Applied Statistical Methods II

Estimation in Mixed effects model

A general mixed effects model

Assume $\boldsymbol{Y} \in \mathbb{R}^n$ and let $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ be a design matrix. Suppose

$$E(\boldsymbol{Y}) = \boldsymbol{X}\beta, \quad \mathrm{Var}(\boldsymbol{Y}) = \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s$$

$\boldsymbol{X}$, $\boldsymbol{B}_s$ are known and non-random. Goal: estimate β, and maybe θ_s.

- Examples:
  - $Y_{ij} = \mu + \delta_i + \epsilon_{ij}$, $\delta_i \overset{i.i.d}{\sim} (0, \sigma_\delta^2)$ and $\epsilon_{ij} \overset{i.i.d}{\sim} (0, \sigma^2)$. Here, i is individual, j is replicate.
    - $\boldsymbol{B}_1 = I_n$,
      $$[\boldsymbol{B}_2]_{rs} = \begin{cases} 1 & \text{samples } r, s \text{ come from same individual} \\ 0 & \text{o/w} \end{cases}.$$
  - $Y_{ij} = \tau_i + \epsilon_{ij}$, $i = 1, \ldots, r$ is trt index, $j = 1, \ldots, n_i$ is replicate, τ_i is a fixed effect, $\epsilon_{ij} \sim (0, \sigma_i^2)$.
    - $\boldsymbol{B}_i = \mathrm{diag}\left(\mathbf{0}_{n_1}, \ldots, \mathbf{0}_{n_{i-1}}, \mathbf{1}_{n_i}, \mathbf{0}_{n_{i+1}}, \ldots, \mathbf{0}_{n_r}\right)$.
  - And many others...
- Question: how do we fit this??

A general mixed effects model

Assume $\boldsymbol{Y} \in \mathbb{R}^n$ and let $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ be a design matrix. Suppose

$$E(\boldsymbol{Y}) = \boldsymbol{X}\beta, \quad \mathrm{Var}(\boldsymbol{Y}) = \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s$$

$\boldsymbol{X}$, $\boldsymbol{B}_s$ are known and non-random. Goal: estimate β, and maybe θ_s.

- Examples:
  - $Y_{ij} = \mu + \delta_i + \epsilon_{ij}$, $\delta_i \overset{i.i.d}{\sim} (0, \sigma_\delta^2)$ and $\epsilon_{ij} \overset{i.i.d}{\sim} (0, \sigma^2)$. Here, i is individual, j is replicate.
    - $\boldsymbol{B}_1 = I_n$,
      $$[\boldsymbol{B}_2]_{rs} = \begin{cases} 1 & \text{samples } r, s \text{ come from same individual} \\ 0 & \text{o/w} \end{cases}.$$
  - $Y_{ij} = \tau_i + \epsilon_{ij}$, $i = 1, \ldots, r$ is trt index, $j = 1, \ldots, n_i$ is replicate, τ_i is a fixed effect, $\epsilon_{ij} \sim (0, \sigma_i^2)$.
    - $\boldsymbol{B}_i = \mathrm{diag}\left(\boldsymbol{0}_{n_1}, \ldots, \boldsymbol{0}_{n_{i-1}}, \boldsymbol{1}_{n_i}, \boldsymbol{0}_{n_{i+1}}, \ldots, \boldsymbol{0}_{n_r}\right)$.
  - And many others...
- Question: how do we fit this??

Maximum quasi-likelihood

$$E(\boldsymbol{Y}) = \boldsymbol{X}\beta, \quad \mathrm{Var}(\boldsymbol{Y}) = \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s$$

- Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_b)^T$, $\boldsymbol{V}_\theta = \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s$
- Maximize likelihood, assume $\boldsymbol{Y}$ is normally distributed:

$$\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\theta}\in\Theta\subset\mathbb{R}^b}{\mathrm{argmax}} \ \ell(\boldsymbol{\beta}, \boldsymbol{\theta})$$

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2}\log\{\det(\boldsymbol{V}_\theta)\} - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\beta)^T \boldsymbol{V}_\theta^{-1} (\boldsymbol{Y} - \boldsymbol{X}\beta)$$

Maximum quasi-likelihood (cont.)

$$\hat{\beta}, \hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p,\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^b}{\operatorname{argmax}} \ \ell(\boldsymbol{\beta}, \boldsymbol{\theta})$$

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} \log \{\det(\boldsymbol{V}_\theta)\} - \frac{1}{2} (\boldsymbol{Y} - \boldsymbol{X}\beta)^T \boldsymbol{V}_\theta^{-1} (\boldsymbol{Y} - \boldsymbol{X}\beta)$$

- Can solve for $\hat{\beta}$ for fixed $\boldsymbol{V}_\theta$:

$$\nabla_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{X}^T \boldsymbol{V}_\theta^{-1} (\boldsymbol{Y} - \boldsymbol{X}\beta)$$

$$\hat{\beta}_\theta = \left(\boldsymbol{X}^T \boldsymbol{V}_\theta^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{V}_\theta^{-1} \boldsymbol{Y}$$

- To determine $\hat{\theta}$, maximize **profile likelihood**

$$\tilde{\ell}(\beta) = -\frac{1}{2} \log \{\det(\boldsymbol{V}_\theta)\} - \frac{1}{2} \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_\theta\right) \boldsymbol{V}_\theta^{-1} \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_\theta\right)$$

- Maximize this with gradient descent/other numerical method

Maximum quasi-likelihood (cont.)

$$\hat{\beta}, \hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p,\, \boldsymbol{\theta}\in\Theta\subset\mathbb{R}^b}{\text{argmax}} \ \ell(\boldsymbol{\beta}, \boldsymbol{\theta})$$

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2}\log\{\det(\boldsymbol{V}_\theta)\} - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\beta)^T \boldsymbol{V}_\theta^{-1}(\boldsymbol{Y} - \boldsymbol{X}\beta)$$

- Can solve for $\hat{\beta}$ for fixed $\boldsymbol{V}_\theta$:

$$\nabla_{\boldsymbol{\beta}}\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{X}^T \boldsymbol{V}_\theta^{-1}(\boldsymbol{Y} - \boldsymbol{X}\beta)$$

$$\hat{\beta}_\theta = \left(\boldsymbol{X}^T \boldsymbol{V}_\theta^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{V}_\theta^{-1} \boldsymbol{Y}$$

- To determine $\hat{\boldsymbol{\theta}}$, maximize **profile likelihood**

$$\tilde{\ell}(\beta) = -\frac{1}{2}\log\{\det(\boldsymbol{V}_\theta)\} - \frac{1}{2}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_\theta\right) \boldsymbol{V}_\theta^{-1} \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_\theta\right)$$

- Maximize this with gradient descent/other numerical method

The problem with maximum quasi-likelihood

$$E(\boldsymbol{Y}) = \boldsymbol{X}\beta, \quad \mathrm{Var}(\boldsymbol{Y}) = \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s$$

- Consider standard regression model: $b = 1$, $\boldsymbol{B}_1 = I_n$, $\theta_1 = \sigma^2$.
- Using ML: $\hat{\sigma}^2 = n^{-1}(\boldsymbol{Y} - \boldsymbol{X}\underbrace{\hat{\beta}}_{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}})^T\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}\right)$
- Is $\hat{\sigma}^2$ an unbiased for estimate for σ^2?
- Problem: ML does not account for the uncertainty in the estimate for β when estimating σ^2!

The problem with maximum quasi-likelihood

$$E(\boldsymbol{Y}) = \boldsymbol{X}\beta, \quad \mathrm{Var}(\boldsymbol{Y}) = \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s$$

- Consider standard regression model: $b = 1$, $\boldsymbol{B}_1 = I_n$, $\theta_1 = \sigma^2$.
- Using ML: $\hat{\sigma}^2 = n^{-1}(\boldsymbol{Y} - \boldsymbol{X}\underbrace{\hat{\beta}}_{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}})^T\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}\right)$
- Is $\hat{\sigma}^2$ an unbiased for estimate for σ^2?
- Problem: ML does not account for the uncertainty in the estimate for β when estimating σ^2!

The problem with maximum quasi-likelihood

$$E(\boldsymbol{Y}) = \boldsymbol{X}\beta, \quad \mathrm{Var}(\boldsymbol{Y}) = \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s$$

- Consider standard regression model: $b = 1$, $\boldsymbol{B}_1 = I_n$, $\theta_1 = \sigma^2$.
- Using ML: $\hat{\sigma}^2 = n^{-1}(\boldsymbol{Y} - \boldsymbol{X}\underbrace{\hat{\beta}}_{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}})^T\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}\right)$
- Is $\hat{\sigma}^2$ an unbiased for estimate for σ^2?
- Problem: ML does not account for the uncertainty in the estimate for β when estimating σ^2!

Restricted maximum likelihood (REML)

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\boldsymbol{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

- Idea behind REML: make the estimation of $\theta_1, \ldots, \theta_b$ invariant to β
- Solution: regress out $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ from $\boldsymbol{Y}$:
  - Define $\boldsymbol{A} \in \mathbb{R}^{n \times (n-p)}$ s.t. the columns of $\boldsymbol{A}$ form a basis for $\ker\left(\boldsymbol{X}^T\right)$
  - i.e. $\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T = I_n - \boldsymbol{H}$
  - Multiply $\boldsymbol{Y}$ by $\boldsymbol{A}^T\boldsymbol{Y}$ (akin to regressing out $\boldsymbol{X}$ from $\boldsymbol{Y}$):

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^T\boldsymbol{Y} = \underbrace{\boldsymbol{A}^T\boldsymbol{X}}_{=\boldsymbol{0}}\beta + \boldsymbol{A}^T\epsilon \sim \left(\boldsymbol{0}_{n-p}, \sum_{s=1}^{b} \theta_s \boldsymbol{A}^T\boldsymbol{B}_s\boldsymbol{A}\right)$$

- Estimate θ by running ML on $\tilde{\boldsymbol{Y}}$, which now has mean $\boldsymbol{0}_{n-p}$ and does not depend on β!
- Set $\hat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{V}_{\hat{\theta}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{V}_{\hat{\theta}}^{-1}\boldsymbol{Y}$, the GLS estimator

Restricted maximum likelihood (REML)

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\boldsymbol{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

- Idea behind REML: make the estimation of $\theta_1, \ldots, \theta_b$ invariant to β
- Solution: regress out $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ from $\boldsymbol{Y}$:
  - Define $\boldsymbol{A} \in \mathbb{R}^{n \times (n-p)}$ s.t. the columns of $\boldsymbol{A}$ form a basis for $\ker\left(\boldsymbol{X}^T\right)$
  - i.e. $\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T = I_n - \boldsymbol{H}$
  - Multiply $\boldsymbol{Y}$ by $\boldsymbol{A}^T\boldsymbol{Y}$ (akin to regressing out $\boldsymbol{X}$ from $\boldsymbol{Y}$):

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^T\boldsymbol{Y} = \underbrace{\boldsymbol{A}^T\boldsymbol{X}}_{=\boldsymbol{0}}\beta + \boldsymbol{A}^T\epsilon \sim \left(\boldsymbol{0}_{n-p}, \sum_{s=1}^{b} \theta_s \boldsymbol{A}^T\boldsymbol{B}_s\boldsymbol{A}\right)$$

- Estimate θ by running ML on $\tilde{\boldsymbol{Y}}$, which now has mean $\boldsymbol{0}_{n-p}$ and does not depend on β!
- Set $\hat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{V}_{\hat{\theta}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{V}_{\hat{\theta}}^{-1}\boldsymbol{Y}$, the GLS estimator

Restricted maximum likelihood (REML)

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\mathbf{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

- Idea behind REML: make the estimation of $\theta_1, \ldots, \theta_b$ invariant to β
- Solution: regress out $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ from $\boldsymbol{Y}$:
  - Define $\boldsymbol{A} \in \mathbb{R}^{n \times (n-p)}$ s.t. the columns of $\boldsymbol{A}$ form a basis for $\ker\left(\boldsymbol{X}^T\right)$
  - i.e. $\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T = I_n - \boldsymbol{H}$
  - Multiply $\boldsymbol{Y}$ by $\boldsymbol{A}^T\boldsymbol{Y}$ (akin to regressing out $\boldsymbol{X}$ from $\boldsymbol{Y}$):

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^T\boldsymbol{Y} = \underbrace{\boldsymbol{A}^T\boldsymbol{X}}_{=\mathbf{0}}\beta + \boldsymbol{A}^T\epsilon \sim \left(\mathbf{0}_{n-p}, \sum_{s=1}^{b} \theta_s \boldsymbol{A}^T\boldsymbol{B}_s\boldsymbol{A}\right)$$

- Estimate $\boldsymbol{\theta}$ by running ML on $\tilde{\boldsymbol{Y}}$, which now has mean $\mathbf{0}_{n-p}$ and does not depend on β!
- Set $\hat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{V}_{\hat{\theta}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{V}_{\hat{\theta}}^{-1}\boldsymbol{Y}$, the GLS estimator

Restricted maximum likelihood (REML)

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\mathbf{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

- Idea behind REML: make the estimation of $\theta_1, \ldots, \theta_b$ invariant to β
- Solution: regress out $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ from $\boldsymbol{Y}$:
  - Define $\boldsymbol{A} \in \mathbb{R}^{n \times (n-p)}$ s.t. the columns of $\boldsymbol{A}$ form a basis for ker $\left(\boldsymbol{X}^T\right)$
  - i.e. $\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T = I_n - \boldsymbol{H}$
  - Multiply $\boldsymbol{Y}$ by $\boldsymbol{A}^T\boldsymbol{Y}$ (akin to regressing out $\boldsymbol{X}$ from $\boldsymbol{Y}$):

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^T\boldsymbol{Y} = \underbrace{\boldsymbol{A}^T\boldsymbol{X}}_{=\mathbf{0}}\beta + \boldsymbol{A}^T\epsilon \sim \left(\mathbf{0}_{n-p}, \sum_{s=1}^{b} \theta_s \boldsymbol{A}^T\boldsymbol{B}_s\boldsymbol{A}\right)$$

- Estimate θ by running ML on $\tilde{\boldsymbol{Y}}$, which now has mean $\mathbf{0}_{n-p}$ and does not depend on β!
- Set $\hat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{V}_{\hat{\theta}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{V}_{\hat{\theta}}^{-1}\boldsymbol{Y}$, the GLS estimator

REML objective function

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\mathbf{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

- $\tilde{\boldsymbol{Y}} = \boldsymbol{A}^T \boldsymbol{Y} \sim \left(\mathbf{0}_{n-p}, \sum_{s=1}^{b} \theta_s \boldsymbol{A}^T \boldsymbol{B}_s \boldsymbol{A}\right)$
- Estimate $\boldsymbol{\theta}$ assuming $\tilde{\boldsymbol{Y}}$ in normally distributed (maximum quasi-likelihood):

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p}{\operatorname{argmax}} \, \ell_{REML}(\boldsymbol{\theta})$$

$$\ell_{REML}(\boldsymbol{\theta}) = -\frac{1}{2} \log\left\{\det\left(\sum_s \theta_s \boldsymbol{A}^T \boldsymbol{B}_s \boldsymbol{A}\right)\right\} - \frac{1}{2} \tilde{\boldsymbol{Y}}^T \left(\sum_s \theta_s \boldsymbol{A}^T \boldsymbol{B}_s \boldsymbol{A}\right)^{-1} \tilde{\boldsymbol{Y}}$$

- $\Theta \subset \mathbb{R}^p$ is the parameter set. It is typically a **convex set.**
  - OLS example: $b = 1$, $\boldsymbol{B}_1 = I_n$, and $\Theta = \{\theta \in \mathbb{R} : \theta \geq 0\}$

REML: choice of A

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\mathbf{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^T\boldsymbol{Y} \sim \left(\mathbf{0}_{n-p}, \sum_{s=1}^{b} \theta_s \boldsymbol{A}^T\boldsymbol{B}_s\boldsymbol{A}\right)$$

- Recall we had to choose $\boldsymbol{A} \in \mathbb{R}^{n\times(n-p)}$ s.t. $\boldsymbol{X}^T\boldsymbol{A}$.
- Does estimate for θ depend on choice of $\boldsymbol{A}$?
- If $\boldsymbol{A}_1, \boldsymbol{A}_2 \in \mathbb{R}^{n\times(n-p)}$ form bases for $\ker\left(\boldsymbol{X}^T\right)$, we must have $\boldsymbol{A}_2 = \boldsymbol{A}_1\boldsymbol{R}$ for some invertible $\boldsymbol{R} \in \mathbb{R}^{(n-p)\times(n-p)}$
- Using REML with $\boldsymbol{A}_2$:

$$\ell_{A_2}(\theta) = -\frac{1}{2}\log\{\det(\sum_s \theta_s \boldsymbol{A}_2^T\boldsymbol{B}_s\boldsymbol{A}_2)\} - \frac{1}{2}\boldsymbol{Y}^T\boldsymbol{A}_2(\sum_s \theta_s \boldsymbol{A}_2^T\boldsymbol{B}_s\boldsymbol{A}_2)^{-1}\boldsymbol{A}_2^T\boldsymbol{Y}$$

$$= \cdots = C + \ell_{A_1}(\theta)$$

where C does not depend on θ.

REML: choice of $\boldsymbol{A}$

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\mathbf{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^T \boldsymbol{Y} \sim \left(\mathbf{0}_{n-p}, \sum_{s=1}^{b} \theta_s \boldsymbol{A}^T \boldsymbol{B}_s \boldsymbol{A}\right)$$

- Recall we had to choose $\boldsymbol{A} \in \mathbb{R}^{n\times(n-p)}$ s.t. $\boldsymbol{X}^T\boldsymbol{A}$.
- Does estimate for $\boldsymbol{\theta}$ depend on choice of $\boldsymbol{A}$?
- If $\boldsymbol{A}_1, \boldsymbol{A}_2 \in \mathbb{R}^{n\times(n-p)}$ form bases for $\ker\left(\boldsymbol{X}^T\right)$, we must have $\boldsymbol{A}_2 = \boldsymbol{A}_1\boldsymbol{R}$ for some invertible $\boldsymbol{R} \in \mathbb{R}^{(n-p)\times(n-p)}$
- Using REML with $\boldsymbol{A}_2$:

$$\ell_{A_2}(\boldsymbol{\theta}) = -\frac{1}{2}\log\{\det(\sum_s \theta_s \boldsymbol{A}_2^T \boldsymbol{B}_s \boldsymbol{A}_2)\} - \frac{1}{2}\boldsymbol{Y}^T\boldsymbol{A}_2(\sum_s \theta_s \boldsymbol{A}_2^T \boldsymbol{B}_s \boldsymbol{A}_2)^{-1}\boldsymbol{A}_2^T\boldsymbol{Y}$$

$$= \cdots = C + \ell_{A_1}(\boldsymbol{\theta})$$

where C does not depend on $\boldsymbol{\theta}$.

Some facts about REML

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\mathbf{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^T \boldsymbol{Y} \sim \left(\mathbf{0}_{n-p}, \sum_{s=1}^{b} \theta_s \boldsymbol{A}^T \boldsymbol{B}_s \boldsymbol{A}\right)$$

- REML is the preferred method to estimate $\boldsymbol{\theta}$ and β in mixed effect models.
- From HW7, you now know how to perform REML with gradient descent.
    - More efficient maximization methods exist (Newton's method, BFGS, etc)
    - BFGS (my favorite) can be performed in `R` with `optim` and `constrOptim`
    - Can also estimate $\boldsymbol{\theta}$ with `lme4` (see vignette on blackboard) in `R`.
- In usual setting when $b = 1$, $\boldsymbol{B}_1 = I_n$ and $v_1 = \sigma^2$: REML is equivalent to ordinary least squares (HW9)

Inference on β using REML

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\boldsymbol{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

$$\hat{\beta} = \left(\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{Y}, \quad \boldsymbol{V}_{\hat{\theta}} = \sum_{s=1}^{b} \hat{\theta}_s \boldsymbol{B}_s$$

- Under regularity assumptions, $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}$ and $\left(\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X}\right)^{-1/2} \left(\hat{\beta} - \beta\right) \xrightarrow{\mathcal{D}} N\left(0, I_p\right)$
  - For large sample size n, suffices to assume $\hat{\beta} \sim N(\beta, (\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X})^{-1}) \Rightarrow$ Wald-type inference
- What about for small sample sizes: like OLS, need to account for uncertainty in $\hat{\theta}$
  - In OLS, accounted for uncertainty in $\hat{\theta}$ with a t-distribution.
  - There are many proposed solutions (we will look at 3), but this is a long-standing problem in the statistics community.

Inference on β using REML

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\boldsymbol{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

$$\hat{\beta} = \left(\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{Y}, \quad \boldsymbol{V}_{\hat{\theta}} = \sum_{s=1}^{b} \hat{\theta}_s \boldsymbol{B}_s$$

- Under regularity assumptions, $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}$ and $\left(\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X}\right)^{-1/2} \left(\hat{\beta} - \beta\right) \xrightarrow{\mathcal{D}} N\left(0, I_p\right)$
  - For large sample size n, suffices to assume $\hat{\beta} \sim N(\beta, (\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X})^{-1}) \Rightarrow$ Wald-type inference
- What about for small sample sizes: like OLS, need to account for uncertainty in $\hat{\boldsymbol{\theta}}$
  - In OLS, accounted for uncertainty in $\hat{\boldsymbol{\theta}}$ with a t-distribution.
  - There are many proposed solutions (we will look at 3), but this is a long-standing problem in the statistics community.

Inference on β using REML: Moment matching

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim \left(\mathbf{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$$

$$\hat{\beta} = \left(\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{Y}, \quad \boldsymbol{V}_{\hat{\theta}} = \sum_{s=1}^{b} \hat{\theta}_s \boldsymbol{B}_s$$

- Based on the ad-hoc moment matching procedure in M. Kenward & J. Roger (1997)
- Idea: $\hat{\text{Var}}\left(\hat{\beta}\right) = (\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X})^{-1}$ **underestimates** $\text{Var}\left(\hat{\beta}\right)$
- Use a Taylor expansion to inflate $\hat{\text{Var}}\left(\hat{\beta}\right)$ and better estimate $\text{Var}\left(\hat{\beta}\right)$.
- Match the moments of Wald statistics to determine appropriate degrees of freedom to approximate F- or t-distribution.
- At best an ad-hoc procedure, although it works reasonably well.
- Available through `nlme` in `R`.

Inference on β using REML: likelihood ratio test

Assume $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N\left(\boldsymbol{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$

- Suppose, for simplicity, we want to test $H_0 : \beta_j = 0$, $H_A : \beta_j \neq 0$
- Gives us two models:

  $H_A : \boldsymbol{Y} \sim N\left(\boldsymbol{X}_j\beta_j + \boldsymbol{X}_{(-j)}\boldsymbol{\beta}_{(-j)}, \boldsymbol{V}_\theta\right), \quad H_0 : \boldsymbol{Y} \sim N\left(\boldsymbol{X}_{(-j)}\boldsymbol{\beta}_{(-j)}, \boldsymbol{V}_\theta\right)$

  Want to make estimation of $\boldsymbol{\theta}$ in both models as invariant as possible to $\boldsymbol{\beta}$
- Let $\boldsymbol{Q}_{(-j)} \in \mathbb{R}^{n \times (n-p+1)}$ be a basis for ker $\left\{\boldsymbol{X}_{(-j)}^T\right\}$:

$$H_A : \boldsymbol{Q}_{(-j)}^T \boldsymbol{Y} \sim N\left(\boldsymbol{Q}_{(-j)}^T \boldsymbol{X}_j \beta_j, \boldsymbol{Q}_{(-j)}^T \boldsymbol{V}_\theta \boldsymbol{Q}_{(-j)}\right)$$

$$H_0 : \boldsymbol{Q}_{(-j)}^T \boldsymbol{Y} \sim N\left(\boldsymbol{0}_{n-p+1}, \boldsymbol{Q}_{(-j)}^T \boldsymbol{V}_\theta \boldsymbol{Q}_{(-j)}\right)$$

- Since H_0 is nested within H_A, can test H_0 with likelihood ratio statistic (LRT). Degrees of freedom?
- Can actually use **score test** here: only need to estimate parameters under H_0 (much faster).

Inference on β using REML: likelihood ratio test

Assume $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N\left(\boldsymbol{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$

- Suppose, for simplicity, we want to test $H_0 : \beta_j = 0$, $H_A : \beta_j \neq 0$
- Gives us two models:

  $$H_A : \boldsymbol{Y} \sim N\left(\boldsymbol{X}_j \beta_j + \boldsymbol{X}_{(-j)} \boldsymbol{\beta}_{(-j)}, \boldsymbol{V}_\theta\right), \quad H_0 : \boldsymbol{Y} \sim N\left(\boldsymbol{X}_{(-j)} \boldsymbol{\beta}_{(-j)}, \boldsymbol{V}_\theta\right)$$

  Want to make estimation of $\boldsymbol{\theta}$ in both models as invariant as possible to $\boldsymbol{\beta}$
- Let $\boldsymbol{Q}_{(-j)} \in \mathbb{R}^{n \times (n-p+1)}$ be a basis for ker $\left\{\boldsymbol{X}_{(-j)}^T\right\}$:

  $$H_A : \boldsymbol{Q}_{(-j)}^T \boldsymbol{Y} \sim N\left(\boldsymbol{Q}_{(-j)}^T \boldsymbol{X}_j \beta_j, \boldsymbol{Q}_{(-j)}^T \boldsymbol{V}_\theta \boldsymbol{Q}_{(-j)}\right)$$

  $$H_0 : \boldsymbol{Q}_{(-j)}^T \boldsymbol{Y} \sim N\left(\boldsymbol{0}_{n-p+1}, \boldsymbol{Q}_{(-j)}^T \boldsymbol{V}_\theta \boldsymbol{Q}_{(-j)}\right)$$

- Since H_0 is nested within H_A, can test H_0 with likelihood ratio statistic (LRT). Degrees of freedom?
- Can actually use **score test** here: only need to estimate parameters under H_0 (much faster).

Inference on $\boldsymbol{\beta}$ using REML: likelihood ratio test

Assume $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N\left(\boldsymbol{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$

- Suppose, for simplicity, we want to test $H_0 : \beta_j = 0$, $H_A : \beta_j \neq 0$
- Gives us two models:

  $H_A : \boldsymbol{Y} \sim N\left(\boldsymbol{X}_j\beta_j + \boldsymbol{X}_{(-j)}\boldsymbol{\beta}_{(-j)}, \boldsymbol{V}_\theta\right), \quad H_0 : \boldsymbol{Y} \sim N\left(\boldsymbol{X}_{(-j)}\boldsymbol{\beta}_{(-j)}, \boldsymbol{V}_\theta\right)$

  Want to make estimation of $\boldsymbol{\theta}$ in both models as invariant as possible to $\boldsymbol{\beta}$
- Let $\boldsymbol{Q}_{(-j)} \in \mathbb{R}^{n\times(n-p+1)}$ be a basis for ker $\left\{\boldsymbol{X}_{(-j)}^T\right\}$:

$$H_A : \boldsymbol{Q}_{(-j)}^T\boldsymbol{Y} \sim N\left(\boldsymbol{Q}_{(-j)}^T\boldsymbol{X}_j\beta_j, \boldsymbol{Q}_{(-j)}^T\boldsymbol{V}_\theta\boldsymbol{Q}_{(-j)}\right)$$
$$H_0 : \boldsymbol{Q}_{(-j)}^T\boldsymbol{Y} \sim N\left(\boldsymbol{0}_{n-p+1}, \boldsymbol{Q}_{(-j)}^T\boldsymbol{V}_\theta\boldsymbol{Q}_{(-j)}\right)$$

- Since H_0 is nested within H_A, can test H_0 with likelihood ratio statistic (LRT). Degrees of freedom?
- Can actually use **score test** here: only need to estimate parameters under H_0 (much faster).

Inference on β using REML: parametric bootstrap

Assume $\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim N\left(\boldsymbol{0}_n, \sum_{s=1}^{b} \theta_s \boldsymbol{B}_s\right)$

- Primarily for confidence intervals, although can also be used for testing
- Estimate $\boldsymbol{\theta}$ using REML, set $\hat{\beta} = \left(\boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{V}_{\hat{\theta}}^{-1} \boldsymbol{Y}$.
- Compute Wald-type t-statistics $t_1, \ldots, t_p$.
- Simulate new dataset:
  $$\boldsymbol{Y}^{(boot)} = \boldsymbol{X}\hat{\beta} + \epsilon^{(boot)}, \quad \epsilon^{(boot)} \sim N\left(\boldsymbol{0}_n, \sum_{s=1}^{b} \hat{\theta}_s \boldsymbol{B}_s\right)$$
- Re-compute $t_1^{(boot)}, \ldots, t_p^{(boot)}$
- Get $\alpha/2$ and $1 - \alpha/2$ quantiles of bootstrap $t_j^{(boot)}$'s.
- Use quantiles when determining $1 - \alpha$ CI.

Output of `nlme` and `lme4`

`nlme`:

- Outputs everything you would expect (coefficient estimates, variance estimates, etc.)
- Also outputs p-values using Kenward and Roger (1997) moment matching method
- p-values are ad-hoc and should be questioned in real data, although they are betting than the naive normal approximation
- Outputs AIC
  - AIC: developed with the goal of picking the model that will predict best.
  - (small) penalty for extra parameters
  - usually defined: -2 loglik + 2 $\times$ # parameters
  - here: number of parameters is 4 (2 fixed effects and 2 variance multipliers)
  - Usage?

Output of `nlme` and `lme4` (cont.)

`nlme`:

- Outputs BIC
  - developed with the goal of picking correct model
  - usually defined: -2 loglik + log(n) $\times$ # parameters
  - in `nlme::lme`, use log(n- # fixed effects) in place of log(n). Justification?
  - # parameters is stille # fixed effects + # variance multipliers
  - Usage?

BLUPs (best linear unbiased predictors)

- Consider the model $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\delta + \epsilon$, where $\delta \sim N\left(0, \sigma_\delta^2 I_r\right)$ and $\epsilon \sim N\left(0, \sigma^2 I_n\right)$
- $\mathrm{Var}(\boldsymbol{Y}) = \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T + \sigma^2 I_n$
- What is our best guess for δ?
- We will use the BLUP!
- Properties of BLUP: unbiased, has minimal variance.
- BLUP here: the conditional expectation $\hat{\delta} = E\left(\delta \mid \boldsymbol{Y}\right)$. This is a random variable. What is random here? Why is this the BLUP?
  - $\begin{pmatrix} bm\delta \\ \boldsymbol{Y} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \boldsymbol{X}\beta \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 I_r & \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T \\ \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T & \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T + \sigma^2 I_n \end{pmatrix}\right)$
  - $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\begin{pmatrix} \mu_1 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \Rightarrow$ $X_1 \mid X_2 = x_2 \sim N\left(\mu_{1|2}, \Sigma_{1|2}\right)$
  - $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(x_2 - \mu_2\right)$, $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{1|2}\Sigma_{22}^{-1}\Sigma_{1|2}^T$

BLUPs (best linear unbiased predictors)

- Consider the model $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\delta + \epsilon$, where $\delta \sim N\left(0, \sigma_\delta^2 I_r\right)$ and $\epsilon \sim N\left(0, \sigma^2 I_n\right)$
- $\text{Var}(\boldsymbol{Y}) = \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T + \sigma^2 I_n$
- What is our best guess for δ?
- We will use the BLUP!
- Properties of BLUP: unbiased, has minimal variance.
- BLUP here: the conditional expectation $\hat{\delta} = E\left(\delta \mid \boldsymbol{Y}\right)$. This is a random variable. What is random here? Why is this the BLUP?
- $\begin{pmatrix} bm\delta \\ \boldsymbol{Y} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \boldsymbol{X}\beta \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 I_r & \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T \\ \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T & \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T + \sigma^2 I_n \end{pmatrix}\right)$
- $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\begin{pmatrix} \mu_1 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \Rightarrow$
  $X_1 \mid X_2 = x_2 \sim N\left(\mu_{1|2}, \Sigma_{1|2}\right)$
- $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(x_2 - \mu_2\right), \Sigma_{1|2} = \Sigma_{11} - \Sigma_{1|2}\Sigma_{22}^{-1}\Sigma_{1|2}^T$

BLUPs (best linear unbiased predictors)

- Consider the model $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{\delta} + \epsilon$, where $\boldsymbol{\delta} \sim N\left(0, \sigma_\delta^2 I_r\right)$ and $\epsilon \sim N\left(0, \sigma^2 I_n\right)$
- $\text{Var}(\boldsymbol{Y}) = \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T + \sigma^2 I_n$
- What is our best guess for $\boldsymbol{\delta}$?
- We will use the BLUP!
- Properties of BLUP: unbiased, has minimal variance.
- BLUP here: the conditional expectation $\hat{\boldsymbol{\delta}} = E\left(\boldsymbol{\delta} \mid \boldsymbol{Y}\right)$. This is a random variable. What is random here? Why is this the BLUP?
  - $\begin{pmatrix} bm\delta \\ \boldsymbol{Y} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \boldsymbol{X}\beta \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 I_r & \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T \\ \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T & \sigma_\delta^2 \boldsymbol{Z}\boldsymbol{Z}^T + \sigma^2 I_n \end{pmatrix}\right)$
  - $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\begin{pmatrix} \mu_1 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \Rightarrow$
    $X_1 \mid X_2 = x_2 \sim N\left(\mu_{1|2}, \Sigma_{1|2}\right)$
  - $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(x_2 - \mu_2\right)$, $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{1|2}\Sigma_{22}^{-1}\Sigma_{1|2}^T$

Inference for variance components

- Consider the model $\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim N\left(\mathbf{0}_n, \sigma^2 I_n + \phi^2 \boldsymbol{B}\right)$
- $\boldsymbol{B}$ a partition matrix (i.e. groups samples by individuals); this is a typical model
- Will assume $\phi^2 \geq 0$
- How do we test $H_0 : \phi^2 = 0$?
- In some cases, we saw it was possible to use an F-test (i.e. homework)
- Suppose we want to estimate parameters with REML. How can we test H_0 with that output?

Inference for variance components

- Consider the model $\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon, \quad \epsilon \sim N\left(\boldsymbol{0}_n, \sigma^2 I_n + \phi^2 \boldsymbol{B}\right)$
- $\boldsymbol{B}$ a partition matrix (i.e. groups samples by individuals); this is a typical model
- Will assume $\phi^2 \geq 0$
- How do we test $H_0 : \phi^2 = 0$?
- In some cases, we saw it was possible to use an F-test (i.e. homework)
- Suppose we want to estimate parameters with REML. How can we test H_0 with that output?

Inference for variance components

- Under certain conditions (see for example Stram and Lee, 1994), LRT distribution is a 50:50 mixtures of χ^2_0 and χ^2_1
- The asymptotic results require that data can be partitioned into independent subsets and that these subsets increase with sample size.
- Asymptotic approximation is often poor (Crainiceanu and Ruppert, 2004).
- General case even more complicated!
- Can use parametric bootstrap to obtain distribution of statistics under H_0.