

<b>1</b>	<b>Model statement</b>
<b>2</b>	<b>Likelihood</b>
<b>3</b>	<b>BLUE and BLUP</b>
<b>4</b>	<b>General terms</b>
<b>5</b>	<b>Coefficients</b>
<b>6</b>	<b>Kernel trick</b>
<b>7</b>	<b>Information</b>
<b>8</b>	<b>Variance components</b>

### 1) Model statement

A linear Gauss-Markov models is presented as:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon} \\ \mathbf{y} &\sim N(\mathbf{X}\mathbf{b}, \mathbf{V}) \\ \boldsymbol{\varepsilon} &\sim N(0, \mathbf{V}) \end{aligned}$$

And independent random terms other than the residuals:

$$\boldsymbol{\varepsilon} = \sum_{i=1}^{i=1} \mathbf{Z}_i \mathbf{u}_i + \mathbf{e}$$

And the variance-covariance matrix is defined by

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \mathbf{V} = \sum_{i=1}^{i=1} \mathbf{Z}_i \mathbf{K}_i \mathbf{Z}_i' \sigma_i^2 + \mathbf{I} \sigma_e^2$$

Reshape the model the model into:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{b} + \sum_{i=1}^{i=1} \mathbf{Z}_i \mathbf{u}_i + \mathbf{e} \\ \mathbf{u} &\sim N(0, \mathbf{Z}_i \mathbf{K}_i \mathbf{Z}_i' \sigma_i^2) \\ \mathbf{e} &\sim N(0, \mathbf{I} \sigma_e^2) \end{aligned}$$

### 2) Likelihood

Gaussian likelihood is defined as:

$$L = 2\pi^{-0.5n} |\mathbf{V}|^{-0.5} \exp(-0.5(\mathbf{y} - \mathbf{X}\mathbf{b})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\mathbf{b}))$$

Log likelihood:

$$\ln L = -\frac{1}{2} [\ln |\mathbf{V}| + \mathbf{y}' \mathbf{P} \mathbf{y}]$$

And restricted log likelihood (log REML) is defined as:

$$\ln L = -\frac{1}{2} [\ln |\mathbf{V}| + \ln |\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}| + \mathbf{y}' \mathbf{P} \mathbf{y}]$$

Where the projection matrix  $\mathbf{P}$  is defined by:

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}$$

Properties and equalities of projection matrices ( $\mathbf{S}$  and  $\mathbf{P}$ ):

$$\begin{aligned} \mathbf{S}\mathbf{X} &= \mathbf{P}\mathbf{X} = 0, & \mathbf{P}\mathbf{y} &= \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}), & \mathbf{P}\mathbf{P} &= \mathbf{P} \\ \mathbf{P}\mathbf{V}\mathbf{P} &= \mathbf{P}, & \mathbf{y}' \mathbf{P} \mathbf{y} &= \mathbf{y}' \mathbf{e} \sigma_e^{-2}, & \mathbf{P} &= \mathbf{V}^{-1} \mathbf{S} \\ \mathbf{S} &= \mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}', & \mathbf{S}\mathbf{y} &= \mathbf{y} - \mathbf{X}\mathbf{b}, & \mathbf{S}\mathbf{S} &= \mathbf{S} \\ \mathbf{Z}\mathbf{A}\mathbf{Z}\mathbf{P}\mathbf{y} &= \mathbf{Z}\mathbf{u} \sigma_u^{-2}, & \mathbf{P}\mathbf{y} &= (\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{u}) \sigma_e^{-2} \end{aligned}$$

The inverse of  $\mathbf{V}$  can be expressed as follows

$$\mathbf{V}^{-1} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}' \mathbf{Z} + \lambda \mathbf{K}^{-1})^{-1} \mathbf{Z}'$$

### 3) BLUE and BLUP

The solution for the best linear unbiased estimators ( $\mathbf{b}$ , BLUE) and best linear unbiased predictor ( $\mathbf{u}$ , BLUP) are commonly defined through the equations:

$$\begin{aligned} \mathbf{b} &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \\ \mathbf{u} &= \sigma_i^2 \mathbf{K}_i \mathbf{Z}_i' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\mathbf{b}) \end{aligned}$$

Conditional to the random terms, BLUE coefficients can be estimated from solving the system of equation:

$$(\mathbf{X}' \mathbf{X}) \mathbf{b} = \mathbf{X}' (\mathbf{y} - \sum \mathbf{Z}_i \mathbf{u}_i)$$

And coefficients of the  $i^{\text{th}}$  random effect, conditional to fixed effects and other random effects, as:

$$(\mathbf{Z}_i' \mathbf{Z}_i + \mathbf{K}_i^{-1} \lambda_i) \mathbf{u} = \mathbf{Z}_i' (\mathbf{y} - \mathbf{X}\mathbf{b} - \sum \mathbf{Z}_{-i} \mathbf{u}_{-i})$$

Where  $\lambda_i = \sigma_e^2 \sigma_i^{-2}$ .

### 4) General terms

In a more generalized notation, the mixed model equation is described as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_{i=1}^{i=1} \mathbf{Z}_i \mathbf{u}_i + \mathbf{e} = \mathbf{W}\mathbf{g} + \mathbf{e}$$

Where the design matrices and coefficients are unified as:

$$\begin{aligned} \mathbf{W} &= \mathbf{X} | \mathbf{Z}_1 | \mathbf{Z}_2 | \dots | \mathbf{Z}_I \\ \mathbf{g} &= \mathbf{b} | \mathbf{u}_1 | \mathbf{u}_2 | \dots | \mathbf{u}_I \end{aligned}$$

And the mixed model can be written as:

$$\begin{aligned} \mathbf{y} &= \mathbf{W}\mathbf{g} + \mathbf{e} \\ \mathbf{y} &\sim N(\mathbf{W}\mathbf{g}, \mathbf{I} \sigma_e^2) \end{aligned}$$

Where the coefficients are, in probabilistic terms, described as a normal distribution with a block diagonal variance:

$$\mathbf{g} \sim \text{MVN} \left( \begin{bmatrix} b \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \infty & 0 & \dots & 0 \\ 0 & \mathbf{K}_1 \sigma_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{K}_I \sigma_I^2 \end{bmatrix} \right)$$

Thus, the mixed model equation is generalized as:

$$\begin{aligned} \mathbf{W}' \mathbf{W} + \boldsymbol{\Sigma} &= \mathbf{W}' \mathbf{y} \\ \boldsymbol{\Sigma} &= \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & \mathbf{K}_1^{-1} \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{K}_I^{-1} \lambda_I \end{bmatrix} \end{aligned}$$

And the mixed model equation makes the complex problem a simple " $\mathbf{A}\mathbf{x} = \mathbf{b}$ " mathematic problem:

$$\begin{aligned} \mathbf{C} &= \mathbf{g}\mathbf{r} \\ \mathbf{C} &\rightarrow \mathbf{W}' \mathbf{W} + \boldsymbol{\Sigma} \\ \mathbf{r} &\rightarrow \mathbf{W}' \mathbf{y} \end{aligned}$$

## 5) Coefficients

The conditional expectation of each regression coefficient, assuming  $\mathbf{K}_i = \mathbf{I}$ , is expressed as:

$$g_j = \frac{\mathbf{w}_j'(\mathbf{y} - \mathbf{W}_{-j}\mathbf{g}_{-j})}{\mathbf{w}_j'\mathbf{w}_j + \lambda_j} = \frac{\mathbf{w}_j'\tilde{\mathbf{y}}}{\mathbf{w}_j'\mathbf{w}_j + \lambda_j}$$

Where  $w_j$  from fixed effect coefficients assume  $\lambda = 0$ .

Conditional to all other terms via *Gauss-Seidel residual update* (GSRU), coefficients can be efficiently estimated in two steps:

$$g_j^{t+1} = \frac{\mathbf{w}_j'\mathbf{e}^t + \mathbf{w}_j'\mathbf{w}_j g_j^t}{\mathbf{w}_j'\mathbf{w}_j + \lambda_j}$$

$$\mathbf{e}^{t+1} = \mathbf{e}^t - \mathbf{w}_j(g_j^{t+1} - g_j^t)$$

## 6) Kernel trick

It is also possible to expand this framework to scenarios where  $\mathbf{K}_i \neq \mathbf{I}$ , that is through the diagonalization of  $\mathbf{K}_i$  by spectral decomposition:

$$\mathbf{K}_i = \mathbf{U}_i \mathbf{D}_i \mathbf{U}_i'$$

Where  $\mathbf{U}$  is a matrix of Eigenvectors and  $\mathbf{D}$  is a diagonal matrix of Eigenvalues. With properties

$$\begin{aligned}\mathbf{K}^{-1} &= \mathbf{U}\mathbf{D}^{-1}\mathbf{U}' \\ \mathbf{K}'\mathbf{K} &= \mathbf{U}\mathbf{D}^2\mathbf{U}' \\ \mathbf{D} &= \mathbf{U}'\mathbf{K}\mathbf{U} \\ \mathbf{I} &= \mathbf{U}'\mathbf{U} \\ \mathbf{U}' &= \mathbf{U}^{-1} \\ \text{tr}(\mathbf{K}) &= \text{tr}(\mathbf{D})\end{aligned}$$

Such that:

$$\begin{aligned}\mathbf{u}_i &\sim N(0, \mathbf{Z}_i \mathbf{K}_i \mathbf{Z}_i' \sigma_i^2) \\ \mathbf{u}_i &\sim \mathbf{U} N(0, \mathbf{Z}_i \mathbf{D}_i \mathbf{Z}_i' \sigma_i^2) \\ \mathbf{a}_i &\sim N(0, \mathbf{D}_i \sigma_i^2) \\ E[\mathbf{u}_i] &= \mathbf{U}_i \mathbf{a}_i \\ \mathbf{Z}_i \mathbf{u}_i &= \mathbf{Z}_i \mathbf{U}_i \mathbf{a}_i\end{aligned}$$

Assume the notation:

$$\mathbf{V}_i = \mathbf{Z}_i \mathbf{U}_i$$

Then, for  $i^{\text{th}}$  random effect and  $j^{\text{th}}$  Eigen-pair, the regression coefficient has expectation (GSRU):

$$g_{ij} = \frac{\mathbf{v}'\tilde{\mathbf{y}}}{\mathbf{v}'\mathbf{v} + \lambda_j d_j^{-1}}, \quad \mathbf{v} = \mathbf{Z}_i \mathbf{U}_{ij}$$

## 7) Information matrix

$$\mathbf{V}_{(Z_u)} = \frac{\partial \mathbf{V}}{\partial \sigma_i^2} = \mathbf{Z}_i \mathbf{K}_i \mathbf{Z}_i' \quad \text{and} \quad \mathbf{V}_{(e)} = \frac{\partial \mathbf{V}}{\partial \sigma_e^2} = \mathbf{I}$$

$$\text{Information} = \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} = \mathbf{P} \mathbf{V}_i \mathbf{P} \mathbf{V}_j$$

## 8) Variance components

First derivative 1: Expectation-maximization

$$\sigma_e^2 = \frac{\mathbf{y}'\mathbf{P}\mathbf{y}}{\text{tr}(\mathbf{P})} = \frac{\mathbf{e}'\mathbf{e} + \text{tr}(\mathbf{W}\mathbf{C}^{-1}\mathbf{W}')\sigma_e^{2^{t-1}}}{n} = \frac{\mathbf{y}'\mathbf{e}}{n - r}$$

$$\sigma_i^2 = \frac{\mathbf{y}'\mathbf{P}\mathbf{V}^{-1}\mathbf{P}\mathbf{y}}{\text{tr}(\mathbf{P}\mathbf{Z}_i \mathbf{K}_i \mathbf{Z}_i')} = \frac{\mathbf{u}_i' \mathbf{K}_i^{-1} \mathbf{u}_i + \text{tr}(\mathbf{K}_i^{-1} \mathbf{C}^{ii})\sigma_e^2}{q_i} = \frac{\mathbf{u}_i' \mathbf{K}_i^{-1} \mathbf{u}_i}{q_i - \lambda \text{tr}(\mathbf{K}_i^{-1} \mathbf{C}^{ii})}$$

$n$  = number of observations  
 $r$  = rank of  $\mathbf{X}$  (number of columns of  $\mathbf{X}$ )  
 $q_i$  = number of columns of  $\mathbf{Z}_i$

First derivative 2: Gibbs sampling (flat priors:  $S = 0, v = -2$ )

$$\sigma_e^2 \sim \frac{\mathbf{e}'\mathbf{e} + S_e v_e}{\chi_{n+v_e}^2} \quad \text{and} \quad \sigma_i^2 \sim \frac{\mathbf{u}_i' \mathbf{K}_i^{-1} \mathbf{u}_i + S_i v_i}{\chi_{q_i+v_i}^2}$$

First derivative 3: Pseudo-expectation

$$\sigma_e^2 = \frac{\mathbf{y}'\mathbf{e}}{n - r} \quad \text{and} \quad \sigma_i^2 = \frac{(\mathbf{y} - \mathbf{X}\mathbf{b})'\mathbf{Z}_i \mathbf{u}_i}{\text{tr}(\mathbf{S}\mathbf{Z}_i \mathbf{K}_i \mathbf{Z}_i')}$$

Second derivative 1: MINQUE

$$\begin{bmatrix} \text{tr}(\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}') & \text{tr}(\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}) \\ \text{tr}(\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}) & \text{tr}(\mathbf{P}) \end{bmatrix} \begin{bmatrix} \sigma_u^2 \\ \sigma_e^2 \end{bmatrix} = \begin{bmatrix} \mathbf{y}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}\mathbf{y} \\ \mathbf{y}'\mathbf{P}\mathbf{y} \end{bmatrix}$$

Second derivative 2: Fisher scoring

$$\mathbf{v}^{t+1} = \mathbf{v}^t - \mathbf{I}^{-1}\mathbf{s}, \quad \mathbf{v} = \begin{bmatrix} \sigma_u^2 \\ \sigma_e^2 \end{bmatrix}$$

$$\mathbf{I} = \begin{bmatrix} \text{tr}(\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}') & \text{tr}(\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}) \\ \text{tr}(\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}) & \text{tr}(\mathbf{P}) \end{bmatrix}$$

$$\mathbf{s} = \begin{bmatrix} \text{tr}(\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}') - \mathbf{y}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}\mathbf{y} \\ \text{tr}(\mathbf{P}) - \mathbf{y}'\mathbf{P}\mathbf{y} \end{bmatrix}$$

Second derivative 3: Average information

$$\mathbf{v}^{t+1} = \mathbf{v}^t - \mathbf{A}\mathbf{I}^{-1}\mathbf{s}, \quad \mathbf{v} = \begin{bmatrix} \sigma_u^2 \\ \sigma_e^2 \end{bmatrix}$$

$$\mathbf{A}\mathbf{I} = 0.5 \times \begin{bmatrix} \text{tr}(\mathbf{y}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}\mathbf{y}) & \text{tr}(\mathbf{y}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}) \\ \text{tr}(\mathbf{y}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}) & \text{tr}(\mathbf{y}'\mathbf{P}\mathbf{y}) \end{bmatrix}$$

$$\mathbf{s} = \begin{bmatrix} \text{tr}(\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}') - \mathbf{y}'\mathbf{P}\mathbf{Z}\mathbf{K}\mathbf{Z}'\mathbf{P}\mathbf{y} \\ \text{tr}(\mathbf{P}) - \mathbf{y}'\mathbf{P}\mathbf{y} \end{bmatrix}$$

Derivation-free method: EMMA

$$\mathbf{Z}\mathbf{K}\mathbf{Z}' = \mathbf{U}\mathbf{D}\mathbf{U}'$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

$$\sigma_e^2 = \frac{(\mathbf{y} - \mathbf{X}\mathbf{b})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})}{n - r}$$

$$\text{argmax } L(\lambda) = -0.5 [\ln|\mathbf{V}| + \ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y}]$$

$$\ln|\mathbf{V}| = \sum \ln(d_j \lambda^{-1} + 1)$$

$$\ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| = \ln|\mathbf{X}'\mathbf{U}(\mathbf{D}^{-1}\lambda^{-1} + \mathbf{I})\mathbf{U}'\mathbf{X}|$$

$$\mathbf{y}'\mathbf{P}\mathbf{y} = \mathbf{y}'\mathbf{e}$$