Recall: Mixed Model and Normal Equations

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

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:

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

BLUE/BLUP of $(\beta, \mathbf{b})$

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

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 (1)

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

$$\ell(\beta, \boldsymbol{b}) = \ell(\mathbf{Y}|\mathbf{b}) - \frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{D}^{-\mathbf{1}}\mathbf{b}$$
$$= -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b})^{\mathbf{T}}\mathbf{R}^{-\mathbf{1}}(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b}) - \frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{D}^{-\mathbf{1}}\mathbf{b} + c$$

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

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \int e^{\ell(\mathbf{Y}|\mathbf{b}) + \ell(\mathbf{b})} d\mathbf{b} \\ &= |\mathbf{D}|^{-\frac{1}{2}} \int e^{\ell(\mathbf{Y}|\mathbf{b}) - \frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{D}^{-\mathbf{1}}\mathbf{b}} d\mathbf{b} \end{aligned}$$

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

Note: Empirical Bayes estimator is also called Stein estimator. You will show this in your homework. **b** is not a parameter: the dimension should not change for parameters.

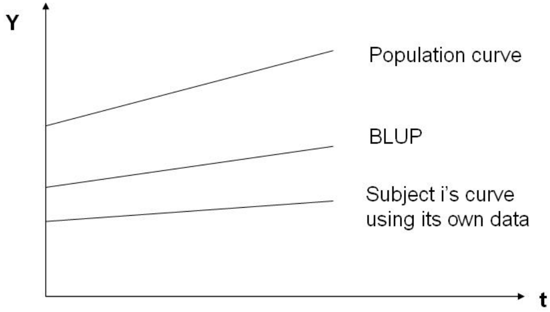
Remarks (2)

- 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:

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

Remarks (3)

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



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

Remarks (4)

- Note

$$cov(\hat{\beta} - \beta) = cov(\hat{\beta})$$

but

$$cov(\hat{\mathbf{b}} - \mathbf{b}) \neq cov(\hat{\mathbf{b}})$$

- Example (longitudinal data):

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

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

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

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

- 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.