Generalized Linear Mixed Models: GEE or Random Effects?

Chris Cox and Mike Schneider

Linear Mixed Models for Longitudinal Data

$$Y_{it} = x_{it}\beta + z_{it}b + \varepsilon_{it}$$

$1 \le i \le n$ subject $1 \le t \le n_i$ time

Assumptions for linear mixed models:

1) **b ~ MVN(0,G)**
2) ε_i **~ MVN(0,**$\mathbf{R}_i$**)**
3) **b and** ε_i **are independent**
   Y_i **and** Y_j **are independent**

When $\mathbf{R}_i = \sigma^2 \mathbf{I}$ and $\mathbf{z}_i = \mathbf{0}$, the mixed model reduces to the standard linear model.

Linear Mixed Models

The variance of Y_i is $\mathbf{V}_i = \mathbf{z}_i\mathbf{G}\mathbf{z}_i' + \mathbf{R}_i$

In fact we know the *marginal likelihood* of the *observed* data!

$$Y_i \sim \mathrm{MVN}(\mathbf{x}_i\boldsymbol{\beta},\ \mathbf{z}_i\mathbf{G}\mathbf{z}_i' + \mathbf{R}_i)$$

Estimation by **maximum likelihood**:
SAS: proc mixed; Stata: xtmixed

Example: Longitudinal Data with a Random Subject Effect (Intercept)

$$Y_{it} = \mathrm{x}_{it}\beta + \mathrm{u}_i + \varepsilon_{it}$$

$$\mathrm{u}_i \sim \mathrm{IID}\,N(0,\sigma_u^2) \quad 1 \le i \le n$$

$$\mathrm{b} = \left(\mathrm{u}_1, \mathrm{u}_2, \ldots \mathrm{u}_n\right)' \qquad \mathrm{G} = \sigma_u^2 \mathrm{I}$$

$$\varepsilon_{it} \sim \mathrm{IID}\,N(0,\sigma^2) \quad 1 \le t \le n_i \qquad \mathrm{R} = \sigma^2 \mathrm{I}$$

$$\mathrm{corr}(Y_{is}, Y_{jt}) = 0 \qquad i \ne j$$

$$\sigma_u^2 / \left(\sigma_u^2 + \sigma^2\right) \qquad i = j$$

Generalized Linear Mixed Models for Longitudinal Data

$$E(Y_{it} \mid \mathrm{b}) = \mathrm{h}\left(\mathrm{x}_{it}\beta + \mathrm{z}_{it}\mathrm{b}\right)$$

$1 \le i \le n$ subject $1 \le t \le n_i$ time

Assumptions for generalized linear mixed models:

1) **The conditional distribution** $p_\beta(Y \mid \mathrm{b})$ **is a generalized linear model (binomial, Poisson, multinomial)**
2) **h is the link function**
3) **b ~ MVN(0, G)**

When $\mathbf{z}_i = \mathbf{0}$, the mixed model reduces to the standard generalized linear model.

Example: Logistic Regression Model for Longitudinal Data: Random Subject Effect

$$p_\beta(Y_{it} = 1 \mid \mathrm{u}_i) = 1/\left\{1 + \exp\left[-(\mathrm{x}_{it}\beta + \mathrm{u}_i)\right]\right\}$$

$$\mathrm{u}_i \sim \mathrm{IID}\, N(0, \sigma^2) \quad 1 \le i \le n$$

$$\mathrm{corr}(Y_{is}, Y_{jt}) = 0 \quad i \ne j$$

$$> 0 \quad i = j$$

Marginal Likelihood of the Observed Data

$$\int_{-\infty}^{\infty} p_{\beta}(Y \mid u)\phi(u / \sigma)du$$

- Compute (for each observation) by numerical integration (adaptive Gauss-Hermite quadrature): integral replaced by weighted sum
- Estimation by maximum likelihood, including numerical derivatives
- Standard errors from the information or empirical (sandwich) estimate
- Prediction of random effects

Software for Generalized Linear Mixed Models

SAS: proc glimmix

- **Pseudo-likelihood method preferred - Four fitting algorithms linearize the model (not recommended)**
- **For ML, number of integration points in quadrature approximation can matter**
- **One point method (Laplace) often works well**
- **Choice of G matrix same as mixed**
- **Multiple random statements – some restrictions**
- **Spline effects also available - *experimental***
- **R side effects available (quasi-likelihood - GEE)**

Software for Generalized Linear Mixed Models

Stata: xtmelogit, xtmepoisson

- Maximum likelihood
- Numerical quadrature: number quadrature points
- Designed for heirarchical models (nested random effects (children in classrooms in schools)
- Choice of G matrix limited
- Syntax is similar to xtmixed
- Exponentiated coefficients available
- Robust SE's use (prefix command) bootstrap: (slow)

Software for Generalized Linear Mixed Models

Stata: gllamm

- gllamm is *contributed*, manual available
- Reference: Rabe-Hesketh & Skrondal *Multilevel and Longitudinal Modeling Using Stata*, 2nd ed. Stata Press, 2008
- Useful for ordinal regression (proportional odds model)

Application: Growth Hormone Use in Children with Chronic Kidney Disease

51 children with short stature, not using at baseline, up to 5 visits, N=188

Outcome:

- **-- Growth hormone therapy use at visit (Y/N)**

Exposures (baseline except as indicated):

- **-- Age-sex specific height z-score (at previous visit)**
- **-- Race (White vs. Non-White)**
- **-- Insurance (Private vs. Not Private)**
- **-- Education (≥ college vs. < college)**
- **-- Household Income (>75K vs. ≤75K)**
- **-- Tanner Stage (II – V vs. I)**
- **-- Age (>10, 6 – 10 vs. <6)**
- **-- GFR**
- **-- visit number**