STAT 770 Nov. 18 Lecture 23
GLMs with Random (Intercept) Effects

Reading and Topics for this and next lecture: Chapter 13 Sections 13.1 – 13.3 and Handout 1(ii) from course web-page.

(1) Random effects model for clustered GLM data

(2) Matched Pairs as an Example of Clusters

(3) Other Examples of GLMMs – Binary outcome models, Item-response models

(4) GLMM likelihood – cluster means, within-cluster correlations, and predictors

(5) Strategies for ML and approximate ML calculation **next class**

What is a GLMM ?

General form of GLMM with cluster-level random intercepts:

$$Y_{ij} \sim f(y \,|\, \theta_{ij}, \underline{X}_{ij}, \epsilon_i) = h(y)\,\text{exp}(\theta'_{ij} y - c(\theta_{ij}))\ , \quad \epsilon_i \overset{iid}{\sim} F(\cdot, \sigma)$$

$$\nabla_\theta\, c'(\theta_{ij}) = \mu_{ij} = g^{-1}(\underline{X}'_{ij}\,\beta + \epsilon_i)$$

$i = 1, \ldots, m$ indexes cluster, with cluster random-effect ϵ_i

$j = 1, \ldots, n_i$ indexes observation within cluster i,

conditionally *iid* given ϵ_i

- $\beta \in \mathbb{R}$, $\sigma > 0$ parameters are shared across cluster
- ϵ_i a random cluster-specific offset

Examples

Idea: cluster combines observations with common experimental setting

Examples: families, schools, clinical centers within which units share some common unmodeled feature, modeled as independent variate differing across cluster

If m were small and offsets ϵ_i of direct interest, they would be viewed as non-random fixed-effect parameters.

Dimension of GLMM parameter (β, σ) is $p+1$.
Model dimension with fixed effects $= p + m$.

Additional Keywords for Models with Random Effects

Empirical Bayes Models: these GLMMs can be interpreted as a way to fit/predict cluster predictive-scores $X'_{ij}\beta + \epsilon_i$ for individual (i, j): **empirical** because of the identical model relationships across cluster, **Bayes** because of the independence of observations (identical within cluster) conditional on 'prior' effects ϵ_i.

Meta-Analysis: think of the clusters (eg clinical centers) as separate (small but not too tiny) studies; one might report estimates $\widehat{\beta}_i$, $\widehat{\mathrm{Var}}(\widehat{\beta}_i)$ from those studies and analyze those as estimates of common β, with ϵ_i modeling differences between studies. GLMM is what one would fit to the unified dataset if it were available.

Special Logistic Mixed Models

Logit-Normal Mixed-model:

$$Y_{ij} \sim \text{Binom}(r_{ij}, \texttt{plogis}(\underline{X}'_{ij}\,\beta + \epsilon_i)) \;\; \text{given} \;\; \epsilon_i \; , \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Binary Matched Pairs: $n_i \equiv 2$, $\underline{X}_{ij} = \underline{x}_i$

similarly, Litter-Studies or Family-studies

Rasch Item-Response Model: $n_i \equiv \nu \geq 2$, $\underline{X}_{ij} = \underline{X}_i \in \mathbb{R}^\nu$ dummy-column for factor with levels $j = 1, \ldots, \nu$, so $\underline{X}'_{ij}\beta \equiv \beta_j$,

$$\text{logit}\; P(Y_{ij} = 1 \,|\, \epsilon_i) \;=\; \beta_j + \epsilon_i$$

j indexes study subject with skill-level β_j, i item with difficulty ϵ_i

Within-cluster Correlation

General Formula: $\text{var}(W) = E(\text{var}(W|Z)) + \text{var}(E(W|Z))$

Consequence: after using the formula for $W = W_1 + W_2$ and $W = W_1 - W_2$ and subtracting 2nd from 1st, dividing by 4:

$$\text{cov}(W_1, W_2) = E\Big(\text{cov}(W_1, W_2 \,|\, Z)\Big) + \text{cov}\Big(E(W_1 \,|\, Z) \cdot E(W_2 \,|\, Z)\Big)$$

Fix cluster index i and $\underline{X}_{ij}$'s (by conditioning on them if random)

$$\text{cov}(Y_{i,j_1}, Y_{i,j_2}) = E\Big(\text{cov}(Y_{i,j_1}, Y_{i,j_2} \,|\, \epsilon_i)\Big) + \text{cov}\Big(E(Y_{i,j_1} \,|\, \epsilon_i) \cdot E(Y_{i,j_2} \,|\, \epsilon_i)\Big)$$

$$= \text{cov}\Big(E(Y_{i,j_1} \,|\, \epsilon_i) \cdot E(Y_{i,j_2} \,|\, \epsilon_i)\Big) \quad > 0$$

Next slide presents an argument why this inequality is true.

Proof of Positive Within-Cluster Correlation

This argument applies to scalar random-intercept GLMM with g, c' strictly increasing.

- $Y \sim f(y, \theta) = h(y)\,\exp(\theta y - c(\theta))$ is stochastically increasing as function of θ, i.e. , for $\theta_1 < \theta_2$ and all y, $\quad F(y, \theta_1) \geq F(y, \theta_2)$

- Y is stochastically increasing in θ if and only if $E_\theta(h(Y)) \nearrow$ in θ for any monotone increasing function h

Pf. for GLMM $\frac{\partial}{\partial\theta} \int h(y) f(y, \theta) dy = \int (y - c'(\theta)) h(y) f(y, \theta) dy > 0$ because $h(y) > h(c'(\theta) > h(z)$ whenever $y > c'(\theta) > z$.

- In GLMM, $g_j(\epsilon_i) \equiv E(Y_{i,j} \,|\, \epsilon_i) \nearrow$ in ϵ_i (varies with $\underline{X}_{ij}$)

- cov $= E\Big[g_{j_1}(\epsilon_i) g_{j_2}(\epsilon_i)\Big] - \Big[E(g_{j_1}(\epsilon_i)) \cdot E(g_{j_2}(\epsilon_i))\Big]$

$$= \int (g_{j_1}(t) - E(g_{j_1}(\epsilon_i))\, g_{j_2}(t) f_\epsilon(t) dt > 0$$

Unit-level Prediction (Logit-Normal Model)

Fix $i, j, \quad \eta = \underline{X}'_{i,j}\,\beta$:

$$P(Y_{ij} = 1) \;=\; E\Big(\texttt{plogis}(\eta + \epsilon_i)\Big) \;=\; \int \frac{e^{\eta+\sigma z}}{1 + e^{\eta+\sigma z}}\,\phi(z)\,dz$$

Remark. If $\underline{X}_{ij} = \underline{X}_i$ depends only on i, not j,

• the displayed quantity is the inverse of a modified link-function g^*_σ if σ is fixed

• the last inequality on the previous slide is the Cauchy-Schwarz inequality

GLMM Likelihood

Consider logit-normal model, and for simplicity restrict to case where all Y_{ij} are binary, and cluster i has n_i unit observations.

Then

$$P(Y_{i,j} = y_j,\ 1 \le j \le n_i) \ = \ E\Big(P(Y_{i,j} = y_j,\ 1 \le j \le n_i \,\Big|\, \epsilon_i)\Big)$$

$$= \int \exp\Big(\sum_{j=1}^{n_i} y_j\,(\underline{X}_{ij}^{tr}\,\beta + \sigma z)\Big)\Big[\prod_{j=1}^{n_i}(1 + e^{\underline{X}_{ij}^{tr}\,\beta + \sigma z})\Big]^{-1}\phi(z)\,dz$$

We will consider strategies for maximizing in β, σ the sum of log's over i, mostly in the simpler case where all covariates $\underline{X}_{i,j} = \underline{X}_i$ are at cluster level

Software Comments

glmer in `R` package `lme4`

`PROC NLINMIX` or `GLIMMIX` in `SAS`

EM, MCMC and **Adaptive Gaussian Quadrature** ideas for logLik maximization