Model Families for Correlated Data

LMMs: Conditional Model

Conditional/Hierarchical specification of LMM

$$Y_{ij} = \mathbf{X}_{ij}^T \beta + \mathbf{Z}_{ij}^T \mathbf{b}_i + \epsilon_{ij}$$

- Y_{ij} : the *jth* outcome of the *ith* subject.
- β : regression coefficient vector $(p \times 1)$.
- $\mathbf{b}_i$: random effects for the *ith* subject, $\mathbf{b}_i \sim N\{0, \mathbf{D}(\boldsymbol{\theta})\}$
- $\boldsymbol{\theta}$ is a $q \times 1$ vector of variance components.
- ϵ_{ij} : residual, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \cdots, \epsilon_{in_i})^T \sim N\{0, \mathbf{R}(\boldsymbol{\theta})\}$.
- $(\mathbf{X}_{ij}, \boldsymbol{Z}_{ij})$: covariate design matrices.

Equivalently:

$$\begin{aligned} \mathbf{Y}_i | \mathbf{b}_i &\sim N(\mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{b}, \mathbf{R}_i) \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \end{aligned}$$

LMMs: Marginal Model

$$f_i(\mathbf{y}) = \int f_i(\mathbf{y}_i|\mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i$$

Then the marginal model is

$$\mathbf{Y}_i \sim N(\mathbf{Z}_i\beta, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{R})$$

- Estimation and Inference are derived from the marginal model
- Nearly seamless/interchangeable with conditional model
    - Some constraints on the variance components

Model Families: Gaussian Case

Marginal Model:

$$E[Y_{ij}|\mathbf{X}_{ij}] = \mathbf{X}_{ij}\beta \tag{1}$$

Conditional Model:

$$E[Y_{ij}|\mathbf{b}_i, \mathbf{X}_{ij}] = \mathbf{X}_{ij}\beta + \mathbf{Z}_{ij}\mathbf{b}_i \tag{2}$$

Transition Model:

$$E[Y_{ij}|Y_{i,j-1}, \ldots, Y_{i,1}, \mathbf{X}_{ij}] = \mathbf{X}_{ij}\beta + \alpha Y_{i,j-1} \tag{3}$$

- (2) follows directly from (1) $\Rightarrow$ β has marginal AND conditional interpretation, simultaneously
  - Marginalize over $\mathbf{b}_i$ or condition on $\mathbf{b}_i = \mathbf{0}$

Non-normal data: Connection between marginal/conditional models is no longer straightforward!

Model Families: General Case

- **Marginal Model**:
  - Responses modeled marginalized over all other responses
  - (usually) GEEs
  - (possibly) likelihood based models
- **Conditionally Specified Models**:
  - Responses in sequence are conditioned upon other outcomes
  - (e.g.) Transition models
- **Subject-Specific (Conditional) Model**:
  - Responses independent conditionally on subject-specific parameters
  - (usually) Mixed models
  - (possibly) fixed subject specific effects; conditional logistic model