Generalized Linear Mixed Models

- These models extend the conceptual approach of linear mixed effects models.
- The basic premise is that we assume natural heterogeneity across individuals in a subset of the regression coefficients.
- Conditional on the random effects, we assume the responses for a single subject are independent and follow a distribution in the exponential family.

Model Assumption

- Given the subject-specific effects $\boldsymbol{b}_i \in \mathbb{R}^q$, Y_{ij} are independent and follow the EF distribution

Y_{i1} vs Y_{i2}

$$Y_{ij}|\boldsymbol{b}_i \sim f(Y_{ij}|\boldsymbol{b}_i, \theta)$$

fixed

random

- $g(\mathbb{E}(Y_{ij}|\boldsymbol{b}_i)) = \boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{b}_i$
- Correlations are captured by the random effects $\boldsymbol{b}_i \sim N(\mathbf{0}, G)$, assumed to vary independently from subject to subject.

$\begin{cases} g \to \text{identity link} \\ f(.) \to \phi(\ ) \end{cases}$

reduces LMM

Example: Continuous Response

When $Y_{ij}|\boldsymbol{b}_i$ follows a Gaussian distribution, GLMM reduces to the linear mixed effects model:

- $Y_{ij} = \boldsymbol{X}_{ij}\beta + \boldsymbol{Z}_{ij}\boldsymbol{b}_i + \epsilon_{ij}$
- $\boldsymbol{b}_i \sim N(\boldsymbol{0}, G)$
- $\epsilon_{ij} \sim_{iid} N(0, \sigma^2)$

$$Y_{ij}|b_i = E(Y_{ij}|b_i) + \varepsilon_{ij}$$
$$= X_{ij}\beta + Z_{ij} \cdot b_i + \varepsilon_{ij}$$

Example: Logistic Model

Binary logistic model with random intercepts:

$$logit(\mathbb{E}(Y_{ij}|b_i)) = (b_i + \beta_1) + X_{ij}\beta_2$$

For the Bernoulli distribution, we have the variance function as $\text{var}(Y_{ij}|b_i) = \mathbb{E}(Y_{ij}|b_i)(1 - \mathbb{E}(Y_{ij}|b_i))$, and $b_i \sim N(0, \sigma_b^2)$.

$Y_{ij} | b_i \sim \text{Bernoulli}(\pi_{ij})$

$\pi_{ij} = E(Y_{ij}| b_i)$

$= \frac{e^{b_i+\beta_1+X_{ij}\beta_2}}{1+e^{b_i+\beta_1+X_{ij}\beta_2}}$

Example: Poisson Model

Random intercept and slope Poisson regression:

$Y_{ij} | b_i \overset{ind}{\sim} \text{Poisson}(\_\_)$

$$\log \mathbb{E}(Y_{ij}|\boldsymbol{b}_i) = (b_{1i} + \beta_1) + (b_{2i} + \beta_2)t_{ij}$$

For the Poisson distribution, we have the variance function as $\text{var}(Y_{ij}|\boldsymbol{b}_i) = \mathbb{E}(Y_{ij}|\boldsymbol{b}_i)$ and $\boldsymbol{b}_i \sim N(\mathbf{0}, G)$, where G is a 2×2 covariance matrix for (b_{1i}, b_{2i}).

Estimation

$$\begin{cases} Y_{ij} \mid b_i \sim \underline{EF}(\quad) \\ \underline{b_i} \sim N(0, G) \end{cases}$$

$$Y_{ij} \sim$$

- Conditional likelihood approach
- Full likelihood approach $\underline{\text{(EM)}}$
- In general, computations are difficult and require numerical or Monte Carlo integration techniques

Parameter Interpretation

- GLMMs are most useful when the scientific objective is to make inferences about *individuals* rather than population averages.
- Regression parameters β measure the change in expected value of response while holding other covariates and random effects constant.
- For example, in the logistic model

$$logit(\mathbb{E}(Y_{ij}|b_i)) = b_i + \beta_1 + X_{ij}\beta_2$$

  - β_2 measures the change in the log odds of a positive response per unit change in X_{ij}, for the same subject.
  - When we consider two subjects with one unit difference in X_{ij}, the difference in log odds is $\beta_2 + b_i - b_{i'}$.

Contrasting Marginal Models and GLMM

- Not equivalent, unless the link function is linear (Gaussian response).

$$\mathbb{E}(Y_{ij}) = \mathbb{E}(\mathbb{E}(Y_{ij}|b_i)) = \mathbb{E}(g^{-1}(X_{ij}\beta + Z_{ij}b_i))$$

- Unlike marginal models, GLMM provides a potential explanation of the sources of association between repeated measures (via the introduction of random effects).

- Interpretation of parameters is different.
  - Marginal models focus on inferences about the study population.
  - GLMMs focus on inferences about individuals.

Example

Epileptic Seizure (revisit)

- Interested in the effect of treatment with progabide on changes in an individual's rate of seizures.
- GLMM with random intercept and slope (for period of treatment).
- Model without interaction term:

$$\log \mathbb{E}(Y_{ij}|\boldsymbol{b}_i) = \log(2) + (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})\text{period}_{ij} + \beta_3 \text{trt}_i$$

$(\beta_2 + b_{2i})$ period_{ij}

$(\beta_2 + b_{2i})$ $(\text{period}_{ij} + 1)$

- Model with interaction term:

$$\log \mathbb{E}(Y_{ij}|\boldsymbol{b}_i) = \log(2) + (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})\text{period}_{ij} + \beta_3 \text{trt}_i + \beta_4 \overset{1}{\text{trt}_i} \times \text{period}_{ij}$$