Lecture 16:
Models with multiple random effects: Repeated measures and maternal effects

Bruce Walsh lecture notes
Synbreed course
version 5 July 2013

Often there are several vectors of random effects

- Repeatability models
  - Multiple measures
- Common family effects
  - Cleaning up residual covariance structure
- Maternal effects models
  - Maternal effect has a genetic (i.e., breeding value) component

Multiple random effects

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

$\mathbf{y}$ is a n x 1 vector of observations

β is a q x 1 vector of fixed effects

$\mathbf{a}$ is a p x 1 vector of random effects

$\mathbf{u}$ is a m x 1 vector of random effects

$\mathbf{X}$ is n x q, $\mathbf{Z}$ is n x p, $\mathbf{W}$ is n x m

$\mathbf{y}$, $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}$ observed. β, $\mathbf{a}$, $\mathbf{u}$, $\mathbf{e}$ to be estimated

Covariance structure

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

Defining the covariance structure key in any mixed-model

Suppose $\mathbf{e} \sim (0, \sigma_e^2 \mathbf{I})$, $\mathbf{u} \sim (0, \sigma_u^2 \mathbf{I})$, $\mathbf{a} \sim (0, \sigma_A^2 \mathbf{A})$, as with breeding values

These covariances matrices are not sufficient to fully describe the model, as we have yet to give describe the relationship between **e**, **a**, and **u**. If they are independent:

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{u} \\ \mathbf{e} \end{pmatrix} \sim \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_A^2 \cdot \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \cdot \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_e^2 \cdot \mathbf{I} \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{u} + \mathbf{e} \qquad \begin{pmatrix} \mathbf{a} \\ \mathbf{u} \\ \mathbf{e} \end{pmatrix} \sim \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_A^2 \cdot \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \cdot \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_e^2 \cdot \mathbf{I} \end{pmatrix}$$

Covariance matrix for the vector of observations **y**

$$\mathbf{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{A}\mathbf{Z}^T\sigma_A^2 + \mathbf{W}\mathbf{W}^T\sigma_u^2 + \mathbf{I}\,\sigma_e^2$$

Note that if we ignored the second vector u of random effects, and assumed **y = Xβ + Za + e***, then **e* = Wu + e**, with Var(**e***) = σ_e^2 **I** + σ_u^2 **WW**^{T}

Consequence of ignoring random effects is that these are incorporated into the residuals, potentially compromising its covariance structure

Mixed-model Equations

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} & \mathbf{X}^T\mathbf{W} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z}+\lambda_A\mathbf{A}^{-1} & \mathbf{Z}^T\mathbf{W} \\ \mathbf{W}^T\mathbf{X} & \mathbf{W}^T\mathbf{Z} & \mathbf{W}^T\mathbf{W}+\lambda_u\mathbf{I} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{a}} \\ \widehat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \\ \mathbf{W}^T\mathbf{y} \end{pmatrix}$$

where

$$\lambda_A = \frac{\sigma_e^2}{\sigma_A^2} \quad \text{and} \quad \lambda_u = \frac{\sigma_e^2}{\sigma_u^2}$$

The repeatability model

- Often, multiple measurements (aka "records") are collected on the same individual
- Such a record for individual k has three components
  - Breeding value a_k
  - Common (permanent) environmental value p_k
  - Residual value for ith observation e_{ki}
- Resulting observation is thus
  - $z_{ki} = \mu + a_k + p_k + e_{ki}$
- The repeatability of a trait is $r = (\sigma_A^2+\sigma_p^2)/\sigma_z^2$
- Resulting variance of the residuals is $\sigma_e^2 = (1-r)\,\sigma_z^2$

Resulting mixed model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{Z}\mathbf{p} + \mathbf{e}$$

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} \sim \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_A^2 \cdot \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_p^2 \cdot \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_e^2 \cdot \mathbf{I} \end{pmatrix}$$

Notice that we could also write this model as

$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}(\mathbf{a} + \mathbf{p}) + \mathbf{e} = \mathbf{X}\beta + \mathbf{Z}\mathbf{v} + \mathbf{e}$, $\mathbf{v} = \mathbf{a}+\mathbf{p}$

In class question: Why can we obtain separate estimates of **a** and **p**?

The careful reader might notice that the two vectors of random effects, the breeding values $\mathbf{a}$ and permanent environment effects $\mathbf{p}$, enter the model as $\mathbf{Za}$ and $\mathbf{Zp}$, respectively. Why then do we simply not combine these, e.g., $\mathbf{Zu}$ where $\mathbf{u} = \mathbf{a} + \mathbf{p}$? The reason we cannot do this (and indeed the reason we can estimate $\mathbf{a}$ and $\mathbf{p}$ separately!) is that $\mathbf{a}$ and $\mathbf{p}$ have *different covariance structures*, $\sigma^2_A \mathbf{A}$ versus $\sigma^2_p \mathbf{I}$. Thus, we assume that permanent environment effects are uncorrelated across individuals and are homoscedastic. On the other hand, breeding values generate covariances in relatives. Again, the critical importance of the covariance matrix to a mixed model analysis is apparent.

The incident matrix Z

Suppose we have a total of 7 observations/records, with 3 measures from individual 1, 2 from individual 2, and 2 from individual 3. Then:

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}, \quad \mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

Why? Matrix multiplication. Consider y_{21}.

$$y_{21} = \mu + A_2 + p_2 + e_{21}$$

Consequences of ignoring p

- Suppose we ignored the permanent environment effects and assumed the model $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{e}^*$
  - Then $\mathbf{e}^* = \mathbf{Z}\mathbf{p} + \mathbf{e}$,
  - $\mathrm{Var}(\mathbf{e}^*) = \sigma_e^2 \mathbf{I} + \sigma_p^2 \mathbf{Z}\mathbf{Z}^T$
- Assuming that $\mathrm{Var}(\mathbf{e}^*) = \sigma_e^2 \mathbf{I}$ gives an incorrect model
- We could either
  - use $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{e}^*$ with the correct error structure (covariance) for $e^* = \sigma_e^2 \mathbf{I} + \sigma_p^2 \mathbf{Z}\mathbf{Z}^T$
  - Or use $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{Z}\mathbf{p} + \mathbf{e}$, where $e = \sigma_e^2 \mathbf{I}$

The repeatability model was used by Estany et al. (1989) to examined the selection response for litter size in rabbits. Their model assumed two groups of fixed effects, d_t the year-season (environmental) effect which had 22 levels in this experiment and the reproductive state l_i of the doe (l has three levels: l_1 for primiparious does, l_2 for lactating does, and l_2 for non-primiparious and non-lactating does). Since only two of these l_x factors are estimable, l_1 was assigned a value zero. Their model had three random effects, a_k and p_k for the additive genetic and permanent environmental effect of the kth doe, and the residual e, giving the overall model as

$$y_{tk\ell i} = \mu + l_i + d_t + a_k + p_k + e_{tk\ell i}$$

where $y_{tk\ell i}$ denotes the litter size for the ℓth litter of doe k in reproductive state i in season-year t.

In matrix form, the mixed-model becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Za} + \mathbf{Zp} + \mathbf{e}$$

where $\mathbf{a}$ and $\mathbf{p}$ are $n \times 1$ vectors corresponding to the n does, $\mathbf{Var}(\mathbf{a}) = \sigma_A^2\,\mathbf{A}$, $\mathbf{Var}(\mathbf{p}) = \sigma_p^2\,\mathbf{I}$, and $\mathbf{Var}(\mathbf{e}) = \sigma_e^2\,\mathbf{I}$. $\mathbf{X}$ and $\mathbf{Z}$ are incident matrices, and the vector of fixed effects is

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ l_1 \\ l_2 \\ d_1 \\ \vdots \\ d_{22} \end{pmatrix}$$

Resulting mixed-model equations

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z}+\lambda_A\mathbf{A}^{-1} & \mathbf{Z}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} & \mathbf{Z}^T\mathbf{Z}+\lambda_u\mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix}$$

where

$$\lambda_A = \frac{\sigma_e^2}{\sigma_A^2} = \frac{1-r}{h^2} \quad \text{and} \quad \lambda_u = \frac{\sigma_e^2}{\sigma_p^2} = \frac{1-r}{r-h^2}$$

Common family effects

- Sibs in the same family also share a common environment
  - Cov(full sibs) = $\sigma_A^2/2 + \sigma_D^2/4 + \sigma_{ce}^2$
- Hence, if the model assumes $y_i = \mu + a_i + e_i$, with $\mathbf{a} \sim 0, \sigma_A^2 \mathbf{A}$, $c \sim 0, \sigma_{cf}^2 I$. If there are records for different sibs from the same family, Var(**e**) is no longer of the form $\sigma_e^2 \mathbf{I}$
- $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{c} + \mathbf{e}$
- Again, if common family effect ignored (we assume $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{e}^*$) the error structure is $e^* = \sigma_e^2 I + \sigma_{cf}^2 \mathbf{W}\mathbf{W}^T$
  - Where $\sigma_{cf}^2 = \sigma_D^2/4 + \sigma_{ce}^2$
  - The common family effect may contain both environment and non-additive genetic components

Example: Measure 7 individuals, first five are from family one, last two from family 2

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{c} + \mathbf{e}$$

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix}, \quad \mathbf{Z} = \mathbf{I}, \quad \mathbf{a} = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Z = I as every individual has a single record.

If there are missing and/or repeated records,

Z does not have this simple structure

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{c} + \mathbf{e}$$

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix}, \quad \mathbf{Z} = \mathbf{I}, \quad \mathbf{a} = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Again, matrix multiplication gives us the form of the **Z** and **W** matrices. Consider y_6:

$$y_6 = \mu + A_6 + c_2 + e_6$$

Maternal effects with genetic components

- The phenotype of an offspring can be influenced by its mother beyond her genetic contribution
- For example, two offspring with identical genotypes will still show potentially significant differences in size if they receive different amounts of milk from their mothers
- Such maternal effects can be quite important
- While we have just discussed models with common family effects, these are potentially rather different that maternal effects models
  - Common family environmental effects are assumed not to be inherited across generations.

- Consider milk yield. The heritability for this trait is around 30% and the milk yield of the mother has a significant impact on the weight of her offspring
- Offspring with high breeding values for milk will tend to have daughters with above-average milk yield, and hence above-average maternal effects
- The value of an offspring can be considered to consist of two components
  - A direct effect (intrinsic breeding value)
  - A maternal contribution

Phenotypic value = direct value + maternal value

$$P_z = P_d + P_m$$

Observable

Latent (unseen) values

Both of the latent values can be further decomposed into breeding plus residual (environmental + non- additive genetic) values

$$P_d = \mu + A_d + E_d, \qquad P_m = \mu + A_m + E_m,$$

The direct breeding value A_d appears in the phenotype of its carrier

The maternal breeding value A_m DOES NOT appear in the phenotype of its carrier, but **rather in the phenotype of her offspring**

Direct vs. maternal breeding values

- The direct and maternal contributions are best thought of as **two separate, but potentially correlated**, traits.
  - Hence, we need to consider $\sigma(A_d,A_m)$ in addition to $\sigma^2(A_d)$ and $\sigma^2(A_m)$. This changes the form of the mixed-model equations
- The direct BV (A_d) is expressed in the individual carrying it
- The maternal BV (A_m) is only expressed in the offspring trait value (and only mom's A_m appears)

Covariance structure

$$\begin{pmatrix} \mathbf{a}_d \\ \\ \mathbf{a}_m \end{pmatrix} \sim \begin{pmatrix} \mathbf{0} \\ \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2(A_d)\,\mathbf{A} & \sigma(A_d, A_m)\,\mathbf{A} \\ \\ \sigma(A_d, A_m)\,\mathbf{A} & \sigma^2(A_m)\,\mathbf{A} \end{pmatrix}$$

This is often written using the Kronecker (or direct) product:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & {}_{mn}\mathbf{B} \end{pmatrix}$$

Giving

$$\begin{pmatrix} \mathbf{a}_d \\ \\ \mathbf{a}_m \end{pmatrix} \sim \begin{pmatrix} \mathbf{0} \\ \\ \mathbf{0} \end{pmatrix}, \mathbf{G} \otimes \mathbf{A} \qquad \mathbf{G} = \begin{pmatrix} \sigma^2(A_d) & \sigma(A_d, A_m) \\ \\ \sigma(A_d, A_m) & \sigma^2(A_m) \end{pmatrix}$$

The mixed-model becomes

Direct effects
breeding values

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_d\mathbf{a}_d + \mathbf{Z}_m\mathbf{a}_m + \mathbf{e}$$

Maternal effects
breeding values

The error structure needs a little care, as the direct E_d and maternal E_m residual values can be correlated*. Initially, we will assume $Var(\mathbf{e}) \sim \sigma_e^2 \mathbf{I}$

*See Bijma 2006 J. Anim. Sci. 84:800-806 for treatment of correlated environmental residuals under this model

The resulting mixed-model equations become

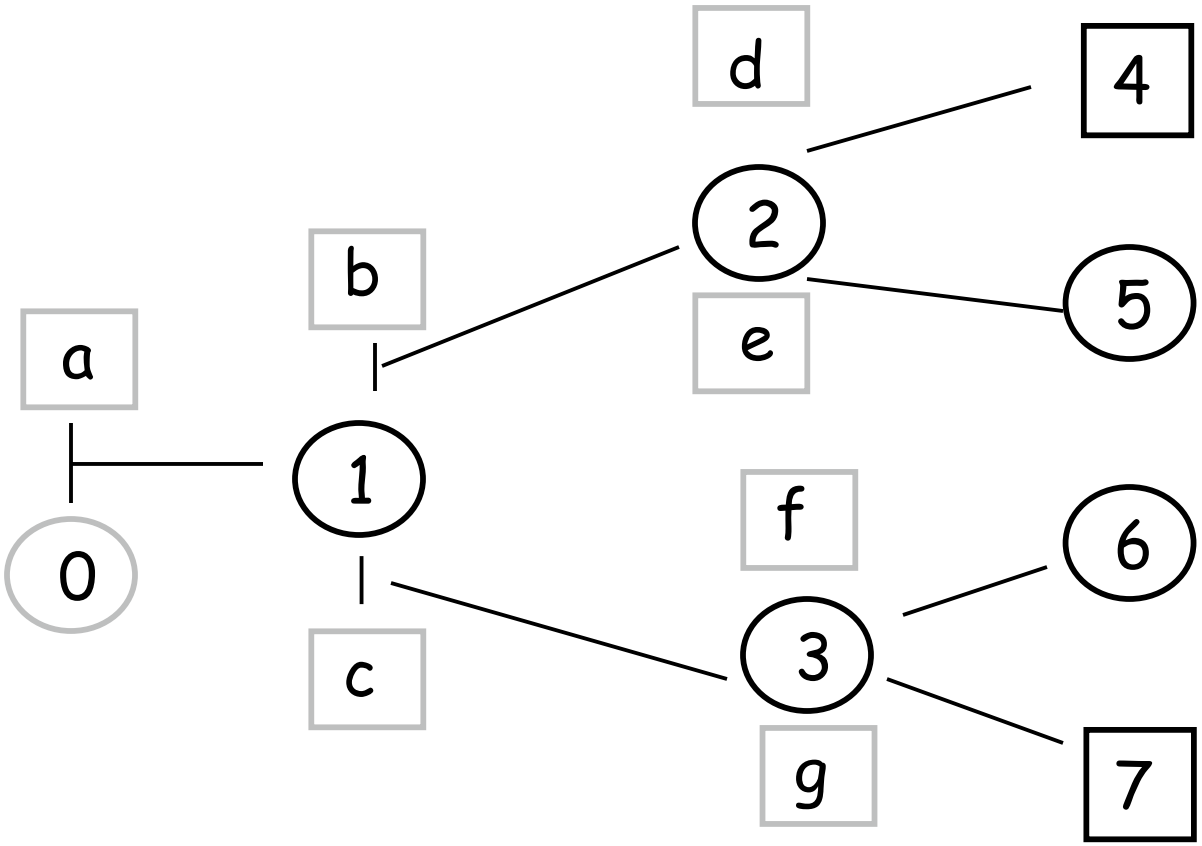
$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z}_d & \mathbf{X}^T\mathbf{Z}_s \\ \mathbf{Z}_d\mathbf{X}^T & \mathbf{Z}_d^T\mathbf{Z}_d + \lambda_1\mathbf{A}^{-1} & \mathbf{Z}_d^T\mathbf{Z}_m + \lambda_2\mathbf{A}^{-1} \\ \mathbf{Z}_m\mathbf{X}^T & \mathbf{Z}_m^T\mathbf{Z}_d + \lambda_2\mathbf{A}^{-1} & \mathbf{Z}_m^T\mathbf{Z}_m + \lambda_3\mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{a}_d \\ \mathbf{a}_m \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}_d^T\mathbf{y} \\ \mathbf{Z}_m^T\mathbf{y} \end{pmatrix}$$

where the weights λ_i are related to elements in the inverse of $\mathbf{G}$, viz.,

$$\begin{pmatrix} \lambda_1 & \lambda_2 \\ \lambda_2 & \lambda_3 \end{pmatrix} = \sigma_e^2\,\mathbf{G}^{-1} = \sigma_e^2 \begin{pmatrix} \sigma^2(A_d) & \sigma(A_d, A_m) \\ \sigma(A_d, A_m) & \sigma^2(A_m) \end{pmatrix}^{-1}$$

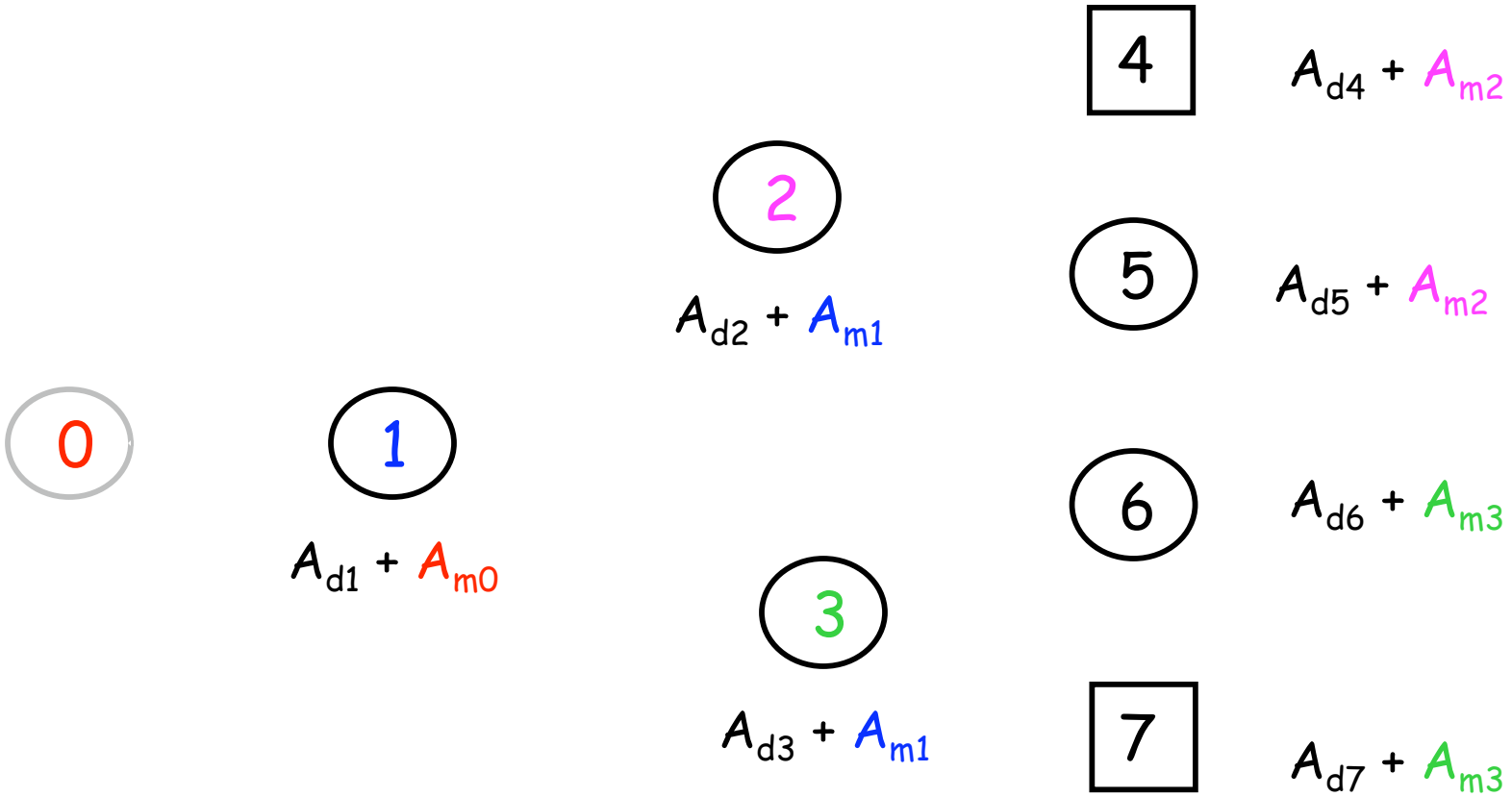
Filling out the maternal effects incident matrix $\mathbf{Z}_m$

A little bookkeeping care is needed when filling out $\mathbf{Z}_m$, because the $\mathbf{A}_m$ associated with a record (measured individual) is that of their mother.



1-7 have records

All sires unrelated



The observed values are y_1 through y_7.
What we can estimate are A_{d1} through A_{d7},
A_{m0} through A_{m3}

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix}, \qquad \mathbf{a}_d = \begin{pmatrix} A_{d,1} \\ A_{d,2} \\ A_{d,3} \\ A_{d,4} \\ A_{d,5} \\ A_{d,6} \\ A_{d,7} \end{pmatrix}, \qquad \mathbf{Z}_d = \mathbf{I}, \qquad \mathbf{a}_m = \begin{pmatrix} A_{m.o} \\ A_{m,1} \\ A_{m,2} \\ A_{m,3} \end{pmatrix}$$

Note that we estimate A_{m0} even though we don't have a record (observation) on her.

Since $\mathbf{Z_m a_m}$ must be a 7 x 1 matrix, $\mathbf{Z_m}$ is 7 x 4 (as $\mathbf{a_m}$ is 4 x 1)

Record 1 is associated with A_{m0}

Records 2 and 3 are associated with A_{m1}

Records 4 and 5 are associated with A_{m2}

Records 6 and 7 are associated with A_{m3}

Record 1 is associated with A_{m0}

Records 2 and 3 are associated with A_{m1}

Records 4 and 5 are associated with A_{m2}

Records 6 and 7 are associated with A_{m3}

$$\mathbf{Z}_m = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \text{as} \quad \mathbf{Z}_m \mathbf{a}_m = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_{m,0} \\ A_{m,1} \\ A_{m,2} \\ A_{m,3} \end{pmatrix} = \begin{pmatrix} A_{m,0} \\ A_{m,1} \\ A_{m,1} \\ A_{m,2} \\ A_{m,2} \\ A_{m,3} \\ A_{m,3} \end{pmatrix}$$

What about A_{m4} through A_{m7}?

Although we have records that only directly relate A_{m0} to A_{m3}, through the use of A we can (in theory) also estimate the maternal breeding values for individuals 4 through 7. Note this includes the maternal BVs for the two males (5 & 7), as they can pass this onto their daughters.

$$\mathbf{Z}_m^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{a}_m^* = \begin{pmatrix} A_{m,0} \\ A_{m,1} \\ A_{m,2} \\ A_{m,3} \\ A_{m,4} \\ A_{m,5} \\ A_{m,6} \\ A_{m,7} \end{pmatrix}$$

Note that

$$\mathbf{Z}_m^* \mathbf{a}_m^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} A_{m,0} \\ A_{m,1} \\ A_{m,2} \\ A_{m,3} \\ A_{m,4} \\ A_{m,5} \\ A_{m,6} \\ A_{m,7} \end{pmatrix} = \begin{pmatrix} A_{m,0} \\ A_{m,1} \\ A_{m,1} \\ A_{m,2} \\ A_{m,2} \\ A_{m,3} \\ A_{m,3} \end{pmatrix}$$

All this raises the question about what can, and cannot, be estimated from the data (**y**) and the design ($\mathbf{Z}_m$, $\mathbf{Z}_d$)?

First issue: Is the structure of the design such that we can estimate all of the variance components. This is the issue of identifiability.

Estimability vs. Identifiability

Details: Identifiability of Variance Components

Due to potential confounding of effects, any particular design might not allow for all variables of interest to be uniquely estimated. For the vector $\boldsymbol{\beta}$ of fixed effects, this is the concept of **estimability** (LW Chapter 26). For $\mathbf{z} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, the vector of fixed effects is estimable (all have unique values) if $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$ exists. Otherwise, some of the fixed effects are confounded and cannot be separated by the design ($\mathbf{X}$) being used. With (co)variance components (often called **dispersal parameters**), a similar concept, **identifiability**, also exists. If variance components are not identifiable in the design, then BLUPs for their associated vectors of random effects do not exist.

Conditions for identifiability of REML estimates of (co)variance components are given by Rothenberg (1971), Jiang (1996), and Cantet and Cappa (2008). Before presenting these, we first review a few details about REML. Recall (LW Chapter 27) that REML estimates are those that maximize that part of the likelihood function that is independent of the fixed effects (this is often stated as being the **translation invariant** part). Let $\mathbf{V}$ be the covariance matrix of $\mathbf{z}$, which is a function of its variance components. As detailed in LW Chapter 27, Harville (1977) shows that (if it exists) the transformation provided by the matrix

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1} \tag{1a}$$

plays a critical role in REML estimates. That this matrix can remove fixed effects can be seen by noting that

$$\mathbf{P}\mathbf{z} = \mathbf{V}^{-1}\left(\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) \tag{1b}$$

yields a vector that is the data vector adjusted by the (estimated) fixed effects. Now consider covariance structures of the form

$$\mathbf{V} = \sum_{i=1}^{n} \mathbf{V}_i \theta_i \tag{2a}$$

where $\mathbf{V}_i$ is a matrix of known constants and the θ_i are unknown variances and covariances to be estimated.

The equations to maximize the likelihood over the restricted space (the REML estimates) are given by LW Equations 27.18 and 27.19, and are solved iteratively. These equations involve the **trace** (sum of the diagonal elements) of matrix products involving $\mathbf{P}$ and the $\mathbf{V}_i$. Recall (LW Appendix 4) that for a vector $\boldsymbol{\Theta}$ of n unknowns, the Fisher information matrix $\mathbf{F}$ (the matrix of second partial derivatives of the likelihood with respect to the parameters) can be used to provide large-sample standard errors. The resulting $n \times n$ information matrix for REML estimates of the unknown θ_i in Equation 2a is

$$F_{ij} = \text{trace}\,(\mathbf{P}\mathbf{V}_i\mathbf{P}\mathbf{V}_j) \tag{2b}$$

Much in the same fashion that the existence of $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$ informs us that all fixed effects are estimable in a given design, all variance components θ_i are identifiable if all of the eigenvalues of $\mathbf{F}$ are positive, that is, that $\mathbf{F}$ is positive-definite (Rothenberg 1971, Jiang 1996). For the maternal effects mixed model, Equation 2a becomes

$$\mathbf{V} = \mathbf{V}_1\,\sigma^2(A_d) + \mathbf{V}_2\,\sigma(A_d, A_s) + \mathbf{V}_3\,\sigma^2(A_s) + \mathbf{V}_4\,\sigma_e^2 \tag{3a}$$

where

$$\mathbf{V}_1 = \mathbf{Z}_d\mathbf{A}\mathbf{Z}_d^T, \quad \mathbf{V}_2 = \left(\mathbf{Z}_d\mathbf{A}\mathbf{Z}_m^T + \mathbf{Z}_m\mathbf{A}\mathbf{Z}_d^T\right), \quad \mathbf{V}_3 = \mathbf{Z}_m\mathbf{A}\mathbf{Z}_s^T, \quad \mathbf{V}_4 = \mathbf{I} \tag{3b}$$

Substituting Equations 1a and 3b into Equation 2b fills out the $\mathbf{F}$ matrix (which is only 4×4 in this case given the four unknown variance components). For any particular design, the eigenvalues of this matrix can be computed to determine if the variance components are all identifiable.

Second issue, connectivity

Even if the design is such that we can estimate all the genetic variances, whether we can estimate all of the β, $\mathbf{a}_d$, and $\mathbf{a}_m$ in the model depends on whether a unique inverse exists for the MME:

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z}_d & \mathbf{X}^T\mathbf{Z}_s \\ \mathbf{Z}_d\mathbf{X}^T & \mathbf{Z}_d^T\mathbf{Z}_d + \lambda_1\mathbf{A}^{-1} & \mathbf{Z}_d^T\mathbf{Z}_m + \lambda_2\mathbf{A}^{-1} \\ \mathbf{Z}_m\mathbf{X}^T & \mathbf{Z}_m^T\mathbf{Z}_d + \lambda_2\mathbf{A}^{-1} & \mathbf{Z}_m^T\mathbf{Z}_m + \lambda_3\mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \mathbf{a}_d \\ \mathbf{a}_m \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}_d^T\mathbf{y} \\ \mathbf{Z}_m^T\mathbf{y} \end{pmatrix}$$

Unique estimates of all the β require $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$ exists

If $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$ does not exist, a generalized inverse is used which can uniquely estimate k linear combinations of the β where k is the rank of $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$ (LW Appendix 3).

Likewise, if the MME equation does not have an inverse (and this is not due to constraints on β), then a generalized inverse can be used to estimate unique estimates of certain linear combinations of the $\mathbf{a}_d$ and $\mathbf{a}_m$.

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z}_d & \mathbf{X}^T\mathbf{Z}_s \\ \mathbf{Z}_d\mathbf{X}^T & \mathbf{Z}_d^T\mathbf{Z}_d + \lambda_1\mathbf{A}^{-1} & \mathbf{Z}_d^T\mathbf{Z}_m + \lambda_2\mathbf{A}^{-1} \\ \mathbf{Z}_m\mathbf{X}^T & \mathbf{Z}_m^T\mathbf{Z}_d + \lambda_2\mathbf{A}^{-1} & \mathbf{Z}_m^T\mathbf{Z}_m + \lambda_3\mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \mathbf{a}_d \\ \mathbf{a}_m \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}_d^T\mathbf{y} \\ \mathbf{Z}_m^T\mathbf{y} \end{pmatrix}$$

A key role in ensuring that unique estimates of $\mathbf{a}_d$ and $\mathbf{a}_m$ exist is played by the relationship matrix A. If individuals with records and individuals without records are sufficiently well connected (non-zero entries in A for their pair-wise relatedness), then we usually can estimate values of un-observed individuals (although their precision is another issue).