

Additional Aspects of BLUP

Peter von Rohr

2020-11-13

Aspects

Recap: BLUP Animal Model, means that we use a linear mixed effect model with the breeding values of all animals in the pedigree as additional random effect.

Solutions for fixed effects and for predicted breeding values are obtained by Mixed Model Equations (MME)

- ▶ Accuracy
 - ▶ Results from MME are estimates of fixed effects and predictions of breeding values
 - ▶ Need statement about quality of estimates and predictions
- ▶ Confidence Intervals
- ▶ Decomposition of Predicted Breeding values

Accuracy

- ▶ One property of BLUP was that variance of prediction error is minimal
- ▶ How can we measure the variance of the prediction error
- ▶ Fixed effects **By definition of a fixed effect, the true value of the fixed effect, does not have any variance. This means $\text{var}(\beta) = 0$**

$$\text{var}(\beta - \hat{\beta}) = \text{var}(\hat{\beta})$$

- ▶ Random effects

The true value of a random effect (e.g. the breeding value u) is expected to show a certain variation (for u , $\text{var}(u) = A * \sigma_u^2$). This means, the predicted breeding values, they are expected to have a variation that should as close as possible to the variance of the true breeding values.

$$\text{var}(u - \hat{u}) = \text{var}(u) - 2 * \text{cov}(u, \hat{u}) + \text{var}(\hat{u}) = \text{var}(u) - \text{var}(\hat{u}) = \text{PEV}(\hat{u})$$

because with BLUP: $\text{cov}(u, \hat{u}) = \text{var}(\hat{u})$

PEV: Prediction error variance

PEV

- ▶ PEV depends on inverse of coefficient matrix of MME

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$$

- ▶ For predicted breeding values \hat{u}

$$PEV(\hat{u}) = var(u) - var(\hat{u}) = C^{22}$$

For real-world datasets, PEV can not be computed exactly, but it has to be approximated.

Single Animal i

$$PEV(\hat{u}_i) = (C)_{ii}^{22}$$

where $(C)_{ii}^{22}$ is the i -th diagonal of C^{22}

- ▶ Accuracy measured by correlation

Reliability (Bestimmtheitsmass)
is the square of the accuracy

$$r_{u_i, \hat{u}_i} = \frac{\text{cov}(u_i, \hat{u}_i)}{\sqrt{\text{var}(u_i) * \text{var}(\hat{u}_i)}} = \sqrt{\frac{\text{var}(\hat{u}_i)}{\text{var}(u_i)}}$$

- ▶ Combining

$$PEV(\hat{u}_i) = (C)_{ii}^{22} = \text{var}(u_i) - \text{var}(\hat{u}_i) = \text{var}(u_i) - r_{u_i, \hat{u}_i}^2 \text{var}(u_i)$$

Accuracy B_i

Reliability

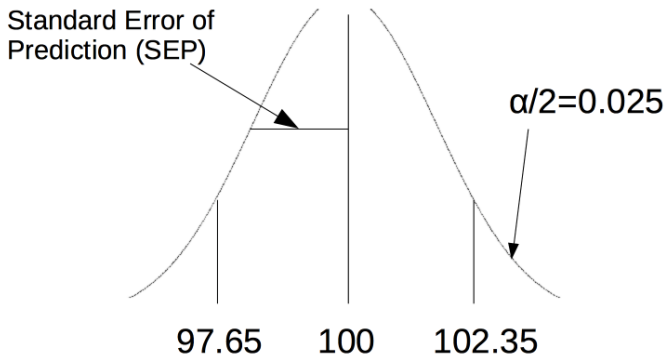
$$B_i = r_{u_i, \hat{u}_i}^2 = \frac{\text{var}(u_i) - (C)_{ii}^{22}}{\text{var}(u_i)} = 1 - \frac{PEV(\hat{u}_i)}{\text{var}(u_i)} = 1 - \frac{(C)_{ii}^{22}}{\text{var}(u_i)}$$

- ▶ B_i is large for small $PEV(\hat{u}_i)$
- ▶ In the limit $B_i \rightarrow 1$ for $PEV(\hat{u}_i) \rightarrow 0$
- ▶ For $PEV(\hat{u}_i) \rightarrow 0$ we must have $\text{var}(\hat{u}_i) \rightarrow \text{var}(u_i)$
- ▶ Therefore, the closer $\text{var}(\hat{u}_i)$ is to $\text{var}(u_i)$, the more accurate the predicted breeding value

Confidence Intervals of \hat{u}_i

- ▶ Predicted breeding value (\hat{u}_i) is a function of the data (y)
- ▶ Hence \hat{u}_i is a random variable with a distribution

Distribution



$$SEP(\hat{u}_i) = \sqrt{PEV(\hat{u}_i)} = \sqrt{(1 - r_{u_i, \hat{u}_i}^2) * var(u_i)}$$

Widths Of Confidence Intervals

Breeding values are predicted on different sources of information:

Calf first has a parental average: B is around 0.2 - 0.3

Table 1: Widths of Confidence Intervals for Given Accuracies

With own performance records, then B depends on the heritability B is 0.3-0.4

Accuracy	Interval Width
0.40	36.44
0.50	33.26
0.60	29.75
0.70	25.76
0.80	21.04
0.90	14.88
0.95	10.52
0.99	4.70

For animals with progeny, B starts to grow above 0.6

The advantage of genomic selection is that already for a young animal, B values of around 0.4 to 0.5 can be reached

with $\hat{u}_i = 100$, $var(u_i) = 144$ and $\alpha = 0.05$

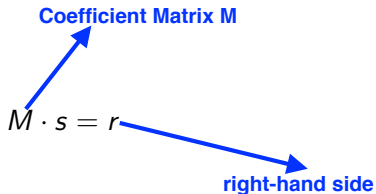
Decomposition of Predicted Breeding Value

- ▶ Write MME as

$$M \cdot s = r$$

Coefficient Matrix M

right-hand side



with

$$s = \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}$$

- ▶ $\hat{\beta}$ has length p
- ▶ \hat{u} has length q

Simplified Model

$$y_i = \mu + u_i + e_i$$

- where
- y_i Observation for animal i
 - u_i breeding value of animal i with a variance of $(1 + F_i)\sigma_u^2$
 - e_i random residual effect with variance σ_e^2
 - μ single fixed effect

Data

- ▶ all animals have an observation
- ▶ animal i has
 - ▶ parents s and d
 - ▶ n progeny k_j (with $j = 1, \dots, n$)
 - ▶ n mates l_j (with $j = 1, \dots, n$).
- ▶ progeny k_j has parents i and l_j .

Example

Animal	Sire	Dam	WWG
1	NA	NA	4.5
2	NA	NA	2.9
3	NA	NA	3.9
4	1	2	3.5
5	4	3	5.0

Variance components $\sigma_e^2 = 40$ and $\sigma_u^2 = 20$.

Model Components

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$X^T X = [5], X^T Z = [1 \ 1 \ 1 \ 1 \ 1]$$

$$Z^T Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Right-hand Side

$$X^T y = \left[\sum_{j=1}^n y_i \right] = 19.8$$

$$Z^T y = \begin{bmatrix} y1 \\ y2 \\ y3 \\ y4 \\ y5 \end{bmatrix} = \begin{bmatrix} 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5 \end{bmatrix}$$

A^{-1}

$$A^{-1} = \begin{bmatrix} 1.5 & 0.5 & 0 & -1 & 0 \\ 0.5 & 1.5 & 0 & -1 & 0 \\ 0 & 0 & 1.5 & 0.5 & -1 \\ -1 & -1 & 0.5 & 2.5 & -1 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}$$

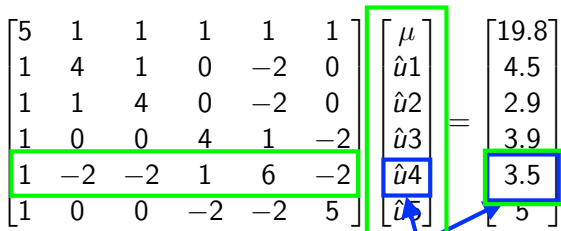
MME

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + A^{-1} * \lambda \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

The diagram illustrates the Mixed Model Equations (MME) with blue arrows pointing from labels to the corresponding parts of the equation:

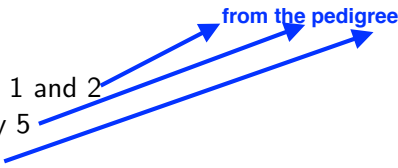
- M** points to the coefficient matrix $\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + A^{-1} * \lambda \end{bmatrix}$.
- s** points to the parameter vector $\begin{bmatrix} \hat{\mu} \\ \hat{u} \end{bmatrix}$.
- r** points to the right-hand side vector $\begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$.

Insert Data

$$\begin{bmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 0 & -2 & 0 \\ 1 & 1 & 4 & 0 & -2 & 0 \\ 1 & 0 & 0 & 4 & 1 & -2 \\ 1 & -2 & -2 & 1 & 6 & -2 \\ 1 & 0 & 0 & -2 & -2 & 5 \end{bmatrix} \begin{bmatrix} \mu \\ \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \end{bmatrix} = \begin{bmatrix} 19.8 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5 \end{bmatrix}$$


Look at the equation that contains y_4 and \hat{u}_4

Animal 4

- ▶ parents 1 and 2
 - ▶ progeny 5
 - ▶ mate 3
 - ▶ inspection of second but last equation in MME where y_4 and \hat{u}_4 occur
 - ▶ Remember from construction of A^{-1} , the variable d^{ii} can assume the following values
- 

$$d^{ii} = \begin{cases} 2 & \text{both parents known} \\ \frac{4}{3} & \text{one parent known} \\ 1 & \text{both parents unknown} \end{cases}$$

Extract Equation

$$y_4 = 3.5 = 1 * \hat{\mu} - 2 * \hat{u}_1 - 2 * \hat{u}_2 + 1 * \hat{u}_3 + 6 * \hat{u}_4 - 2 * \hat{u}_5$$

- ▶ Solving for \hat{u}_4

$$\hat{u}_4 = \frac{1}{6} [y_4 - \hat{\mu} + 2 * (\hat{u}_1 + \hat{u}_2) - \hat{u}_3 + 2\hat{u}_5]$$

- ▶ \hat{u}_4 depends on
 - ▶ own performance record y_4
 - ▶ estimate of fixed effect $\hat{\mu}$ - environment
 - ▶ predicted breeding value of parents 1 and 2, mate 3 and progeny 5

General Equation

$$\hat{u}_i = \frac{1}{1 + \alpha\delta^{(i)} + \frac{\alpha}{4} \sum_{j=1}^n \delta^{(k_j)}} [y_i - \hat{\mu} + \frac{\alpha}{2} \left\{ \delta^{(i)}(\hat{u}_s + \hat{u}_d) + \sum_{j=1}^n \delta^{(k_j)}(\hat{u}_{k_j} - \frac{1}{2}\hat{u}_{l_j}) \right\}]$$

where α ration between variance components σ_e^2/σ_u^2
 $\delta^{(j)}$ contribution for animal j to A^{-1}