

The objective of this laboratory session is to gain familiarity with the mixed linear models that we will be using in the Bayesian analyses later in the course.

Exercise 1

The lecture notes introduced the equations for generalized least squares (GLS). The GLS equation(s) for the model we discussed in the lecture are

$$\hat{\mathbf{b}}^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}), \text{ for } \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}.$$

These equations are useful as \mathbf{V} is typically full rank, but are not practical in many situations where \mathbf{V} is large. In this example with just the mean fitted as the only fixed effect, the GLS equation will be a scalar form.

In order to form \mathbf{V} , you will need to know \mathbf{G} and \mathbf{R} .

Create a small Hendersonian data set by constructing a vector \mathbf{y} of phenotypic observations (no more than 6 observations). Create a corresponding \mathbf{X} matrix to represent the incidence matrix for the fixed effects. This matrix will have as many rows as there are observations in \mathbf{y} , and as many columns as there are fixed effects in \mathbf{b} . Use the minimum configuration for \mathbf{X} which is a vector of 1's that would correspond to a model that included an overall mean. Other alternatives for \mathbf{X} might be to include a vector of covariates (eg age of the animal at measurement) or a class variable such as a fixed effect for the sex of the measured animal, or covariates and classes.

Construct a \mathbf{G} matrix that will be square and have order equal to the number of animals in the pedigree file. For ease of viewing, the order of \mathbf{G} should not exceed 6. The \mathbf{G} matrix is the variance-covariance matrix of the fitted random effects, such as the breeding values. In that case, \mathbf{G} will be the product of the numerator relationship or \mathbf{A} matrix, and the scale additive genetic variance. You could form \mathbf{A} for some simple pedigree and assume a value of the additive genetic variance, or create a small pedigree, and use Julia to form \mathbf{A}^{-1} directly and invert it to inspect \mathbf{A} . Note that the pedigree might contain some animals that do not have observed phenotypes, so the length of \mathbf{y} may be less than the order of \mathbf{G} .

Construct an incidence matrix \mathbf{Z} , that relates the observations in \mathbf{y} to the corresponding breeding value in \mathbf{u} . The matrix \mathbf{Z} may be an identity matrix if all animals in the pedigree have a phenotypic record. More typically, \mathbf{Z} has as many rows as there are records in \mathbf{y} , and as many columns as there are animals in \mathbf{u} (and therefore the \mathbf{G} matrix).

Construct \mathbf{R} , the variance-covariance matrix for the residual effects, which for independent and identically distributed residual effects will be an identity matrix of order equal to the length of \mathbf{y} , multiplied by the scalar residual variance. Recall that

the heritability is the ratio of the genetic variance over the phenotypic variance, and the phenotypic variance in this model is the sum of the additive genetic and residual variances, so the values you assume will imply a particular heritability.

Lastly, construct \mathbf{y} using `MvNormal` to produce the vectors \mathbf{u} and \mathbf{e} . Remember these vectors may be different lengths if some animals in the pedigree do not have observations.

Given defined values for all these vectors, matrices and constants, calculate the phenotypic variance-covariance matrix \mathbf{V} , and then solve the GLS equations to obtain best linear unbiased estimates (BLUEs) of the fixed effects. Use the BLUEs to adjust the phenotypic records and form deviations, that you can then use to compute the best linear unbiased predictions (BLUP) of the random effects as a linear function of these deviations, as described below. Note that this form of obtaining BLUP works with a singular \mathbf{G} matrix.

The equations to obtain BLUP estimates are

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}^0).$$

Be sure to save all your steps so you can immediately repeat your calculations with a modified dataset or different parameters. Print out and inspect the results of all your calculations.

Exercise 2

Repeat the same exercise as above, but this time estimate the BLUEs and predict the BLUPs by setting up and solving the mixed model equations. The answers should be identical to those you obtained using GLS. The mixed model equations are shown below.

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}^0 \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Exercise 3

Obtain the variance of the estimated BLUP effects, and the prediction error variance. These values require elements of the inverse of the mixed model coefficient matrix. We will use the following notation

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

and the corresponding partitions of the inverse are

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

In relation to random effects, we need only concern ourselves with the \mathbf{C}^{22} partition of the inverse coefficient matrix. Note however that the entire coefficient matrix must be inverted to obtain the partition of interest. From this partition you have the prediction error variance-covariance matrix. That is,

$$\text{var}[\mathbf{u} - \hat{\mathbf{u}}] = \mathbf{C}^{22}$$

$$\text{var}[\hat{\mathbf{u}}] = \mathbf{G} - \mathbf{C}^{22}, \text{ and recall that } \text{var}[\mathbf{u}] = \mathbf{G}.$$

A common unitfree measure of how well we have estimated the BLUP is the square of the correlation between the true and estimated effect. Since the true effects are not known, this cannot be calculated directly, but is a function of the \mathbf{G} and \mathbf{C}^{22}

matrices. Specifically, $r^2 = \frac{\text{var}[\hat{\mathbf{u}}]}{\text{var}[\mathbf{u}]} = \frac{\text{diag}[\mathbf{G} - \mathbf{C}^{22}]}{\text{diag}[\mathbf{G}]}$ for best linear predictions (BLP) and best linear unbiased predictions (BLUP).

Exercise 4

In many circumstances we are interested in linear combinations of random effects. For example, we might want to know the BLUP and the r^2 of a team of sires rather than an individual. Alternatively, we might be interested in the contrast or difference between one or more alternative sires or teams. To compute these, we need to construct a relevant vector of contrasts that we will denote as \mathbf{k} . For

example, to predict the superiority of sire 1 over sire 2, for $\mathbf{u}' = [u_1 \ u_2 \ u_3 \ u_4]$,

we would form $\mathbf{k}' = [1 \ -1 \ 0 \ 0]$. To compare a team of the first two sires to

the second two sires we would use $\mathbf{k}' = [0.5 \ 0.5 \ -0.5 \ -0.5]$. Both of these contrasts can be considered simultaneously by stacking them up the rows of \mathbf{k}'

together in a matrix, $\mathbf{K} = \begin{bmatrix} \mathbf{1} & -\mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0.5} & \mathbf{0.5} & -\mathbf{0.5} & -\mathbf{0.5} \end{bmatrix}$

The BLUP of $\mathbf{k}'\mathbf{u}$ is simply obtained as $\mathbf{k}'\hat{\mathbf{u}}$, and $\text{var}(\mathbf{k}'\mathbf{u}) = \mathbf{k}'\mathbf{G}\mathbf{k}$,

$$\text{var}(\mathbf{k}'\hat{\mathbf{u}}) = \mathbf{k}'[\mathbf{G} - \mathbf{C}^{22}]\mathbf{k}.$$

Construct some linear combinations, and estimate the prediction error variance and r^2 for these linear combinations.