STAT 770 Nov. 30 Lecture 25

R Calculations with GLMM

Reading and Topics for this lecture: Sections 13.1–13.6, and `GLMMscript.RLog`, under Handout 9.

(1) R Scripts (most of lecture)

(2) Logit-Normal Likelihood with Cluster-level Covariates

(3) Posterior Expectations in Logit-Normal Settings

(4) "Small Area Estimation" example (Table 13.2)

GLMM Likelihood

Consider logit-normal model, restricted to case where all Y_{ij} are binary, cluster i has n_i unit observations, and $\underline{X}_{ij} \equiv \underline{X}_i$.

Then

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

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

Last time we discussed 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.

logLik and Posterior Expectation

Restrict to cluster-level covariates, let Y_{i+} be cluster total. Then

$$Y_{i+} \sim \text{Binom}\Big(n_i, \, \texttt{plogis}(\underline{X}_i^{tr}\,\beta + \epsilon_i)\Big) \;, \qquad \epsilon_i/\sigma \sim \mathcal{N}(0,1)$$

Define $\eta_i = \underline{X}_i^{tr}\,\beta$, and

$$\boxed{g(\eta,\sigma,k,n) \;=\; \int e^{(\eta+z\sigma)k}\,(1+e^{\eta+z\sigma})^{-n}\,\phi(z)\,dz}$$

Then $\;\text{logLik}_i \;=\; \log g(\eta_i,\, \sigma,\, Y_{i+}, n_i)$, and

$$E\Big(\epsilon_i \,\Big|\, Y_{i+} = k\Big) \;=\; \int z\sigma\, e^{(\eta_i+z\sigma)k}\,(1{+}e^{\eta_i+z\sigma})^{-n_i}\,\phi(z)\,dz \Big/ g(\eta_i,\,\sigma,\,k,\,n_i)$$

(integrate by parts)

$$= \sigma \int \frac{d}{dz}\Big[e^{(\eta_i+z\sigma)k}\,(1+e^{\eta_i+z\sigma})^{-n_i}\Big]\,\phi(z)\,dz \Big/ g(\eta_i,\sigma,k,n_i)$$

$$=\; \sigma^2\,\Big(k \;-\; n_i\, g(\eta_i,\sigma,k+1,n_i+1)/g(\eta_i,\sigma,k,n_i)\,\Big)$$

‘Small Area Estimation: Empirical Bayes Model’

Here is a logit-normal ‘area-level’ Empirical Bayes Model:

$$\boxed{Y_{i+} \;=\; \mathrm{Binom}\Big(n_i,\, \mathtt{plogis}(\underline{X}_i^{tr}\,\beta \,+\, v_i)\Big)}$$

Y_{i+} = aggregated total from a (weighted) survey sample

$\underline{X}_i$ = vector of predictors for outcomes Y_{ij} in ‘area’ or cluster i

`Binom` reflects random sampling outcomes

v_i = unmodeled random effect due to differences in areas

Target of inference: $\mathtt{plogis}(X_i^{tr}\,\beta + v_i)$

Typical predictors $\widehat{\beta}$ **MLE**, $\widehat{v}_i$ posterior means for plug-in (as in Agresti); better would be posterior mean of $\mathtt{plogis}(X_i^{tr}\,\beta + v_i)$ given Y_{i+}, at $\beta = \widehat{\beta}$.