LMMs: Matrix Notation

$\mathbf{Y}$: $n \times 1$ outcome vector
$\mathbf{X}$: $n \times p$ design matrix for fixed effects
$\mathbf{Z}$: $n \times q$ design matrix for random effects
β : $p \times 1$ regression coefficient vector
$\mathbf{b}$: $q \times 1$ random effect vector
$\boldsymbol{\theta}$: $c \times 1$ variance component

Model: (Harville, 1977, JASA, pp320-340)

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

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

Ex 1: Clustered Data (Laird and Ware, 1982, Biometrics)

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \text{ for cluster i.}$$

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix}, \qquad \mathbf{X}_i : n_i \times p, \qquad \mathbf{Z}_i : n_i \times q_i.$$

$$\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix} \sim N\{\mathbf{0}, \mathbf{R_0}(\boldsymbol{\theta})\} \text{ , } \mathbf{b}_i \sim N\{\mathbf{0}, \mathbf{D_0}(\theta)\}.$$

Ex 1: Clustered Data - Random Intercept/Slope

Random Intercept:

$$\mathbf{Z}_i = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{b}_i = b_i.$$

Random Intercept & Slope:

$$\mathbf{Z}_i = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix}, \mathbf{b}_i = \begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix}.$$

Ex 1: Clustered Data - Random Intercept/Slope

Then $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix}$, $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{pmatrix}$, $\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & & & \mathbf{0} \\ & \mathbf{Z}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{Z}_m \end{pmatrix}$.

$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_m \end{pmatrix}$, $\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}$, $\mathbf{D}(\boldsymbol{\theta}) = cov(\mathbf{b}) = diag\{\mathbf{D}_0(\boldsymbol{\theta})\}$,

and $\mathbf{R}(\boldsymbol{\theta}) = cov(\boldsymbol{\epsilon}) = diag\{\mathbf{R}_0(\boldsymbol{\theta})\}$.

Ex 2: Hierarchical Data

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{Z}_2\mathbf{b}_2 + \cdots + \mathbf{Z}_c\mathbf{b}_c + \epsilon,$$

- $\mathbf{b}_j$ is a $q_j \times 1$ vector of random effects.
- $\mathbf{b}_j \sim N(\mathbf{0}, \theta_j I_{q_j})$.
- $\mathbf{b}_1, \cdots, \mathbf{b}_c$ are independent.

Ex 2: Hierarchical Data

$\mathbf{b}_1-$ center effect, $\mathbf{b}_2-$ patient effect, suppose each patient has 2 observations, then

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & & \\ \vdots & & \\ 1 & & \\ & 1 & \\ & \vdots & \\ & 1 & \\ & & \ddots \end{pmatrix}, \mathbf{b}_1 = \begin{pmatrix} b_{11} \\ b_{12} \\ \vdots \\ \vdots \end{pmatrix}, \quad \begin{array}{lcl} b_{11} & : & \text{center 1 effect} \\ b_{12} & : & \text{center 2 effect} \end{array}$$

$$\mathbf{Z}_2 = \begin{pmatrix} 1 & & \\ 1 & & \\ & 1 & \\ & 1 & \\ & & \ddots \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} b_{21} \\ b_{22} \\ \vdots \\ \vdots \end{pmatrix}, \quad \begin{array}{lcl} b_{21} & : & \text{patient 1 effect} \\ b_{22} & : & \text{patient 2 effect} \end{array}$$

Objectives with LMMs

- Inference on β - population mean effect.
- Inference on $\boldsymbol{\theta}$ - correlation.
- Estimate (predict) the random effects **b**
  - subject-specific curves
  - center/subject-specific curves

Estimation in LMMs

Recall:

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

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

Marginal distribution of Y:

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V})$$

where $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \mathbf{R}$.

Log likelihood function of (β, θ):

$$\ell(\beta, \theta) = -\frac{1}{2} ln|\mathbf{V}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\beta)$$

Side Note

We don't usually like to work with LMMs in Matrix form:

- $n \times n$ matrices are unnecessarily cumbersome; we often prefer to do things via summations
- Ex: Suppose we have m $= 100$ people and $n_i = 10$ obs/person. If we start from the likelihood directly, $\mathbf{V}$ is 1000×1000. Instead, we can work with block diagonal structure, s.t. we have only v_i as 10×10 which is much easier to take invert.
- The reason we want to write the likelihood this way, is because once we moved away from the normal case, we no longer have the nice orthogonal property of the parameters of mean and variance.

Log likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\theta})$:

$$L(\beta, \boldsymbol{\theta}) = \int L(\mathbf{Y}|\mathbf{b})L(\mathbf{b})d\mathbf{b}$$

where $L(\mathbf{Y}|\mathbf{b}) \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{b}, \mathbf{R})$ and $\mathbf{b} \sim N\{\mathbf{0}, \mathbf{D}(\theta)\}$.

Then

$$L(\beta, \boldsymbol{\theta}) = \frac{1}{|\mathbf{R}|^{\frac{1}{2}}|\mathbf{D}|^{\frac{1}{2}}} \int e^{-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta-\mathbf{Z}\mathbf{b})^T\mathbf{R}^{-1}(\mathbf{Y}-\mathbf{X}\beta-\mathbf{Z}\mathbf{b})-\frac{1}{2}\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}} d\mathbf{b}$$

Score Equations:

Facts:

$$\frac{\partial \mathbf{V}^{-1}}{\partial \theta_j} = -\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}$$

$$\frac{\partial ln|\mathbf{V}|}{\partial \theta_j} = tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j})$$

Score Equations of β and $\boldsymbol{\theta}$:

$$\begin{aligned} U_\beta(\beta, \boldsymbol{\theta}) &= \frac{\partial \ell}{\partial \beta} = \mathbf{X}^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta) = 0 \\ &\Rightarrow \hat{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}, \text{ weighted LS estimator} \end{aligned}$$

$$\begin{aligned} U_{\theta_j}(\beta, \boldsymbol{\theta}) &= -\frac{1}{2}tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}) + \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta) \\ \text{set} &= 0 \end{aligned}$$

Note: we often see this written in terms of sums.

Information matrix:

$$
\begin{aligned}
I_{\beta\beta} &= \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} \\
I_{\beta\theta_j} &= E\{\mathbf{X}^T\frac{\partial \mathbf{V}^{-1}}{\partial\theta_j}(\mathbf{Y}-\mathbf{X}\beta)\} = 0 \\
I_{\theta_j\theta_k} &= \frac{1}{2}tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial\theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial\theta_k})
\end{aligned}
$$

$$
\Rightarrow I = \begin{pmatrix} \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \{\frac{1}{2}tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial\theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial\theta_k})\} \end{pmatrix}
$$

Derivations

$$\frac{\partial \ell}{\partial \theta_j} = -\frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}) + \frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)^T\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\beta)$$

$$\Rightarrow \quad \frac{\partial^2 \ell}{\partial \theta_j \partial \theta_k} = -\frac{1}{2} tr(-\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}) - \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial^2 \mathbf{V}}{\partial \theta_j \partial \theta_k})$$

$$-2 \times \frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)^T\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\beta)$$

$$+\frac{1}{2}(\mathbf{Y}-\mathbf{X}\beta)^T\mathbf{V}^{-1}\frac{\partial^2 \mathbf{V}}{\partial \theta_j \partial \theta_k}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\beta)$$

$$\Rightarrow \quad E(-\frac{\partial^2 \ell}{\partial \theta_j \partial \theta_k}) = -\frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}) + \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial^2 \mathbf{V}}{\partial \theta_j \partial \theta_k})$$

$$+tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}) - \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial^2 \mathbf{V}}{\partial \theta_j \partial \theta_k})$$

$$\Rightarrow \quad I_{\theta_j \theta_k} = \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k})$$

Estimation

LMM:

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

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

Log likelihood function of $(\beta, \boldsymbol{\theta})$:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$$

$$\max_{\boldsymbol{\beta}, \boldsymbol{\theta}} \ell(\boldsymbol{\beta}, \boldsymbol{\theta}) \Rightarrow (\hat{\beta}, \hat{\boldsymbol{\theta}})$$

Estimation of Random Effects

Question: How can we estimate the random effects $\boldsymbol{b}$?

- Subject-specific growth curves?
- Subject-specific CD4 count trajectories?

Answer: Best Linear Unbiased Predictors (BLUPs) / Best Linear Unbiased Estimators (BLUEs)

References:

- Harville (1977, JASA)
- Robinson (1991, Stat Science)

Best Linear Unbiased Estimator/Predictor

- **Objective:** To construct a "best" estimator of $\boldsymbol{\lambda}_1^T \beta + \boldsymbol{\lambda}_2^T \boldsymbol{b}$ for any $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2$ given $\boldsymbol{\theta}$.

- **Form of the estimator:** $\mathbf{C}^\mathbf{T}\mathbf{Y}$ (linear in $\mathbf{Y}$).

- **Properties of the estimator:**
  1. Unbiased: $E(\mathbf{C}^\mathbf{T}\mathbf{Y}) = \boldsymbol{\lambda}_1^T \beta$
  2. Best: Minimize the unconditional mean square error:

$$\min_{\mathbf{C}} E[\mathbf{C}^\mathbf{T}\mathbf{Y} - \boldsymbol{\lambda}_1^T \beta - \boldsymbol{\lambda}_2^T \boldsymbol{b}]^2$$

Such a $(\hat{\beta}, \hat{\mathbf{b}})$ is called the best linear unbiased estimator (predictor) (BLUE/BLUP) of $(\beta, \mathbf{b})$.

Mixed Model Equations

$(\hat{\beta}, \hat{\mathbf{b}})$ satisfies the Mixed Model Equations:

$$\begin{pmatrix} \mathbf{X^T R^{-1} X} & \mathbf{X^T R^{-1} Z} \\ \mathbf{Z^T R^{-1} X} & \mathbf{Z^T R^{-1} Z} + \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{X^T R^{-1} Y} \\ \mathbf{Z^T R^{-1} Y} \end{pmatrix}$$