Lecture 4
BLUP Breeding Values

Guilherme J. M. Rosa

University of Wisconsin-Madison

Mixed Models in Quantitative Genetics

SISG, Seattle

20 - 22 July 2022

Mixed Models in Animal and Plant Breeding

Animal/plant breeding programs are based on the principle that phenotypic observations on related individuals can provide information about their underlying genotypic values

The additive component of genetic variation is the primary determinant of the degree to which offspring resemble their parents, and therefore this is usually the component of interest in artificial selection programs

Mixed Models in Animal and Plant Breeding

Many statistical methods for analysis of genetic data are specific (or more appropriate) for phenotypic measurements obtained from planned experimental designs and with balanced data sets

While such situations may be possible within laboratory or greenhouse experimental settings, data from natural populations and agricultural species are generally highly unbalanced and fragmented by numerous kinds of relationships

Animal Model

Culling of data to accommodate conventional statistical techniques (e.g. ANOVA) may introduce bias and/or lead to a substantial loss of information

The mixed model methodology allows efficient estimation of genetic parameters (such as variance components and heritability) and breeding values while accommodating extended pedigrees, unequal family sizes, overlapping generations, sex-limited traits, assortative mating, and natural or artificial selection

To illustrate such application of mixed models in breeding programs, we consider here the so-called Animal Model in situations with a single trait and a single observation (including missing values) per individual

Animal Model

The animal model can be described as:

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

$\mathbf{y}$ is an $(n \times 1)$ vector of observations (phenotypic scores)

$\boldsymbol{\beta}$ is a $(p \times 1)$ vector of fixed effects (e.g. herd-year-season effects)

$\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{G})$ is a $(q \times 1)$ vector of breeding values (relative to all individuals with record or in the pedigree file, such that q is in general bigger than n)

$\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_n\sigma_e^2)$ represents residual effects, where σ_e^2 is the residual variance

The Matrix *A*

The matrix ***G*** describing the covariances among the random effects (here the breeding values) follows from standard results for the covariances between relatives

It can be shown that the additive genetic covariance between two relatives i and i' is given by $2\theta_{ii'}\sigma^2_a$, where $\theta_{ii'}$ is the coefficient of coancestry between individuals i and i', and σ^2_a is the additive genetic variance in the base population

Hence, under the animal model, $\mathbf{G} = \mathbf{A}\sigma^2_a$, where ***A*** is the additive genetic (or numerator) relationship matrix, having elements given by $a_{ii'} = 2\theta_{ii'}$

The Matrix *A*

For each animal i in the pedigree (i = 1, 2,…,n), going from older to younger animals, compute a_{ii} and a_{ij} (j = 1, 2,…,i-1) as follows:

If both parents (s and d) of animal i are known:

$$a_{ij} = a_{ji} = (a_{js} + a_{jd})/2 \text{ and } a_{ii} = 1 + a_{sd}/2$$

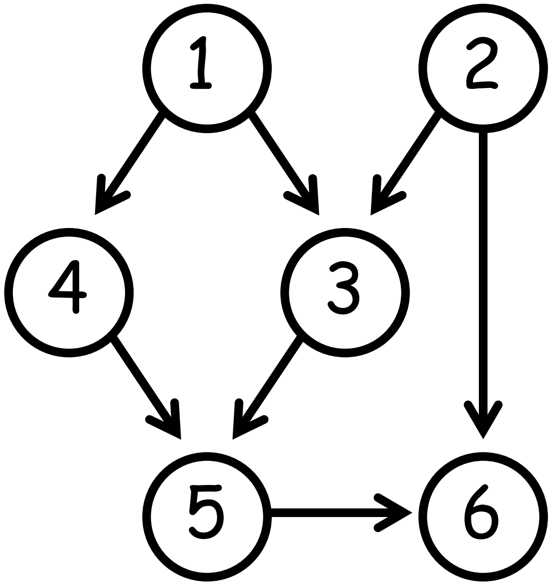
If only one parent (e.g. d) of animal i is known:

$$a_{ij} = a_{ji} = a_{jd}/2 \text{ and } a_{ii} = 1$$

If parents unknown:

$$a_{ij} = a_{ji} = 0 \text{ and } a_{ii} = 1$$

Example



Animal	Sire	Dam
1	-	-
2	-	-
3	1	2
4	1	-
5	4	3
6	5	2

$$A = \begin{bmatrix} 1 & 0 & .5 & .5 & .5 & .25 \\ 0 & 1 & .5 & 0 & .25 & .625 \\ .5 & .5 & 1 & .25 & .625 & .563 \\ .5 & 0 & .25 & 1 & .625 & .313 \\ .5 & .25 & .625 & .625 & 1.125 & .688 \\ .25 & .625 & .563 & .313 & .688 & 1.125 \end{bmatrix}$$

Animal Model

In general, in animal/plant breeding interest is on prediction of breeding values (for selection of superior individuals), and on estimation of variance components and functions thereof, such as heritability

The fixed effects are, in some sense, nuisance factors with no central interest in terms of inferences, but which need to be taken into account (i.e., they need to be corrected for when inferring breeding values)

Animal Model

Since under the animal model $\mathbf{G}^{-1} = \mathbf{A}^{-1}\sigma_a^{-2}$ and $\mathbf{R}^{-1} = \mathbf{I}_n\sigma_e^{-2}$, the mixed model equations can be expressed as:

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

where $\lambda = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2}$, such that:

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

Conditional on the variance components ratio λ, the BLUP of the breeding values are given then by:

$$\hat{\mathbf{u}} = (\mathbf{Z}^{\mathrm{T}}\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1}\mathbf{Z}^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

These are generally referred to as Estimated Breeding Values (EBV)

Alternatively, some breeders associations express their results as Predicted Transmitting Abilities (PTA) (or Estimated Transmitting Abilities (ETA) or Expected Progeny Difference (EPD)), which are equal to half the EBV, representing the portion of an animal's breeding values that is passed to its offspring

The amount of information contained in an animal's genetic evaluation depends on the availability of its own record, as well as how many (and how close) relatives it has with phenotypic information

As a measure of amount of information in livestock genetic evaluations, EBVs are typically reported with its associated accuracies

Accuracy of predictions is defined as the correlation between true and estimated breeding values, i.e., $r_i = \rho(\hat{u}_i, u_i)$

Instead of accuracy, some livestock species genetic evaluations use reliability, which is the squared correlation of accuracy (r_i^2)

Prediction Accuracy

The calculation of $\rho(\hat{u}_i, u_i)$ requires the diagonal elements of the inverse of the MME coefficient matrix, represented as:

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta u} \\ \mathbf{C}^{u\beta} & \mathbf{C}^{uu} \end{bmatrix}$$

It can be shown that the prediction error variance of EBV $\hat{u}_i$ is given by:

$$\mathrm{PEV} = \mathrm{Var}(\hat{u}_i - u_i) = c_i^{uu}\sigma_e^2$$

where c_i^{uu} is the i-th diagonal element of $\mathbf{C}^{uu}$, relative to animal i.

Prediction Accuracy

The PEV can be interpreted as the fraction of additive genetic variance not accounted for by the prediction

Therefore, PEV can be expressed also as:

$$\mathrm{PEV} = (1 - r_i^2)\sigma_a^2$$

such that $c_i^{uu}\sigma_e^2 = (1 - r_i^2)\sigma_a^2$, from which the reliability is obtained as:

$$r_i^2 = 1 - c_i^{uu}\sigma_e^2 / \sigma_a^2 = 1 - \lambda c_i^{uu}$$

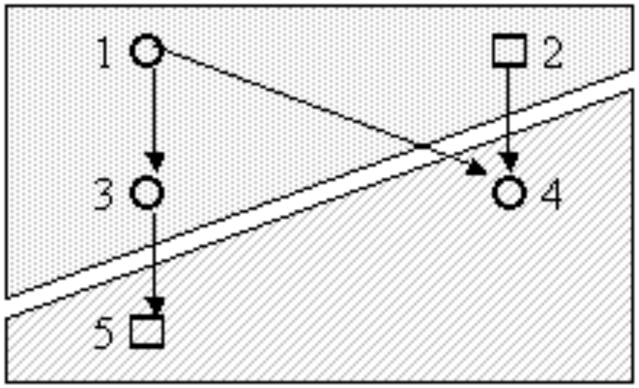
PAUSE

- ⇨ Animal Model
- ⇨ Numerator relationship matrix
- ⇨ EBV and prediction accuracy
- ⇨ Next: Examples

Next PAUSE, slide 21

Animal Model

herd 1



herd 2

Animal	Sire	Dam	Herd	Observation
1	–	–	h1	310
2	–	–	h1	–
3	–	1	h1	270
4	2	1	h2	350
5	–	3	h2	–

$$\underset{y}{\begin{bmatrix}310\\270\\350\end{bmatrix}} = \underbrace{\begin{bmatrix}1 & 0\\1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}h_1\\h_2\end{bmatrix}}_{X\,\beta} + \underbrace{\begin{bmatrix}1 & 0 & 0 & 0 & 0\\0 & 0 & 1 & 0 & 0\\0 & 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\\u_4\\u_5\end{bmatrix}}_{Z\,u} + \underset{e}{\begin{bmatrix}e_1\\e_3\\e_4\end{bmatrix}}$$

Animal Model

Breeding values: $\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \mathbf{A}\sigma_{\mathrm{u}}^2)$, with

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0.25 \\ 0 & 1 & 0 & 0.5 & 0 \\ 0.5 & 0 & 1 & 0.25 & 0.5 \\ 0.5 & 0.5 & 0.25 & 1 & 0.125 \\ 0.25 & 0 & 0.5 & 0.125 & 1 \end{bmatrix}$$

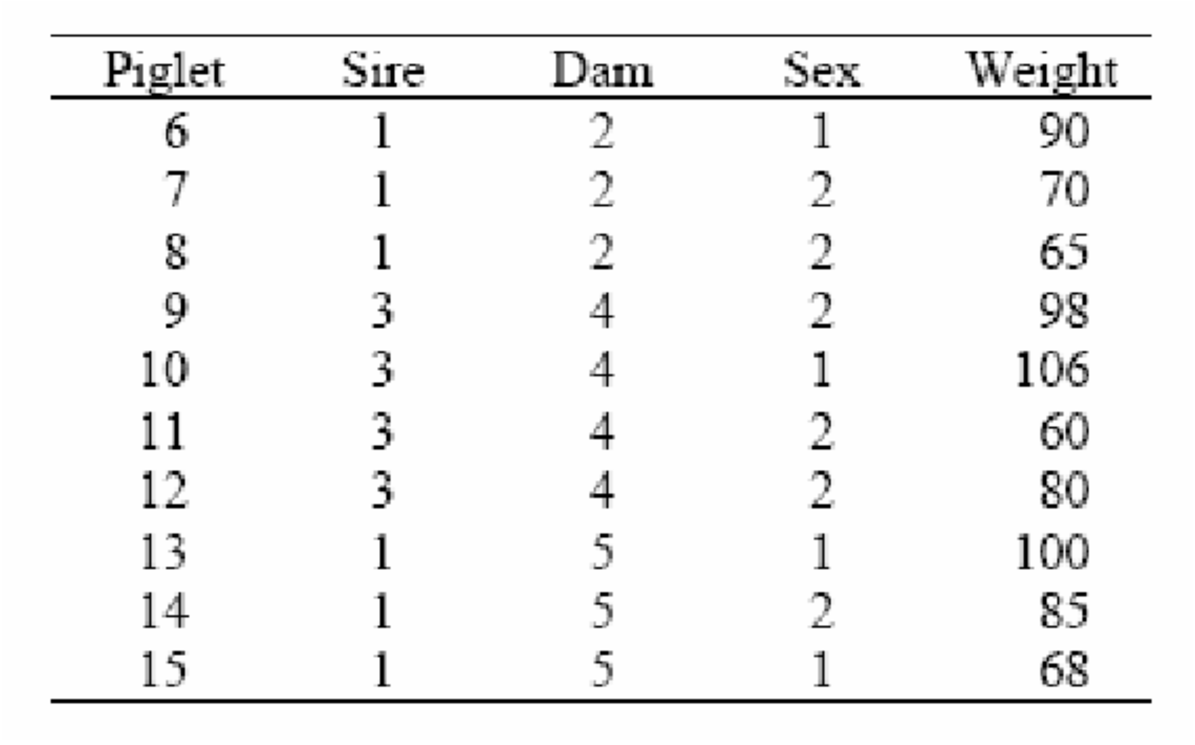
$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{Z} \\ \mathbf{Z}^{\mathrm{T}}\mathbf{X} & \mathbf{Z}^{\mathrm{T}}\mathbf{Z} + \boldsymbol{\lambda}\mathbf{A}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^{\mathrm{T}}\mathbf{y} \\ \mathbf{Z}^{\mathrm{T}}\mathbf{y} \end{bmatrix}$$

$$\boldsymbol{\lambda} = \frac{\boldsymbol{\sigma}_{\mathrm{e}}^2}{\boldsymbol{\sigma}_{\mathrm{u}}^2} = \frac{1 - \mathrm{h}^2}{\mathrm{h}^2}$$

Animal Model

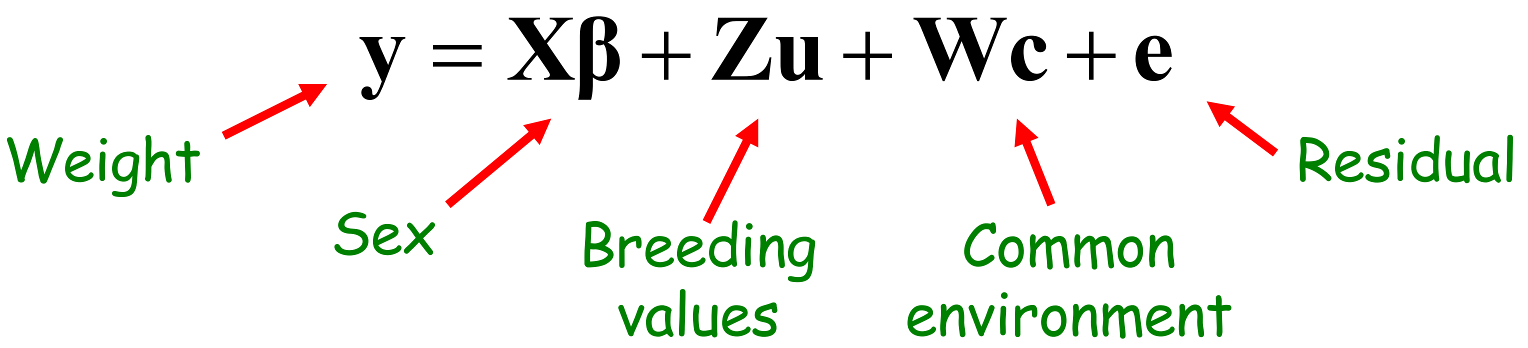
The animal model can be extended to model multiple (correlated) traits, multiple random effects (such as maternal effects and common environmental effects), repeated records (e.g. test day models), and so on

Example (Mrode 1996, pp74-76): Weaning weight (kg) of piglets, progeny of three sows mated to two boars:



Piglet	Sire	Dam	Sex	Weight
6	1	2	1	90
7	1	2	2	70
8	1	2	2	65
9	3	4	2	98
10	3	4	1	106
11	3	4	2	60
12	3	4	2	80
13	1	5	1	100
14	1	5	2	85
15	1	5	1	68

A linear model with the (fixed) effect of sex, and the (random) effects of common environment (related to each litter) and breeding values can be expressed as **X**:



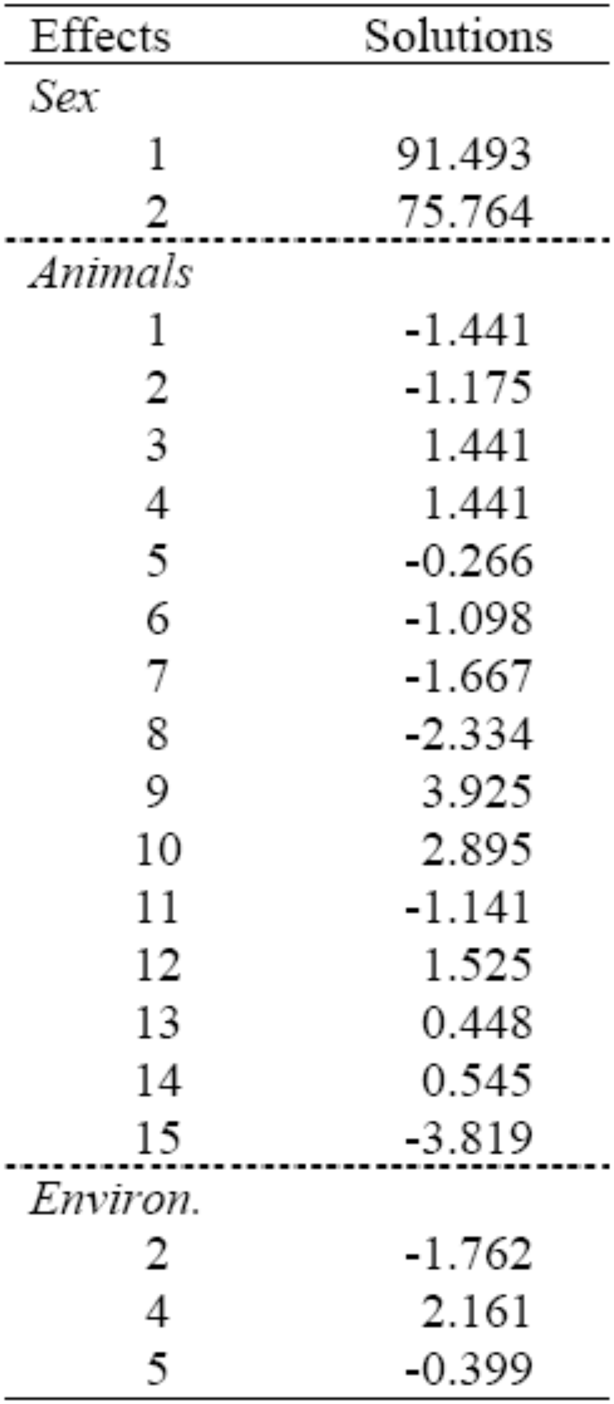
$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{e}$$

Assuming that $\sigma^2_u = 20$, $\sigma^2_c = 15$ and $\sigma^2_e = 65$, the MME are as follows:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} & \mathbf{X}^T\mathbf{W} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z}+\mathbf{A}^{-1}\lambda_1 & \mathbf{Z}^T\mathbf{W} \\ \mathbf{W}^T\mathbf{X} & \mathbf{W}^T\mathbf{Z} & \mathbf{W}^T\mathbf{W}+\mathbf{I}\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol\beta} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \\ \mathbf{W}^T\mathbf{y} \end{bmatrix}$$

where $\lambda_1 = \frac{\sigma^2_e}{\sigma^2_u} = 3.25$ and $\lambda_2 = \frac{\sigma^2_e}{\sigma^2_c} = 4.\dot{3}$

The BLUEs and BLUPs (inverting the numerator relationship matrix) are:



Effects	Solutions
Sex	
1	91.493
2	75.764
Animals	
1	-1.441
2	-1.175
3	1.441
4	1.441
5	-0.266
6	-1.098
7	-1.667
8	-2.334
9	3.925
10	2.895
11	-1.141
12	1.525
13	0.448
14	0.545
15	-3.819
Environ.	
2	-1.762
4	2.161
5	-0.399

PAUSE

- ⇨ Examples with single trait analysis
- ⇨ Models with multiple random effects

Next PAUSE, slide 35

Animal Model Extensions

- Multiple-trait Model
- Repeatability Model
- Maternal Effects
- Generalized Linear Models

Multiple (Correlated) Traits

The animal model can be extended for the joint analysis of multiple traits

Let the model for each of k traits be:

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\varepsilon}_j$$

where j is an index to indicate the trait (j = 1, 2,…,k).

For the joint analysis of the k trait, the model becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon}$$

with design matrices given by:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_k \end{bmatrix} \qquad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_k \end{bmatrix}$$

Multiple (Correlated) Traits

In this case it is assumed that:

$$\mathrm{Var}\begin{bmatrix} \mathbf{a} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \otimes \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \otimes \mathbf{I} \end{bmatrix}$$

where **G** and **Σ** are the genetic and residual variance-covariance matrices, given by:

$$\mathrm{G} = \begin{bmatrix} \sigma^2_{a_1} & \sigma_{a_1 a_2} & \cdots & \sigma_{a_1 a_k} \\ \sigma_{a_1 a_2} & \sigma^2_{a_2} & \cdots & \sigma_{a_2 a_2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{a_1 a_k} & \sigma_{a_2 a_k} & \cdots & \sigma^2_{a_k} \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{\boldsymbol{\varepsilon}_1} & \sigma_{\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_2} & \cdots & \sigma_{\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_k} \\ \sigma_{\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_2} & \sigma^2_{\boldsymbol{\varepsilon}_2} & \cdots & \sigma_{\boldsymbol{\varepsilon}_2 \boldsymbol{\varepsilon}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_k} & \sigma_{\boldsymbol{\varepsilon}_2 \boldsymbol{\varepsilon}_2} & \cdots & \sigma^2_{\boldsymbol{\varepsilon}_k} \end{bmatrix}$$

Note: $\otimes$ represents the direct (Kronecker) product

Multiple (Correlated) Traits

The MME for multi-trait analyses are of the same form as before, i.e.:

$$\begin{bmatrix} \mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X} & \mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z} \\ \mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X} & \mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{Z} + \mathbf{G}^{-1} \otimes \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y} \\ \mathbf{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y} \end{bmatrix}$$

from which the BLUEs and BLUPs of **β** and **a** can be obtained.

Multiple (Correlated) Traits

The dimensionality of multi-trait MME, however, can become a hurdle for solving it when more than two or three traits are considered

An alternative for the analysis of multiple traits is to use a canonical transformation of the traits, which consists of transforming the vectors of correlated traits into a new vector of uncorrelated variables

In such case, each transformed variable can be analyzed independently using standard single trait models, and subsequently the estimated breeding values are transformed back to the original scale of measurement

Repeatability Model

Repeatability Model

For the analysis of repeated measurements, environmental effects can be partitioned into permanent and temporary effects

In this case, the mixed model, usually called ‘repeatability model’, can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Za} + \mathbf{Wp} + \boldsymbol{\varepsilon}$$

where $\mathbf{p} \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$ is the vector of permanent environmental effects, with each level pertaining to a common effect to all observations of each animal

Repeatability Model

It is often assumed that **a**, **p**, and **ε**, which are independent from each other

Under these assumptions, the MME becomes:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\lambda_a\mathbf{A}^{-1} & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W}+\lambda_p\mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix}$$

with $\lambda_a = \sigma_{\boldsymbol{\varepsilon}}^2/\sigma_a^2$ and $\lambda_p = \sigma_{\boldsymbol{\varepsilon}}^2/\sigma_p^2$

Maternal Effects

Maternal Effects

There are some traits of interest in livestock, such as weaning weight in beef cattle, in which progeny performance is affected by the dam's ability to affect the calf's environment, such as in the form of nourishment through her milk production, the quantity and quality of which is in part genetically determined

In such cases, dams contribute to the performance of their progeny not only through the genes passed to the progeny (the "direct genetic effects") but also through their ability to provide a suitable environment (the "indirect genetic effects")

Maternal Effects

Maternally influenced traits can be analyzed by using a model as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Za} + \mathbf{Km} + \mathbf{Wp} + \boldsymbol{\varepsilon}$$

where **m** is a vector of random maternal genetic effects, and **p** is a vector of random maternal permanent environmental effects

It is assumed that **m** ~ $N(\mathbf{0}, \mathbf{A}\sigma_m^2)$ and **p** ~ $N(\mathbf{0}, \mathbf{I}\sigma_p^2)$, and quite often a covariance structure between direct and maternal additive genetic effects is considered, assumed equal to $A\sigma_{a,m}$

Example

Animal	Sire	Dam	CG	Weight
5	1	3	1	156
6	2	3	1	124
7	1	4	1	135
8	2	4	2	163
9	1	3	2	149
10	2	4	2	138

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{p} + \mathbf{e}$$

$$Var\begin{pmatrix} \mathbf{a} \\ \mathbf{m} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{am} & \mathbf{0} & \mathbf{0} \\ \mathbf{A}\sigma_{am} & \mathbf{A}\sigma_m^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_p^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{Z}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{Z}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{Z}_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(Schaeffer, 2017)

Computing Strategies

Solving the MME does not necessary require the inversion of the coefficient matrix ***C***

More computationally convenient alternatives for solving high dimensional systems of linear equations include methods based on iteration on the MME, such as the Jacobi or Gauss-Seidel iteration, and the “iteration on the data” strategy, which is commonly used methodology in national genetic evaluations involving millions of records

PAUSE

- ⇨ Multi-trait analysis
- ⇨ Repeatability Model
- ⇨ Maternal Effects

Next PAUSE, slide 42 (end)

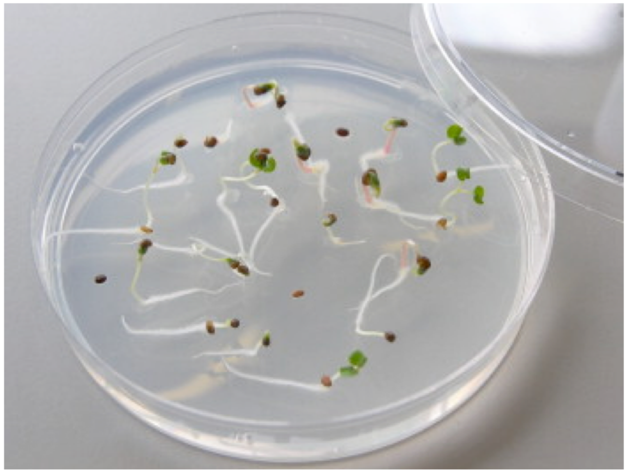
Generalized Linear Models

The models discussed so far assumed a Gaussian (normal) distribution of the phenotypic traits

Often however phenotypic traits are expressed a a binary (e.g., pregnancy in dairy cattle, or germination in seeds) or count variable (e.g., litter size in swine, or fruits in trees)

In such cases the linear (Gaussian) model is not appropriate, and a generalized linear model (GLM) approach is necessary

Generalized Linear Models



Generalized Linear Models

GLM can actually model outcomes (response variables) generated from any distribution from the exponential family, which includes the normal, binomial, Poisson and gamma distributions, among others

The GLM consists of three elements:

1. Probability distribution from the exponential family.
2. Linear predictor $\eta = X\beta$
3. Link function g such that $E(Y) = \mu = g^{-1}(\eta)$.

Generalized Linear Mixed Models

Notice that the Gaussian model is a specific case of the GLM, with the normal distribution and an identity link function

In the case of Generalized Linear Mixed Models, including the applications in animal/plant breeding, the model is defined as:

1. Probability distribution from the exponential family.
2. Linear predictor $\eta = X\beta + Zu$
3. Link function g such that $E(Y|u) = \mu = g^{-1}(\eta)$

GLMM in R

GLMM can be implemented in R using the package lme4

lme4, however, assumes independence between levels of random effects, and as such it is not suitable for many animal/plant breeding applications

pedigreemm is an R package that uses lme4 with a Cholesky decomposition strategy to overcome this problem

pedigreemm

An R package for fitting generalized linear mixed models in animal breeding

$$\boxed{g\left(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}}\right) = \mathbf{Zu} + \mathbf{X}\boldsymbol{\beta}}$$

$$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}} = E\left[\mathbf{Y}|\mathbf{U} = \mathbf{u}\right] \qquad \mathbf{u} \sim N\left(\mathbf{0}, \mathbf{A}\sigma_u^2\right)$$

$$\mathbf{u}^* = \mathbf{L}^{-1}\mathbf{u} \longrightarrow g\left(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}}\right) = \mathbf{ZL}\left(\mathbf{L}^{-1}\mathbf{u}\right) + \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}^*\mathbf{u}^* + \mathbf{X}\boldsymbol{\beta}$$

$$\mathbf{A} = \mathbf{LL}' \qquad \mathbf{u}^* \sim N\left(\mathbf{0}, \mathbf{I}\sigma_u^2\right)$$

(Harville and Callanan 1989)

Technical note: An R package for fitting generalized linear mixed models in animal breeding[1]

A. I. Vazquez,*[2] D. M. Bates,† G. J. M. Rosa,* D. Gianola,*‡ and K. A. Weigel*

*Department of Dairy Science, †Department of Statistics, and ‡Department of Animal Sciences,
University of Wisconsin, Madison 53706

J. Anim. Sci. 2010. 88:497–504
doi:10.2527/jas.2009-1952

Data Set 1. Milk production records of 3,397 lactations from first- through fifth-parity Holsteins were available. These records were from 1,359 cows, daughters of 38 sires in 57 herds. Records are in the *milk* data set in the *pedigreemm* package. The data were downloaded from the USDA site (http://www.aipl.arsusda.gov/). All lactation records represent cows with at least 100 d in milk, with an average of 347 d. Milk yield ranged from 4,065 to 19,345 kg estimated for 305 d, averaging 11,636 kg. There were 1,314, 1,006, 640, 334, and 103 records for first-, second-, third-, fourth-, and fifth-lactation animals, respectively. A 5-generation pedigree of the cows with a total of 6,547 animals was used in the analysis (http://www.aipl.arsusda.gov/). The pedigree information is available in the *pedCows* and *pedCowsR* pedigree objects also included in the package; the second one is a lighter pedigree (with 70% of the information on *pedCows*). The milk production data used in the first 2 examples are described below.