Subject Specific Models

Model Families for Correlated Data

- **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

Subject Specific Models

$$\mathbf{Y}_i|\mathbf{b}_i \sim F_i(\beta, \mathbf{b}_i)$$

$$\begin{aligned}
\mathbf{Y}_i &= n_i \text{ vector of measurements for cluster } i \\
\beta &= \text{parameters common to all clusters} \\
\mathbf{b}_i &= \text{cluster specific effect} \\
F_i(\cdot) &= \text{pre-specified distribution}
\end{aligned}$$

Often assume conditional independence: $Y_{ij} \perp Y_{ik} | \mathbf{b}_i$

- $\Rightarrow F_i$ is product over n_i elements of $\mathbf{Y}_i$
- LMM: $\mathbf{R}_i(\boldsymbol{\theta}) = \sigma^2 \mathbf{I}$

Dealing with Subject Specific Parameters

1. Treat the $\mathbf{b}_i$ as fixed unknown parameters
2. Conditional inference
3. Random effects

Treating $\mathbf{b}_i$ as fixed

If $f_i(\cdot)$ is the density for $\mathbf{Y}_i$, then

$$L(\beta, \mathbf{b}_i) = \prod_i f_i(\mathbf{Y}_i | \mathbf{b}_i, \beta)$$

can be maximized w.r.t. both β and $\mathbf{b}_i$ to obtain "MLE".

Bad Idea

- $\mathbf{b}_i$ is NOT really a parameter
- Neyman & Scott: "MLE" is inconsistent because number of unknowns increases with m
- Example: Logistic regression for matched binary data (Breslow and Day, 1989)
  - Then MLE estimate $\widehat{OR} \to OR^2$ instead of the true OR

Conditional Inference

Idea: Treat $\mathbf{b}_i$ as nuisance, calculate sufficient statistic for $\mathbf{b}_i$, and estimate $\widehat{\beta}$ by maximizing conditional likelihood on the sufficient stat.

Pros:

- No additional assumptions necessary for $\mathbf{b}_i$
- Likelihood has closed form

Cons:

- Need to find a sufficient stat (only in special cases)
- Information on $\mathbf{b}_i$ is lost

Random Effects Models

Subject specific $\rightarrow$ Random effects

Random Effects Models

Idea: Sample subjects i from population AND $\mathbf{b}_i$ from a population of parameters.

- $\mathbf{b}_i$ random vectors with CDF $Q(\mathbf{b}_i)$, mixing distribution
- Estimation and inference based on MLE from marginal model

$$f_i(\mathbf{y}_i|\beta, Q(\cdot)) = \int f_i(\mathbf{y}_i|\mathbf{b}_i, \beta) dQ(\mathbf{b}_i)$$

- Q is typically from parametric family $\Rightarrow$ consistency and normality of β due to ML theory
- Non-parametric MLE (NPMLE) of mixing distribution:
  - Estimate Q by distribution that yields highest likelihood (Kiefer & Wolfowitz, 1956; Laird, 1978; Lindsay, 1983)
  - Computationally expensive
  - Results in discrete estimates necessary (even if true Q is continuous)

Special Cases on Random Effects Models

- LMM: Normal-Normal
- Beta-binomial
- Probit-normal
- Poisson-gamma
- Frailty-model: survival analysis
  - $\lambda_{ij}(t) = \lambda_0(t)e^{X_i\beta + b_i} = \lambda_0(t)b_i^* e^{X_i\beta}$
  - If $b_i^* \sim gamma$, then we have closed form.
- Generalized Linear Mixed Models

Example: Beta-binomial Model

$$Y_{ij} \sim ber(b_i); \qquad b_i \sim beta(\alpha, \beta)$$

$$Z_i = \sum_{j=1}^{n_i} Y_{ij}$$

Then marginal density of Z is

$$\begin{aligned} f_i(z_i|\alpha,\beta) &= \int \begin{pmatrix} n_i \\ z_i \end{pmatrix} b_i^{z_i}(1-b_i)^{n_i-z_i} f(b_i|\alpha,\beta)db_i \\ &= \begin{pmatrix} n_i \\ z_i \end{pmatrix} \frac{B(z_i+\alpha, n_i-z_i+\beta)}{B(\alpha,\beta)} \end{aligned}$$

- Can include covariates as well
- Model is commonly used for over-dispersed binomial data
- Amy Willis recently used this for microbiome data

Generalized Linear Mixed Models (GLMM)

Subject specific $\rightarrow$ Random effects $\rightarrow$ GLMMs

Extending Linear Mixed Models

Recall LMM:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon,$$

where $\mathbf{b} \sim N\{\mathbf{0}, \mathbf{D}(\theta)\}$ and $\epsilon \sim N\{\mathbf{0}, \mathbf{R}(\theta)\}$.

Question: How can we extend linear mixed models to correlated discrete data?

Answer: Add random effects to the GLM linear predictor.

$\Rightarrow$ Generalized Linear Mixed Models (GLMMs)

Reference: Breslow and Clayton, 1993, JASA.

Generalized Linear Mixed Models (GLMMs)

Suppose $Y_1, \cdots, Y_n|\mathbf{b}$ are independent and follow a GLM with

$$\begin{aligned} \mu_i^b &= E(Y_i|\mathbf{b}_i) \\ var(Y_i|\mathbf{b}_i) &= \phi a_i^{-1} v(\mu_i^{\mathbf{b}_i}) \end{aligned}$$

and

$$g(\mu_i^{\mathbf{b}_i}) = \mathbf{X}_i^T \beta + \mathbf{Z}_i^T \mathbf{b}_i$$

where

$$\mathbf{b}_i \sim N\{\mathbf{0}, \mathbf{D}(\theta)\}$$

Quasi-likelihood of $(\beta, \boldsymbol{\theta})$

$$
\begin{aligned}
L(\beta, \boldsymbol{\theta}) &= \int L(Y|\mathbf{b}; \beta) L(\mathbf{b}; \boldsymbol{\theta}) d\mathbf{b} \\
&= e^{\ell(\beta, \boldsymbol{\theta})} \\
&\propto |\mathbf{D}(\boldsymbol{\theta})|^{-\frac{1}{2}} \int exp^{\{\sum_{i=1}^{n} \ell_i(Y_i|\mathbf{b};\beta) - \frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{D}(\boldsymbol{\theta})^{-\mathbf{1}}\mathbf{b}\}} d\mathbf{b}
\end{aligned}
$$

(often no closed form)

where

$$
\begin{aligned}
\ell_i(Y_i|\mathbf{b}; \beta) &\propto \int_{Y_i}^{\mu_i^{\mathbf{b}}} \frac{Y_i - u}{\phi a_i^{-1} v(u)} du \\
&= \text{log quasi-likelihood of } \beta \text{ conditional on } \mathbf{b}
\end{aligned}
$$

Matrix notation

$$g(\mu_i^{\mathbf{b}}) \quad = \quad \mathbf{X}_i^T \beta + \mathbf{Z}_i^T \mathbf{b}$$

where

$$\begin{aligned} \beta &: \quad p \times 1 \text{ vector of regression coefficients.} \\ \mathbf{b} &: \quad q \times 1 \text{ vector of random effects.} \end{aligned}$$

Matrix notation:

$$\begin{aligned} \boldsymbol{\mu}^{\mathbf{b}} &= (\mu_1^{\mathbf{b}}, \mu_2^{\mathbf{b}}, \cdots, \mu_n^{\mathbf{b}}) \\ g(\boldsymbol{\mu}^{\mathbf{b}}) &= \{g(\mu_1^{\mathbf{b}}), g(\mu_2^{\mathbf{b}}) \cdots, g(\mu_n^{\mathbf{b}})\} \end{aligned}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1^T \\ \vdots \\ \mathbf{X}_n^T \end{pmatrix} \quad , \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1^T \\ \vdots \\ \mathbf{Z}_n^T \end{pmatrix}$$

$\Rightarrow$

$$g(\boldsymbol{\mu}^{\mathbf{b}}) \quad = \quad \mathbf{X}\beta + \mathbf{Z}\mathbf{b}$$

Example 1: LMM

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon},$$

where

$$\begin{aligned} \mathbf{b} &\sim N\{\mathbf{0}, \mathbf{D}(\boldsymbol{\theta})\} \\ \boldsymbol{\epsilon} &\sim N\{\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I}\} \end{aligned}$$

$\Leftrightarrow$

$$\begin{aligned} Y_1, \cdots, Y_n|\mathbf{b} \quad \text{indep} &\sim N(\mu_i^{\mathbf{b}}, \sigma^2) \\ \boldsymbol{\mu}^{\mathbf{b}} &= \mathbf{X}\beta + \mathbf{Z}\mathbf{b} \text{ (identity link)} \end{aligned}$$

with

$$\begin{aligned} \mu_i^{\mathbf{b}} &= E(Y_i|\mathbf{b}) \\ var(Y_i|\mathbf{b}) &= \sigma^2 \qquad\qquad (\phi = \sigma^2) \end{aligned}$$

Example 2: GLMM for clustered/longitudinal data

- m clusters
- $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{in_i})^T$: observations for cluster i $(i = 1, \cdots, m)$
- $\mathbf{Y}_1, \cdots, \mathbf{Y}_m$ are independent.
- $\mu_{ij}^{\mathbf{b}} = E(Y_{ij}|b_i)$ and $var(Y_{ij}|b_i) = \phi a_{ij}^{-1} v(\mu_{ij}^{\mathbf{b}})$

$$\Downarrow$$

$$g(\mu_{ij}^{\mathbf{b}}) = \mathbf{X}_{ij}^T \beta + \mathbf{Z}_{ij}^T \mathbf{b}_i$$

where

$$\mathbf{b}_i \sim N\{0, \mathbf{D}_0(\boldsymbol{\theta})\} \quad = \text{cluster-specific random effects.}$$

Example 2: What do the matrices look like?

$\Rightarrow$

$$\boldsymbol{\mu}_i^{\mathbf{b}} = \begin{pmatrix} \mu_{i1}^{\mathbf{b}} \\ \vdots \\ \mu_{in_i}^{\mathbf{b}} \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} \mathbf{X}_{i1}^T \\ \vdots \\ \mathbf{X}_{in_i}^T \end{pmatrix}, \quad \mathbf{Z}_i = \begin{pmatrix} \mathbf{Z}_{i1}^T \\ \vdots \\ \mathbf{Z}_{in_i}^T \end{pmatrix}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1^{\mathbf{b}} \\ \vdots \\ \boldsymbol{\mu}_m^{\mathbf{b}} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{Z}_m \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

$\Rightarrow$

$$\begin{aligned} g(\boldsymbol{\mu}_i^{\mathbf{b}}) &= \mathbf{X_i}\beta + \mathbf{Z_i b_i} \\ g(\boldsymbol{\mu}^{\mathbf{b}}) &= \mathbf{X}\beta + \mathbf{Zb} \end{aligned}$$

where

$$\mathbf{b} \sim N\{\mathbf{0}, \mathbf{D}(\theta)\} \quad , \quad \mathbf{D}(\theta) = \begin{pmatrix} \mathbf{D}_0(\theta) & & \\ & \ddots & \\ & & \mathbf{D}_0(\theta) \end{pmatrix}.$$

Example 2: Specific sub-examples

Examples:

- Random intercept:

$$g(\mu_{ij}^{\mathbf{b}}) = \mathbf{X}_{ij}^T \beta + b_i$$

where $b_i \overset{i.i.d}{\sim} N(0, \theta)$.

- Random intercept and slope:

$$g(\mu_{ij}^{\mathbf{b}}) = \mathbf{X}_{ij}^T \beta + b_{1i} + t_{ij} b_{2i}$$

where

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \overset{i.i.d}{\sim} N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{D}_0(\boldsymbol{\theta}) \right\}$$

This is familiar from discussion of LMM.

Longitudinal Example A: Clustered normal data (CD4 counts)

Assume $Y_{ij}|\boldsymbol{b}_i \sim N(\mu_{ij}^{b}, \sigma^2)$ and

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

$$\Updownarrow$$

LMM:

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

where $\epsilon_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$.

note: CD4 is is usually highly skewed. Y is either $log(CD4)$ or $\sqrt{CD4}$

Longitudinal Example B: Clustered binomial data (infection (yes/no))

Assume $N * Y_{ij}|\boldsymbol{b}_i \sim (overdispersed)\ Binomial(N, \mu_{ij}^b)$ and

$$logit(\mu_{ij}^{\mathbf{b}}) = \mathbf{X}_{ij}^T \beta + \mathbf{Z}_{ij}^T \mathbf{b}_i$$

$$\Downarrow$$

$$\begin{aligned} E(Y_{ij}|\mathbf{b}_i) &= \mu_{ij}^{\mathbf{b}} \\ var(Y_{ij}|\mathbf{b}_i) &= \phi N^{-1} \mu_{ij}^{\mathbf{b}} (1 - \mu_{ij}^{\mathbf{b}}) \end{aligned}$$

Remarks:

- if you have Bernoulli data, ϕ is always 1.
- If you have grouped data, your ϕ may be 1. ($> 1 \rightarrow$ overdispersion.

Longitudinal Example C: Clustered count data (# of seizures)

Assume $Y_{ij}|\mathbf{b_i} \sim$(*overdispersed*) $Poisson(\mu_{ij}^{b})$ and

$$ln(\mu_{ij}^{\mathbf{b}}) = \mathbf{X}_{ij}^{T}\beta + \mathbf{Z}_{ij}^{T}\mathbf{b}_i$$

or

$$ln(\mu_{ij}^{\mathbf{b}}) = ln(N_{ij}) + \mathbf{X}_{ij}^{T}\beta + \mathbf{Z}_{ij}^{T}\mathbf{b}_i$$

where $ln(N_{ij})$ is the offset if $Y_{ij} = \#$ of cases.

$$\Downarrow$$

$$\begin{aligned} E(Y_{ij}|\mathbf{b}_i) &= \mu_{ij}^{\mathbf{b}} \\ var(Y_{ij}|\mathbf{b}_i) &= \phi\mu_{ij}^{\mathbf{b}} \end{aligned}$$

Note: Y_{ij} is the incidence rate.

Example 3: Hierarchical Data

Center	Physician	Patient	Disease
1	1	1	1
1	1	2	0
1	2	3	0
1	2	4	0
2	3	5	1
2	3	6	1
2	4	7	0
2	4	8	1
⋮	⋮	⋮	⋮

We have three levels in this data. But for longitudinal data, we only have two levels.

Example 3: Hierarchical Data (2)

Model:

$$logit(\boldsymbol{\mu}^b) = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2$$

where
$\mathbf{b}_1 = (b_{11}, \cdots, b_{1q_1})^T$ is a $q_1 \times 1$ vector of random center effect and $\mathbf{b}_1 \sim N(0, \theta_1 \mathbf{I}_{q_1})$;
$\mathbf{b}_2 = (b_{21}, \cdots, b_{2q_2})^T$ is a $q_2 \times 1$ vector of random physician effect and $\mathbf{b}_2 \sim N(0, \theta_2 \mathbf{I}_{q_2})$.
$\Leftrightarrow$

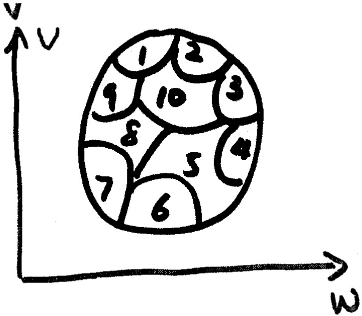
$$\begin{aligned} logit(\boldsymbol{\mu}_\mathbf{i}^\mathbf{b}) &= \mathbf{X}_i^T\beta + \mathbf{Z}_{i1}^T\mathbf{b}_1 + \mathbf{Z}_{i2}^T\mathbf{b}_2 \\ &= \mathbf{X}_i^T\beta + \mathbf{Z}_i^T\mathbf{b} \end{aligned}$$

where

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Z}_{i1} \\ \mathbf{Z}_{i2} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \theta_1 \mathbf{I}_{q_1} & \mathbf{0} \\ \mathbf{0} & \theta_2 \mathbf{I}_{q_2} \end{pmatrix} \right\}$$

Example 4: Spatial Data

county	# of D	PY	**b**
1	Y_1	N_1	$\mathbf{b}_1$
2	Y_2	N_2	$\mathbf{b}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	Y_n	N_n	$\mathbf{b}_n$



Example 4: Spatial Data (2)

$$
\begin{aligned}
& ln(\boldsymbol{\mu}_{\boldsymbol{i}}^{\mathbf{b}}) = ln(N_i) + \mathbf{X}_i^T \beta + b_i \\
\Leftrightarrow \quad & ln(\boldsymbol{\mu}) = ln(\mathbf{N}) + \mathbf{X}\beta + \mathbf{b}
\end{aligned}
$$

where

$$
\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \sim N \left\{ \mathbf{0}, \begin{pmatrix} 1 & \rho^{|\mathbf{d}_1 - \mathbf{d}_2|} & \rho^{|\mathbf{d}_1 - \mathbf{d}_3|} & \dots & \rho^{|\mathbf{d}_1 - \mathbf{d}_n|} \\ & 1 & \rho^{|\mathbf{d}_2 - \mathbf{d}_3|} & \dots & \rho^{|\mathbf{d}_2 - \mathbf{d_n}|} \\ & & 1 & & \\ & * & & \ddots & \\ & & & & 1 \end{pmatrix} \right\}
$$

and $\mathbf{d}_i = (w_i, v_i)$ and w_i is the longitude and v_i is the latitude.

Remarks

- Subject-specific models $\rightarrow$ Random effects models $\rightarrow$ GLMM
- GLMM is a natural combination of LMM and GLM.
- Can use a multi-stage formulation (useful for Bayesian inference)
- GLMM is a possible alternative approach to GEEs for modeling longitudinal data:

Question: Does β in GLMM have the same interpretation as β in GEE?