated nature in data, we use double indices to index the observations. So, for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, m_i$, the responses, p covariates for fixed effect coefficients and q covariates for random effect coefficients are respectively denoted by

$$\begin{pmatrix} y_{11} \\ \cdots \\ y_{1m_1} \\ . \\ . \\ . \\ y_{n1} \\ \cdots \\ y_{nm_n} \end{pmatrix}, \begin{pmatrix} x_{111}\ , x_{112}\ , \ldots, x_{11p} \\ \cdots \\ x_{1m_11}, x_{1m_12}, \ldots, x_{1m_1p} \\ . \\ . \\ . \\ x_{n11}\ , x_{n12}\ , \ldots, x_{n1p} \\ \cdots \\ x_{nm_n1}, x_{nm_n2}, \ldots, x_{nm_np} \end{pmatrix}, \text{ and } \begin{pmatrix} z_{111}\ , z_{112}\ , \ldots, z_{11q} \\ \cdots \\ z_{1m_11}, z_{1m_12}, \ldots, z_{1m_1q} \\ . \\ . \\ . \\ z_{n11}\ , z_{n12}\ , \ldots, z_{n1q} \\ \cdots\cdot \\ z_{nm_n1}, z_{nm_n2}, \ldots, z_{nm_nq} \end{pmatrix}.$$

  The q covariates z may or may not be a part of the p covariate x.

- A *generalized linear mixed effects model* (GLMM) is just a generalized linear model that contains some random effects coefficients. Formally, we specify the following requirements:

  ◦ Given random effects $\mathbf{b}_i = (b_{i0}, b_{i1}, \ldots, b_{iq})^T$, the responses $y_{i1}, y_{i2}, \ldots, y_{im_i}$ are conditionally independent.

  ◦ The conditional distribution of y_{it} (given $\mathbf{b}_i$) satisfied a GLM model with

  $$h(\mathrm{E}\{y_{it}|\mathbf{b}_i\}) = \eta_{it} + \xi_{it}^{(b)}$$

  where h is the link function, the fixed effects linear predictor $\eta_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_p x_{itp}$ and the random effects linear predictor $\xi_{it}^{(b)} = b_{i0} + b_{i1} z_{it1} + \ldots + b_{iq} z_{itq}$.

  ◦ The random vector $\mathbf{b}_i$ follows a specific form of distribution, say $F(\cdot, \theta)$. A common choice is that

  $$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{\Sigma}(\theta)),$$

  where the form of $\mathbf{\Sigma}(\theta)$ is known (except for the unknown parameter θ). For example, $\mathbf{\Sigma}(\theta) = \theta\mathbf{I}$ with $\mathbf{I}$ being the identity matrix of size $(q+1) \times (q+1)$.

- (An example) Logistic mixed effects model: Consider a data set with binary responses y_{it}, for $i = 1, \ldots, n$ and $t = 1, 2 \ldots, m_i$. Suppose, given random coefficient b_i, y_{i1} follows a Bernoulli (or binomial) distribution with success rate

  $$P(y_{it} = 1|b_i) = \frac{e^{\beta_0 + \beta_1 x_i + b_i}}{1 + e^{\beta_0 + \beta_1 x_i + b_i}} \quad \text{where} \quad b_i \sim N(0, \sigma_b^2).$$

  Here, b_i is called random intercept and the model is also known as the random intercept logistic regression model.

- The primary objective in many mixed effects models is to adopt the conventional regression tools, which relates the responses to the explanatory variables (covariates). That is, the regression parameters β are of our primary interest. The Secondary interest is to account for the within subject correlation.

Likelihood Inference in GLMMs

- The joint distribution (density) of the responses y_{it}'s is

$$f(\mathbf{y}|\beta,\theta) = \prod_{i=1}^{n} f(\mathbf{y}_i|\beta,\theta) = ... = \prod_{i=1}^{n} \int \{\prod_{j=1}^{m_i} f(y_{ij}|\mathbf{b}_i,\beta,\theta)\} f(\mathbf{b}_i|\theta) d\mathbf{b}_i.$$

  So, the likelihood function is

$$l(\beta,\theta|\mathbf{y}) = \log f(\mathbf{y}|\beta,\theta) = \sum_{i=1}^{n} \log \left[\int \{\prod_{j=1}^{m_i} f(y_{ij}|\mathbf{b}_i,\beta,\theta)\} f(\mathbf{b}_i|\theta) d\mathbf{b}_i \right].$$

  - In a linear mixed effects model (with normal-distributed responses y_{it} and normal-distributed random effects $\mathbf{b}_i$), this likelihood function has explicit form. But in a general GLMM, it involves a q-dimensional integration.

- (In theory) the inference problem is straightforward. We can use the likelihood inference method: We can obtain the maximum likelihood estimates (MLEs) by maximizing the log-likelihood function

$$(\hat{\beta}_{\text{MLE}}, \hat{\theta}_{\text{MLE}}) = \text{argmax}\, l(\beta,\theta|\mathbf{y}).$$

- (In practice) for a linear mixed effects model (lmm), the likelihood inference approach works well and is computationally feasible. For a GLMM, we need to use numerical method to compute the integrations in the likelihood function. The computing is feasible only when q is small. When q is large, it is going to be very expensive to compute.

  - In SAS we use "proc mixed" or "proc glm"(with option 'repeated' or 'random') and in R/S+ we use lme() function to deal with the linear mixed effects model (lmm).
  [INSERT EXAMPLEMIX.SAS HERE]

- For a GLMM model with many random effects, several different approaches have been explores but with only limited successes. The approaches include the Monte Carlo methods, approximation approaches, Bayes methods and stochastic approximation approaches. In particular, the Bayes methods seem to have some advantages in some special models and Bayesian statisticians have been promoting the BUGS software (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml) for such problems.

- *A hierarchical linear model* (HLM) [1] can be viewed as a special kind of mixed effects model, where the random effects are nested within several layers of model structures. For example, in educational study, within randomly picked districts (districts random effects), there are randomly picked schools; within randomly picked schools (school random effects), there are randomly picked classes; within randomly picked classes (class random effects), the students may have several test scores over a period of time (student random effects). The HLM models is widely used in social science, medical research, and related fields. A special software package for fitting hierarchical linear model HLM (Scientific Software International, Inc.) can be found at http://www.ssicentral.com/hlm/hlm.htm.

2 Marginal Modeling Approach for Data with Repeated Measurements

Marginal models

- Instead of using the random effects, an alternative approach of dealing with batch correlated data is the marginal modeling approach (or better known as the *Generalized estimating Equations* (GEEs) approach).
- In the marginal approach, the regression are modeled through the marginal distribution of a single response and the regression modeling is separated from the modeling of the within-subject correlations. Specifically, we assume:
  - ∘ The marginal expectation of y_{it}, $\mu_{it} = \mathrm{E}y_{it}$, is related to explanatory variables by
$$h(\mu_{it}) = \eta_{it}$$
    where h is the link function and $\eta_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_p x_{itp}$.
  - ∘ The marginal variance $\mathrm{var}(y_{it})$ is a function of marginal mean
$$\mathrm{var}(y_{it}) = k(\mu_{it})\phi,$$

[1] see, for example, Bryk A, Raudenbush SW. *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods.* Newbury Park, CA: Sage;1992.

where ϕ is the scale parameter (in GLM) and k is a known function determined by the link function h and the marginal distribution of y_{it}.

- ∘ The covariance between y_{is} and y_{it} ($s \neq t$, $s, t = 1, 2, \ldots, m_i$) is a function of the marginal means (i.e., μ_{is} and μ_{it}) and additional parameters α:

$$\mathrm{cov}(y_{is}, y_{it}) = C(\mu_{is}, \mu_{it}; \alpha),$$

where C is a known function and α is the unknown parameters.[2]

- (An example) Marginal logistic regression model:

$$\mu_{it} = P(y_{it} = 1) = \frac{e^{\beta_0+\beta_1 x_{it}}}{1 + e^{\beta_0+\beta_1 x_{it}}}, \ \mathrm{var}(y_{it}) = \mu_{it}(1-\mu_{it}), \text{ and } \mathrm{corr}(y_{is}, y_{it}) = \rho.$$

(Here, ρ is the unknown correlation coefficient between observations.)

Marginal model versus mixed effects model

- The regression coefficients (i.e., β) in a marginal model is usually different (but connected) than those in a corresponding mixed effects model. However, the connection is usually complicated except for in Gaussian linear mixed effects/marginal models or in Probit mixed effects/marginal models.
- Assume a Gaussian mixed effects model is true, it is easy to see that the β in the mixed effects model is (happen to be) exactly same as those in the corresponding marginal model.
- Assume a probit mixed effects model is ture, we also can prove that the β in the mixed effects model is a constant times the β in the corresponding marginal model.
- Take home message: Both approaches are valid even though they are studying different sets of regression parameters. Usually, the significant tests of the regression (intercept and slope) parameters in these two approaches agree with each other.
  - ∘ The marginal approach is usually simpler but its inference (see below) is usually not optimal.

[2] The covariance structure specified in a marginal model is not that critical. Even this specification is wrong, we can still have consistent estimators of the regression parameter β.

Statistical inference on marginal models

- Liang and Zeger (1986, Biometrics and Biometrka) proposed to estimate the regression parameters β by solving the following estimating equations (it is known as the *Generalized Estimating Equations* (GEEs)):

$$\sum_{i=1}^{n}\{\frac{\partial\mu_i}{\partial\beta}\}^T\{\mathbf{V}(\alpha)\}^{-1}(\mathbf{y}_i-\mu_i)=0$$

where $\mathbf{y}_i=(y_{i1},y_{i2},\ldots,y_{im_i})^T$, $\mu_i=(\mu_{i1},\ldots,\mu_{im_i})$ and $\mu_{ij}=\mathrm{E}(y_{ij})$. The $m_i\times m_i$ matrix $\mathbf{V}(\alpha)$ is the covariance matrix $\mathrm{cov}(\mathbf{y}_i)$ of the response vector $\mathbf{y}_i$ specified in a marginal model; the matrix $\mathbf{V}(\alpha)$ is also known as the *working covariance matrix.*

- (Theory) We can prove that the regression estimators obtained from the GEEs are consistent (i.e., as $n\to\infty$, $\hat\beta\to\beta$). Furthermore, we have that

$$\sqrt{n}(\hat\beta-\beta)\to N(0,\mathbf{S}).$$

- The variance $\mathbf{S}$ can be consistently estimated by a "sandwich" estimator $\mathbf{A}_n\mathbf{B}_n^{-1}\mathbf{A}_n$, where

$$\mathbf{A}_n=\frac{1}{n}\sum_{i=1}^{n}(\frac{\partial\mu_i}{\partial\beta})^T\{\mathbf{V}(\alpha)\}^{-1}(\frac{\partial\mu_i}{\partial\beta})$$

and

$$\mathbf{B}_n=\frac{1}{n}\sum_{i=1}^{n}(\frac{\partial\mu_i}{\partial\beta})^T\{\mathbf{V}(\alpha)\}^{-1}(\mathbf{y}_i-\mu_i)(\mathbf{y}_i-\mu_i)^T\{\mathbf{V}(\alpha)\}^{-1}(\frac{\partial\mu_i}{\partial\beta}).$$

- The above result still holds, even if the specification of the covariance matrix $\mathbf{V}(\alpha)$ is not correct. But we may suffer some loss of efficiency (e.g., $\hat\beta$ has bigger variance, etc.)

  [INSERT EXAMPLEMIX2 (SAS/R) AND FIGUREMIX HERE]

GEEs inference versus mixed effects model (likelihood) inference

Often (but not always) the conclusions based on a GEE method and a mixed effects modeling approach agree with each other. But there are advantages and disadvantages of each approach in practice.

- The mixed effects models try to mimic the actual sampling collection approach and produce specific types of correlation structures. If we use likelihood inference and the random effects modeling is right, this approach is the most efficient.
- The down side of the mixed effects approach is that when there are too many random effects in a model in a GLM setting, it could be computationally very expensive (could even be not feasible). Also, there is no way to guarrantee (or no way to easily check) that mixed effects model is the "correct" one.
- The GEEs inference approach provides a simple alternative for fitting batch correlated data. The advantages of this approach include (1) it is computationally simple and (2) it is robust against mis-specification of correlation structures.
- The downside of the GEE approach is that it may not be (usually are not) efficient, and the regression parameters in a marginal modeling approach are usually different (with different interpretaions) than those in the mixed effect approaches.