BIOS 755: Generalized Linear Mixed Models II

Alexander McLain

March 27, 2025

Generalized Linear Mixed Model for an Ordinal Response

Suppose Y_{ij} is an ordinal response with K categories $(1, \ldots, K)$. A logistic mixed effects model for the *cumulative response probabilities* is given by:

1. Conditional on a vector of random effects $\boldsymbol{b}_i$, the Y_{ij} are independent and have a multinomial distribution
   - the multinomial covariance is determined by the conditional means (given below) or the conditional response probabilities.
2. The k^{th} cumulative response probability for Y_{ij} depends on fixed and random effect with

$$\alpha_{0k} + \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$$

Generalized Linear Mixed Model for an Ordinal Response

2. (cont.) which is related to the conditional cumulative response probabilities with

$$\log\left\{\frac{\Pr(Y_{ij} \leq k|\boldsymbol{b}_i)}{\Pr(Y_{ij} > k|\boldsymbol{b}_i)}\right\} = \alpha_{0k} + \boldsymbol{X}'_{ij}\beta + \boldsymbol{Z}_i\boldsymbol{b}_i$$

3. The random effects have a bivariate normal distribution $\boldsymbol{b}_i \sim N(0, \boldsymbol{G})$.

This is a proportional odds mixed-effects regression model.

Generalized Linear Mixed Model for Counts

Suppose Y_{ij} is a count.

- Usually, we model counts using a Poisson distribution with a log-link.
- Conditional on the random effects $\boldsymbol{b}_i$, the Y_{ij} are independent and have a Poisson distribution with

$$Var(Y_{ij}|\boldsymbol{b}_i) = E(Y_{ij}|\boldsymbol{b}_i)$$

  or that

$$\sqrt{E(Y_{ij}|\boldsymbol{b}_i)} = StdDev(Y_{ij}|\boldsymbol{b}_i)$$

- This would mean that if the expectation is 100 the standard deviation would be 10.

The Negative Binomial Model

- The negative binomial model allows for extra variance versus what we see in the Poisson model.
- This model also uses a log link for the covariate data.
- Under the negative binomial model

$$Var(Y_{ij}|\boldsymbol{b}_i) = E(Y_{ij}|\boldsymbol{b}_i) + \theta\{E(Y_{ij}|\boldsymbol{b}_i)\}^2$$

  where $\theta \geq 0$.
- This model allows for "over-dispersion" and is often called the **over-dispersed Poisson model**.

Generalized Linear Mixed Model with over-dispersion

- We can actually add over-dispersion to any GLMM.
- Suppose that
$$g\{E(Y_{ij}|\boldsymbol{b}_i)\} = \boldsymbol{X}_i\boldsymbol{\beta} + b_{i0}$$
over-dispersion can be modeled by adding an additional random effect at the observation level.
$$g\{E(Y_{ij}|\boldsymbol{b}_i)\} = \boldsymbol{X}_i\boldsymbol{\beta} + b_{i0} + b_{ij0}$$
- This is not recommended with the Poisson, but can be used there or with the logistic, multinomial, etc.
- Over-dispersion means that the variance is larger than what we would expect.

GEE vs GLMM

- GEEs and GLMMs are both popular statistical techniques used for analyzing correlated data, such as longitudinal data or data with repeated measurements.
- Each method has its strengths and applications, depending on the research questions and the structure of the data.
- However, how do we choose which to use?

GEE vs GLMM

Use GEE When:

- **Population Average Effects Are of Interest:** GEE is focused on estimating the average response over the population, making it suitable when you're interested in the general effect of predictors on the outcome across the population rather than individual-specific effects.
- **Robustness to Model Misspecification:** GEE offers robust standard errors that are consistent even if the working correlation structure is incorrectly specified, making it somewhat more forgiving when the exact nature of within-cluster correlation is uncertain.
- Some missing data considerations that we'll discuss in the coming weeks.

GEE vs GLMM

Use GLMM When:

- **Subject-Specific Effects Are of Interest:** GLMMs are useful when you're interested in individual-specific effects. They can account for both fixed and random effects, allowing for more nuanced analyses of data with hierarchical or nested structures.
- **Predict of Random Effects:** If there's a need to explicitly predict random effects (e.g., area-level disease rates).
- **Handling Overdispersion in Count Data:** When the variance exceeds the mean in count data, GLMMs with random effects can better model the overdispersion compared to GEE.
- **Complexity vs. Interpretability:** GLMMs can be more computationally intensive and complex to specify correctly than GEEs. However, they offer more flexibility in modeling individual-level variability.