

Pedigree-based mixed linear models

The Prediction Problem

Model Equation

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

Other aspects of the model

First moments $\mathbf{E}[\mathbf{u}] = \mathbf{0}, \mathbf{E}[\mathbf{e}] = \mathbf{0}$, therefore $\mathbf{E}[\mathbf{y}] = \mathbf{X}\mathbf{b}$

Second moments $\mathbf{var}[\mathbf{u}] = \mathbf{G}, \mathbf{var}[\mathbf{e}] = \mathbf{R}, \mathbf{cov}[\mathbf{u}, \mathbf{e}'] = \mathbf{0}$

Distributional Assumptions e.g. $\mathbf{u}, \mathbf{e} \sim \text{MVN}$

Want to predict \mathbf{u} or linear functions like $\mathbf{k}'\mathbf{u}$

Original Solution

Generalized Least Squares (GLS)

For estimable $\mathbf{q}'\mathbf{b}$, $\mathbf{q}'\hat{\mathbf{b}}^0$ is BLUE (Best Linear Unbiased Estimator)

where $\hat{\mathbf{b}}^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ for $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$

then $\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}^0)$, is BLUP (BLU Predictor)

(same as Selection Index/BLP except $(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}^0)$ in place of $(\mathbf{y} - \mathbf{X}\mathbf{b})$)

obtained by exploiting (genetic) covariances between animals

In traditional animal breeding practice

\mathbf{G} is large and dense and determined by \mathbf{A} the numerator relationship matrix

\mathbf{V} is too big to compute $\mathbf{X}'\mathbf{V}^{-1}$

BLP vs GLS BLUP

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$\mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\mathbf{u} + \mathbf{e}$, a fully random model

Selection Index Equations $\mathbf{P}\mathbf{b} = \mathbf{G}\mathbf{v}$

$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{v}$, defines the best linear function to predict \mathbf{u}
the "weights" are the same for every animal with the same
sources of information (ie same traits observed)

$$\text{BLP } \hat{\mathbf{u}} = \mathbf{b}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{v}\mathbf{G}\mathbf{P}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\text{cf } \text{GLS BLUP } \hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^0)$$

Henderson's Contributions One

Developed methods to compute **G** and **R** from field data
Henderson's Method I (not his!), II and III
Including circumstances that involved selection

Henderson's Contributions Two

Invented the Mixed Model Equations

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

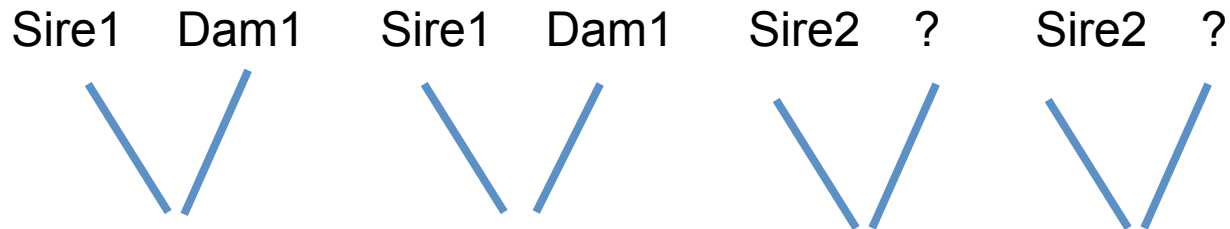
and jointly showed $\mathbf{k}'\hat{\mathbf{b}}^0$ and $\hat{\mathbf{u}}$ were BLUE and BLUP

Computationally tractable if \mathbf{G} and \mathbf{R} assumed diagonal or block-diagonal
(eg sire model with relationships ignored)

(Order 40 matrix takes weeks to invert by hand)

MME typically sparse in national animal evaluation

Example NRM or \mathbf{A} matrix

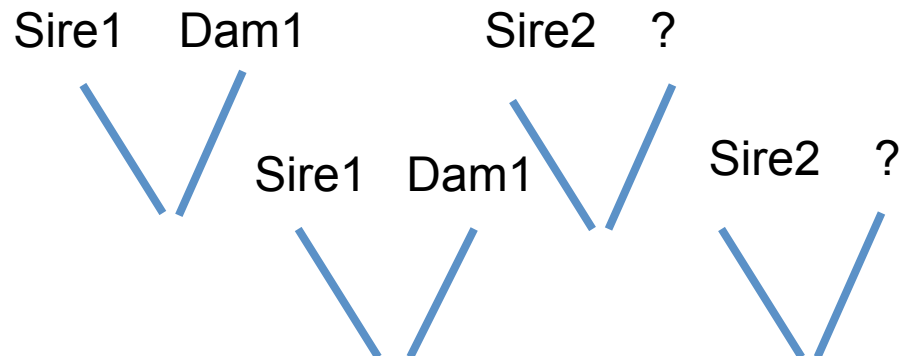


	Offspring ₁	Offspring ₂	Offspring ₃	Offspring ₄
Offspring ₁	1	1/2	0	0
Offspring ₂	1/2	1	0	0
Offspring ₃	0	0	1	1/4
Offspring ₄	0	0	1/4	1

Sires and dams unrelated and non-inbred

Simple calculation of \mathbf{A}_p^{-1} requires including all ancestors and would result in a matrix of order 7 not 4

A^{-1} matrix (animal model)



	Sire ₁	Dam ₁	Sire ₂	Off ₁	Off ₂	Off ₃	Off ₄
Sire ₁	2	1	0	-1	-1	0	0
Dam ₁	1	2	0	-1	-1	0	0
Sire ₂	0	0	1.667	0	0	-0.667	-0.667
Off ₁	-1	-1	0	2	0	0	0
Off ₂	-1	-1	0	0	2	0	0
Off ₃	0	0	-0.667	0	0	1.333	0
Off ₄	0	0	-0.667	0	0	0	1.333

Ancestors w/out records are fitted for simple A^{-1} structure

Henderson's Contributions Three

Invented an algorithm to directly form \mathbf{A}^{-1} from a pedigree list

Then \mathbf{G}^{-1} can be formed as a scalar product or kronecker product

define d to be "mendelian" sampling variance

$d = (1, 3/4, 1/2)$ for 0, 1 or 2 parents known

define $\mathbf{s}' = (-1/2, -1/2, 1)$ to represent sire (if known), dam (if known)

and individual equations

accumulate $\mathbf{s}d^{-1}\mathbf{s}'$ in the sire, dam and individual rows/columns

for every trio of animals in the pedigree list

Consequence of A^{-1} structure

Accumulate for each animal

$$\begin{array}{c}
 \textit{sire} \\
 \textit{dam} \\
 \textit{i}
 \end{array}
 \begin{bmatrix}
 0.25 & 0.25 & -0.5 \\
 0.25 & 0.25 & -0.5 \\
 -0.5 & -0.5 & 1
 \end{bmatrix}
 d^{-1}$$

When both parents are known

Nonparents (ie terminal offspring)

Own equation (ie row) has 2 on diagonal, -1 in sire column -1 in dam column

Parent with one offspring

Own equation has $2+1/2$ on diagonal, -1 in sire and dam columns

in addition to $-1/2$ in the column of its mate, -1 in column of offspring

Parent with many offspring to different mates

accumulates a large diagonal element, many small negative offdiagonals

Consider rearranging the MME

In general,

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

or equivalently $[\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}][\hat{\mathbf{u}}] = [\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}^0)]$

Single trait animal model $\mathbf{R} = \mathbf{I}\sigma_e^2$, $\mathbf{G} = \mathbf{A}\sigma_g^2$, $\mathbf{G}^{-1} = \mathbf{A}^{-1}\sigma_g^{-2}$

or multiplying σ_e^2 , $[\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1}][\hat{\mathbf{u}}] = [\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}^0)]$, with $\lambda = \frac{\sigma_e^2}{\sigma_g^2}$

Consider the MME for a nonparent

$$[\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1}][\hat{\mathbf{u}}] = [\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}^0)]$$

Nonparent animal with one record

$$(1 + 2\lambda)\hat{u}_{animal} - \lambda\hat{u}_{sire} - \lambda\hat{u}_{dam} = adjusted_y$$

$$\hat{u}_{animal} = \frac{2\lambda(\hat{u}_{sire} + \hat{u}_{dam})}{(1 + 2\lambda)2} + \frac{(adjusted_y)}{(1 + 2\lambda)}$$

$$= (1 - w)PA + w(adjusted_y) \text{ for } w = \frac{1}{(1 + 2\lambda)}$$

Consider the MME for a nonparent

$$\hat{u}_{animal} = (1 - w)PA + w(\text{adjusted } y) \text{ for } w = \frac{1}{(1 + 2\lambda)}$$

$$\lambda = \frac{1 - h^2}{h^2} \text{ so for } h^2 = 1, \lambda = 0, w = 1, (\text{no shrinkage})$$

for $h^2 = \text{low}$, $\lambda = \text{big}$, $w = \text{small}$, (shrink the deviation)

Two sources of BV information are pooled

The parent average PA

The individual prediction (shrunk deviation)

with heritability influencing shrinkage

Consider the MME for a nonparent

$$[\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1}][\hat{\mathbf{u}}] = [\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}^0)]$$

Nonparent animal with one record

$$\hat{u}_{animal} = (1 - w)PA + w(\text{adjusted } y)$$

Nonparent animal with no record

$$2\lambda\hat{u}_{animal} - \lambda\hat{u}_{sire} - \lambda\hat{u}_{dam} = 0$$

$$\hat{u}_{animal} = \frac{\lambda(\hat{u}_{sire} + \hat{u}_{dam})}{\lambda 2} = \frac{(\hat{u}_{sire} + \hat{u}_{dam})}{2} = PA$$

Reliability of nonparents

Property of BLP/BLUP is $\text{cov}(u, \hat{u}) = \text{var}(\hat{u})$ so $r^2 = \frac{\text{var}(\hat{u})}{\text{var}(u)}$

but $\hat{u}_{\text{nonparent}} = \frac{\hat{u}_{\text{sire}}}{2} + \frac{\hat{u}_{\text{dam}}}{2}$, for nonparent without a record

so $r_{\text{nonparent}}^2 = \frac{r_{\text{sire}}^2}{4} + \frac{r_{\text{dam}}^2}{4} \leq \frac{1}{2}$

Finally $\Delta G = \frac{i r_{\text{nonparent}} \sigma_g}{L}$, limiting selection response

when candidates at puberty lack phenotypic information

An option to do better

Solution

- We need a different representation of the covariance between relatives, that allows relatives other than parents to directly contribute to the prediction of nonparents without records
- The NRM or **A**-matrix is an expectation of relationships in the context of repeated sampling of the pedigree (conditional on pedigree)

A-matrix

- Relationship with self is $1+F$ (noninbred $F=0$)
- (Additive) relationship of $\frac{1}{2}$ between non-inbred full-sibs and between parents and non-inbred offspring
- Relationship of $\frac{1}{4}$ between non-inbred half-sibs and between grandparents and offspring
- But particular individuals can have greater or lesser values
 - If we know their genotype we can compute relationships conditional on the chromosome regions they inherited

Relationship matrix

A matrix

$$\begin{bmatrix} 1 & 0 & .5 & .5 & .5 & .5 \\ 0 & 1 & .5 & .5 & .5 & .5 \\ .5 & .5 & 1 & .5 & .5 & .5 \\ .5 & .5 & .5 & 1 & .5 & .5 \\ .5 & .5 & .5 & .5 & 1 & .5 \\ .5 & .5 & .5 & .5 & .5 & 1 \end{bmatrix}$$

A-inverse matrix

$$\begin{bmatrix} 3 & 2 & -1 & -1 & -1 & -1 \\ 2 & 3 & -1 & -1 & -1 & -1 \\ -1 & -1 & 2 & 0 & 0 & 0 \\ -1 & -1 & 0 & 2 & 0 & 0 \\ -1 & -1 & 0 & 0 & 2 & 0 \\ -1 & -1 & 0 & 0 & 0 & 2 \end{bmatrix}$$



Consider a sire, dam and 4 full sibs

Relationship matrix

A matrix

$$\begin{bmatrix} 1 & 0 & .5 & .5 & .5 & .5 \\ 0 & 1 & .5 & .5 & .5 & .5 \\ .5 & .5 & 1 & .5 & .5 & .5 \\ .5 & .5 & .5 & 1 & .5 & .5 \\ .5 & .5 & .5 & .5 & 1 & .5 \\ .5 & .5 & .5 & .5 & .5 & 1 \end{bmatrix}$$

G matrix

$$\begin{bmatrix} 1 & 0 & .5 & .5 & .5 & .5 \\ 0 & 1 & .5 & .5 & .5 & .5 \\ .5 & .5 & 1 & .6 & .4 & .4 \\ .5 & .5 & .6 & 1 & .4 & .4 \\ .5 & .5 & .4 & .4 & 1 & .6 \\ .5 & .5 & .4 & .4 & .6 & 1 \end{bmatrix}$$

A-inverse matrix

$$\begin{bmatrix} 3 & 2 & -1 & -1 & -1 & -1 \\ 2 & 3 & -1 & -1 & -1 & -1 \\ -1 & -1 & 2 & 0 & 0 & 0 \\ -1 & -1 & 0 & 2 & 0 & 0 \\ -1 & -1 & 0 & 0 & 2 & 0 \\ -1 & -1 & 0 & 0 & 0 & 2 \end{bmatrix}$$

G-inverse matrix

$$\begin{bmatrix} 3.5 & 2.5 & -1.25 & -1.25 & -1.25 & -1.25 \\ 2.5 & 3.5 & -1.25 & -1.25 & -1.25 & -1.25 \\ -1.25 & -1.25 & 2.1875 & -0.3125 & 0.3125 & 0.3125 \\ -1.25 & -1.25 & -0.3125 & 2.1875 & 0.3125 & 0.3125 \\ -1.25 & -1.25 & 0.3125 & 0.3125 & 2.1875 & -0.3125 \\ -1.25 & -1.25 & 0.3125 & 0.3125 & -0.3125 & 2.1875 \end{bmatrix}$$

Predict the last animal with no data

$$\left[-1.25\hat{u}_{sire} \quad -1.25\hat{u}_{dam} \quad .3125\hat{u}_{sib1} \quad .3125\hat{u}_{sib2} \quad -.3125\hat{u}_{sib3} \quad 2.1875\hat{u}_{candidate} \right] = [0]$$

$$\hat{u}_{candidate} = \frac{1.25(\hat{u}_{sire} + \hat{u}_{dam}) - 0.3125(\hat{u}_{sib1} + \hat{u}_{sib2}) + 0.3125\hat{u}_{sib3}}{2.1875}$$

But to form **G**, we needed to know which loci/QTL contribute to variation in performance

Some MME Results

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

$$\text{var}(\mathbf{g}) = \mathbf{G} \quad \text{var}(\hat{\mathbf{g}}) = \mathbf{G} - \mathbf{C}^{22} \quad \text{var}(\hat{\mathbf{g}} - \mathbf{g}) = \mathbf{C}^{22} \quad r_{g\hat{g}}^2 = \frac{\text{var}(\hat{g})}{\text{var}(g)}$$

$$\text{var}(\mathbf{k}'\mathbf{g}) = \mathbf{k}'\mathbf{G}\mathbf{k} \quad \text{var}(\mathbf{k}'\hat{\mathbf{g}}) = \mathbf{k}'(\mathbf{G} - \mathbf{C}^{22})\mathbf{k}$$