Estimation of Random Effects

- Best Linear Unbiased Predictors (BLUPs)

Recall (LMMs):

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

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

Log likelihood function of (β, θ):

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

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

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. Unbiasness: $E(\mathbf{C}^\mathbf{T}\mathbf{Y}) = \boldsymbol{\lambda}_1^T \beta$
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 $(\beta, \mathbf{b})$.

- $(\hat{\beta}, \hat{\mathbf{b}})$ satisfies

  Mixed Model Equations:

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

Recall

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

where

$$\begin{aligned} \mathbf{b} &\sim N(\mathbf{0}, \mathbf{D}(\theta)) \\ \epsilon &\sim N(\mathbf{0}, \mathbf{R}(\theta)) \\ \text{and } \mathbf{V} &= cov(\mathbf{Y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathbf{T}} + \mathbf{R} \end{aligned}$$

Recall the normal equations:

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

The BLUPs of $(\beta, \mathbf{b})$:

$$
\begin{aligned}
\hat{\beta} &= (\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{Y} \\
\hat{\mathbf{b}} &= \mathbf{D}\mathbf{Z}^{\mathbf{T}}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\widehat{\beta}) \\
&= \mathbf{D}\mathbf{Z}^{\mathbf{T}}\mathbf{P}\mathbf{Y}
\end{aligned}
$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}$.

P is called the projection matrix (projected into the error space).

Remarks

- The BLUPs $(\beta, \hat{\mathbf{b}})$ maximize the penalized log-likelihood

$$\ell(\beta, \boldsymbol{b}) = \ell(\mathbf{Y}|\mathbf{b}) - \frac{1}{2}\mathbf{b}^{\mathsf{T}}\mathbf{D}^{-1}\mathbf{b}$$

$$= -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Zb})^{\mathsf{T}}\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Zb}) - \frac{1}{2}\mathbf{b}^{\mathsf{T}}\mathbf{D}^{-1}\mathbf{b} + c$$

- $\hat{\mathbf{b}}$ is the posterior mean (mode).

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

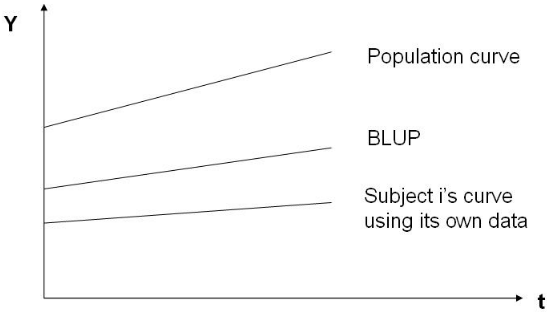
$$= |\mathbf{D}|^{-\frac{1}{2}} \int e^{\ell(\mathbf{Y}|\mathbf{b}) - \frac{1}{2}\mathbf{b}^{\mathsf{T}}\mathbf{D}^{-1}\mathbf{b}} d\mathbf{b}$$

$$\Rightarrow \hat{\mathbf{b}} = E(\mathbf{b}|\mathbf{Y}, \widehat{\beta}, \boldsymbol{\theta}) : \text{ Empirical Bayes estimator.}$$

- The BLUP $\hat{\mathbf{b}}$ is a shrinkage estimator:
  - $\hat{\mathbf{b}}$ is a weighted average of $\mathbf{0}$ (mean of $\mathbf{b}$) and the weighted LSE $\tilde{\mathbf{b}}$ when $\mathbf{b}$ is treated as fixed parameters.
  - If $\mathbf{b}$ is treated as fixed, then $\tilde{\boldsymbol{b}} = (\boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})$.
  - The BLUP:

$$\begin{aligned}
\hat{\mathbf{b}} &= \mathbf{D}\mathbf{Z}^{\mathbf{T}}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\widehat{\beta}) \\
&= (\mathbf{D}^{-1} + \mathbf{Z}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}\widehat{\beta}) \\
&= (\mathbf{D}^{-1} + \mathbf{Z}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{Z})^{-1}[\mathbf{Z}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{Z}\tilde{\mathbf{b}} + \mathbf{D}^{-1}\mathbf{0}]
\end{aligned}$$

- $\hat{\mathbf{b}}$ shrinks $\tilde{\mathbf{b}}$ towards $\mathbf{0}$.



- The estimated curve for subject i borrows strength (information) from the other subjects.

- Covariance of $(\hat{\beta}, \hat{\mathbf{b}})$:

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

$$\stackrel{\triangle}{=} \mathbf{H}^{-1}$$

$\Rightarrow$

$$cov(\hat{\mathbf{b}}-\mathbf{b}) = \mathbf{D} - \mathbf{D}\mathbf{Z}^{\mathbf{T}}\mathbf{P}\mathbf{Z}\mathbf{D}$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}$.

- Remarks:

$$\mathbf{H} = -\frac{\partial^2 \ell_p(\beta, \boldsymbol{b})}{\partial(\beta, \boldsymbol{b})\partial(\beta, \boldsymbol{b})^T}$$

- Note:
$$cov(\hat{\beta}-\beta) = cov(\hat{\beta})$$
but
$$cov(\hat{\mathbf{b}}-\mathbf{b}) \neq cov(\hat{\mathbf{b}})$$
Why?

- Example (longitudinal data):

$$Y_{ij} = \beta_0 + \mathbf{X}_{ij}^T\beta_1 + t_{ij}\beta_2 + b_{1i} + b_{2i}t_{ij} + \epsilon_{ij}$$

$$\begin{aligned}\hat{\beta}, \hat{b}_i \Rightarrow \hat{\mu}_i &= \hat{\beta}_0 + \mathbf{X}_{ij}^T\hat{\beta}_1 + t_{ij}\hat{\beta}_2 + \hat{b}_{1i} + \hat{b}_{2i}t_{ij} \\ &= (\hat{\beta}_0 + \hat{b}_{1i}) + (\hat{\beta}_2 + \hat{b}_{2i})t_{ij} + \mathbf{X}_{ij}^T\hat{\beta}_1\end{aligned}$$

$=$ Estimated subject specific curve.

REML log-likelihood of $\boldsymbol{\theta}$

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

Remarks:

- The MLE of $(\boldsymbol{\theta}, \beta)$ jointly maximizes

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

- The REML estimator of $\boldsymbol{\theta}$ maximizes $\ell_R(\boldsymbol{\theta})$ instead of $\ell(\boldsymbol{\beta}, \boldsymbol{\theta})$.
- The REML estimator of $\boldsymbol{\theta}$ accounts for the loss of degrees of freedom from estimating β, and has a smaller bias and a larger variance compared to its MLE counterpart.
- REML eliminates the nuisance parameter β by using an error contrast.