Best Linear Unbiased Predictor (BLUP)

Haipeng Yu

February 20, 2018

Recall

Model

$$\mathbf{y} = \mathbf{g} + \mathbf{e}$$

- $\mathbf{y}$: observed phenotype
- $\mathbf{g}$: additive genetic values
- $\mathbf{e}$: residuals

$$\begin{bmatrix} \mathbf{g} \\ \mathbf{e} \end{bmatrix} \sim \mathcal{N}\left[\mathbf{0}, \begin{pmatrix} \mathbf{G}{\sigma_u}^2 & 0 \\ 0 & \mathbf{R}{\sigma_e}^2 \end{pmatrix}\right]$$

- $\mathbf{G}$ is genomic relationship matrix
- $\mathbf{R}$ is generally considered as $\mathbf{I}$ matrix

Best Predictor

- **Best** : Minimize Mean Squared Error (**MSE**): $\frac{1}{n}\sum_{i=1}^{n}(\hat{g}_i - g_i)^2$
- **Best predictor** (Henderson, 1976) : predictor that minimize prediction error variance (**PEV**), where

$$PEV = var(\hat{g} - g) = var(g - \hat{g})$$

- **Derivation** :
  - $g = f(y)$
  - $g_i(y_i) = E(g_i|y_i) + e$

Best Predictor

$$
\begin{aligned}
E[(g_i - \hat{g}_i)^2] &= E[(g_i - \widehat{g_i(y_i)})^2] \\
&= E[(g_i - E(g_i|y_i) + E(g_i|y_i) - \widehat{g_i(y_i)})^2] \\
&= E[(g_i - E(g_i|y_i))^2] + E[(E(g_i|y_i) - \widehat{g_i(y_i)})^2] \\
&+ 2E[(g_i - E(g_i|y_i)) \times (E(g_i|y_i) - \widehat{g_i(y_i)})]
\end{aligned}
$$

Decomposition theory : for any random variable g_i , where $g_i = E(g_i|y_i) + e_i$

- $E(e_i|y_i) = 0$
- $E(h(y_i) * e_i) = 0$, where $h(y_i)$ is a function of y_i.

Derivation-continue :
$E[(g_i - \hat{g}_i)^2]$ was minimized when : $E(g_i|y_i) = \widehat{g_i(y_i)}$ (**Conditional Expectation**)

Statistical Model

Model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

- $\mathbf{X}$: Incidence matrix for fixed effects
- $\mathbf{Z}$: Incidence matrix for additive genetic effects
- $\mathbf{b}$: Vector of fixed effects
- $\mathbf{u}$: Vector of additive genetic effects
- $\mathbf{e}$: Vector of residuals

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim \mathcal{N} \left[\mathbf{0}, \begin{pmatrix} \mathbf{G}{\sigma_u}^2 & 0 \\ 0 & \mathbf{R}{\sigma_e}^2 \end{pmatrix} \right]$$

Best Linear Unbiased Predictor

Henderson, 1976

$$\begin{aligned} BLUP(\hat{\mathbf{u}}) &= E(\mathbf{u}|\mathbf{y}) \\ &= E(\mathbf{u}) + Cov(\mathbf{u}, \mathbf{y}')Var(\mathbf{y})^{-1}(\mathbf{y} - E(\mathbf{y})) \\ &= \mathbf{G}\sigma^{\mathbf{2}}_{\mathbf{u}}\mathbf{Z}'\mathbf{V}^{-\mathbf{1}}_{\mathbf{y}}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ &= \mathbf{G}\sigma^{\mathbf{2}}_{\mathbf{u}}\mathbf{Z}'(\mathbf{Z}\mathbf{G}\sigma^{\mathbf{2}}_{\mathbf{u}}\mathbf{Z}' + \mathbf{I}\sigma^{\mathbf{2}}_{\mathbf{e}})^{-\mathbf{1}}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{aligned}$$

Two steps to predict estimated breeding value (**EBV**) of $\hat{\mathbf{u}}$

1. Fit ordinary least square (OLS) to estimate fixed effect $\hat{\mathbf{b}}$
2. Predict EBV ($\hat{\mathbf{u}}$) with BLUP conditioned on estimated fixed effect $\hat{\mathbf{b}}$

Mixed Model Equation (MME)

Henderson (1949; 1950; 1959; 1963; 1975)

Model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$
$$\mathbf{u} \sim \mathcal{N}(0, \mathbf{G}^*)$$
$$\mathbf{e} \sim \mathcal{N}(0, \mathbf{R}^*)$$

Joint distribution :

$$f(\mathbf{y}, \mathbf{u}) = g(\mathbf{y}|\mathbf{u})h(\mathbf{u})$$

Likelihood :

$$\begin{aligned} L(\mathbf{R}^*|\mathbf{y}) &= g(\mathbf{y}|\mathbf{u}) \\ &= 2\pi^{-\frac{1}{2}n}|\mathbf{R}^*|^{-\frac{1}{2}} exp\left\{-\frac{(\mathbf{y}-\mathbf{X}\mathbf{b}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{*-1}(\mathbf{y}-\mathbf{X}\mathbf{b}-\mathbf{Z}\mathbf{u})}{2}\right\} \\ L(\mathbf{G}^*|\mathbf{u}) &= h(\mathbf{u}) \\ &= 2\pi^{-\frac{1}{2}q}|\mathbf{G}^*|^{-\frac{1}{2}} exp\left\{-\frac{\mathbf{u}'\mathbf{G}^{*-1}\mathbf{u}}{2}\right\} \end{aligned}$$

MME

Maximize joint distribution $f(\mathbf{y}, \mathbf{u})$ to derive MME:

$$
\begin{aligned}
f(\mathbf{y}, \mathbf{u}) &= g(\mathbf{y}|\mathbf{u})h(\mathbf{u}) \\
&= 2\pi^{-\frac{1}{2}n-\frac{1}{2}q}|\mathbf{R}^*|^{-\frac{1}{2}}|\mathbf{G}^*|^{-\frac{1}{2}} \\
&\quad exp\left\{-\frac{1}{2}\begin{bmatrix}\mathbf{u}' \\ (\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'\end{bmatrix}\begin{bmatrix}\mathbf{G}^* & 0 \\ 0 & \mathbf{R}^*\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{u} \\ (\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})\end{bmatrix}\right\} \\
&= C.\frac{1}{exp\left\{\frac{1}{2}.\begin{bmatrix}\mathbf{u}' \\ (\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'\end{bmatrix}\begin{bmatrix}\mathbf{G}^* & 0 \\ 0 & \mathbf{R}^*\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{u} \\ (\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})\end{bmatrix}\right\}}
\end{aligned}
$$

To maximize joint distribution, we just need to minimize "denominator".

MME

Set

$$f(\mathbf{b},\mathbf{u}) = \begin{bmatrix} \mathbf{u}' \\ (\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})' \end{bmatrix} \begin{bmatrix} \mathbf{G}^* & 0 \\ 0 & \mathbf{R}^* \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{u} \\ (\mathbf{y}-\mathbf{Xb}-\mathbf{Zu}) \end{bmatrix}$$

$$= \mathbf{u}'\mathbf{G}^{*-1}\mathbf{u} + (\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'\mathbf{R}^{*-1}(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})$$

$$= \mathbf{u}'\mathbf{G}^{*-1}\mathbf{u} + (\mathbf{y}'-\mathbf{b}'\mathbf{X}'-\mathbf{u}'\mathbf{Z}')\mathbf{R}^{*-1}(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})$$

$$= \mathbf{u}'\mathbf{G}^{*-1}\mathbf{u} + \mathbf{y}'\mathbf{R}^{*-1}\mathbf{y} - \mathbf{y}'\mathbf{R}^{*-1}\mathbf{Xb} - \mathbf{y}'\mathbf{R}^{*-1}\mathbf{Zu} - \mathbf{b}'\mathbf{X}'\mathbf{R}^{*-1}\mathbf{y}$$

$$+ \mathbf{b}'\mathbf{X}'\mathbf{R}^{*-1}\mathbf{Xb} + \mathbf{b}'\mathbf{X}'\mathbf{R}^{*-1}\mathbf{Zu} - \mathbf{u}'\mathbf{Z}'\mathbf{R}^{*-1}\mathbf{y} + \mathbf{u}'\mathbf{Z}'\mathbf{R}^{*-1}\mathbf{Xb}$$

$$+ \mathbf{u}'\mathbf{Z}'\mathbf{R}^{*-1}\mathbf{Zu}$$

$$= \mathbf{u}'\mathbf{G}^{*-1}\mathbf{u} + \mathbf{y}'\mathbf{R}^{*-1}\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{R}^{*-1}\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{R}^{*-1}\mathbf{Xb}$$

$$+ 2\mathbf{b}'\mathbf{X}'\mathbf{R}^{*-1}\mathbf{Zu} - 2\mathbf{u}'\mathbf{Z}'\mathbf{R}^{*-1}\mathbf{y} + \mathbf{u}'\mathbf{Z}'\mathbf{R}^{*-1}\mathbf{Zu}$$

MME

Differentiation of $f(\mathbf{b},\mathbf{u})$

$$\frac{\partial f(\mathbf{b},\mathbf{u})}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{R}^{*-\mathbf{1}}\mathbf{y} + 2\mathbf{X}'\mathbf{R}^{*-\mathbf{1}}\mathbf{X}\mathbf{b} + 2\mathbf{X}'\mathbf{R}^{*-\mathbf{1}}\mathbf{Z}\mathbf{u} = 0 \tag{1}$$

$$\frac{\partial f(\mathbf{b},\mathbf{u})}{\partial \mathbf{u}} = 2\mathbf{G}^{*-\mathbf{1}}\mathbf{u} + 2\mathbf{b}'\mathbf{X}'\mathbf{R}^{*-\mathbf{1}}\mathbf{Z} + 2\mathbf{Z}'\mathbf{R}^{*-\mathbf{1}}\mathbf{Z}\mathbf{u} - 2\mathbf{Z}'\mathbf{R}^{*-\mathbf{1}}\mathbf{y} = 0 \tag{2}$$

Solve the two equations

$$(1) : \mathbf{X}'\mathbf{R}^{*-\mathbf{1}}\mathbf{X}\mathbf{b} + \mathbf{X}'\mathbf{R}^{*-\mathbf{1}}\mathbf{Z}\mathbf{u} = \mathbf{X}'\mathbf{R}^{*-\mathbf{1}}\mathbf{y}$$

$$(2) : \mathbf{Z}'\mathbf{R}^{*-\mathbf{1}}\mathbf{X}\mathbf{b} + \mathbf{Z}'\mathbf{R}^{*-\mathbf{1}}\mathbf{Z}\mathbf{u} = \mathbf{Z}'\mathbf{R}^{*-\mathbf{1}}\mathbf{y}$$

MME

Henderson Mixed Model Equation

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{*-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{*-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{*-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{*-1}\mathbf{Z} + \mathbf{G}^{*-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{*-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{*-1}\mathbf{y} \end{bmatrix}$$

Multiply $\mathbf{R}^*$ on both sides

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Solve MME to get $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$ SIMULTANEOUSLY!

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

where, $\lambda = \frac{\sigma_e^2}{\sigma_u^2}$