

# Introduction to Mixed Linear Models

*Vincent Ducrocq*

GABI, Jouy-en-Josas  
France



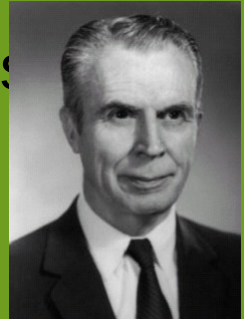
# Historical aspects

- A long time ago...
  - Selection / genetic improvement of livestock = visual appreciation
  - creation of dairy cattle breeds in Europe : 19th century
- Early 20th century: milk recording starts in Europe (Denmark)
  - « selection criterion »: best production (milk or fat) in one day for cows
- Then, daughter-dam comparison for bulls... ( $\hat{b}_{\text{bull}} = \text{daughter} - \text{dam}$ )
- Then, herdmate comparison
  - compare the average production of daughters of a bull with the average production of their herdmates □ significant genetic gain



Quickly, unfair comparison because herd mates of daughters of young bulls are often daughters of selected (better) bulls

□ Modified herd mate comparison □ 1989 in the USA...



Meanwhile: 1949-1952: **C.R. Henderson**'s thesis

The basis of « mixed models » and variance component estimation

1959: a theoretical statistical treatment of mixed models

1963: a first presentation of BLUP

**1973: « Sire evaluation and genetic trends »**

**1976: « A method to compute the inverse of the numerator relationship matrix »**

# C.R. Henderson's framework

We want to estimate/**predict** the additive genetic effect of one ( $a_i$ ) or several ( $\mathbf{a}$ ) animals, from the available data ( $\mathbf{y}$ )

**Case 1:** we know **all genetic parameters** (*genetic variances and covariances* (or equivalently: genetic correlations) and *heritabilities*)  
+ we know how to correct for environmental conditions

(e.g., very large groups of animals with similar characteristics are raised in exactly the same conditions – as in poultry breeding)

□ Take  $\hat{\mathbf{a}}$  = **Best Linear Prediction** of  $\mathbf{a}$  (BLP)

In fact = Selection index theory (Smith, 1936; Hazel,

1943)

# C.R. Henderson's framework

**Case 2:** we know **all genetic parameters** (*genetic variances and covariances* (or equivalently: genetic correlation) and *heritabilities*)

+ ~~we know how to correct for environmental conditions~~

(e.g., in cattle breeding : every month, there are new cows calving, at a different age, and different lactation number, in small herds with different feeding systems)

□ we need to correct the data for environmental effects  $\beta$

# C.R. Henderson's framework

**Case 3** ~~we know all genetic parameters (genetic variances and covariances (or equivalently: genetic correlation) and heritabilities)~~  
(e.g., because we are studying a new breed, a new trait, a new environment)  
• ~~we know how to correct for environmental conditions~~

□ we need to estimate genetic parameters (« variance components »)  
□ Then take  $\hat{\mathbf{a}}$  = **Best Linear Unbiased Prediction** of  $\mathbf{a}$  (**BLUP**)  
assuming that the estimated genetic parameters are the correct ones

This is a very strong assumption, **especially when we don't have much data** to estimate genetic parameters (but, in that case, a Bayesian approach can be used to do both at the same time (estimation of genetic parameters + genetic evaluation))

• With large data sets, it is not so important as long as genetic parameters are estimated by **Restricted Maximum Likelihood** (**REML**)

# BLUP: objective

$$E[\hat{\omega}_i] = E[\omega_i]$$

# BLUP: relationship with BLP





# Mixed Model Equations

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}y \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}y \end{bmatrix} \begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

where:

- $R$  is the matrix of (co)variances between residuals (with relationships accounted)
- $G^{-1}$  = inverse of  $G$ , the matrix of **genetic (co)variances**

Advantages compared to BLP:

- no intermediate computation of  $bi$
- $R^{-1}$  and  $G^{-1}$  have a (known) simple structure (much simpler than  $R$  and  $G$ )
- Size of the system = size of  $\beta$  + size of  $a$  (and not size of  $y$ )

# Mixed Model Equations: example

$$\begin{array}{ccccc} \begin{array}{c} \text{?} \\ \text{?} \\ \text{?} \end{array} X' R^{-1} X & X' R^{-1} Z & \begin{array}{c} \text{?} \beta \text{?} \\ \text{?} \text{?} \text{?} \\ \text{?} \alpha \text{?} \end{array} & \begin{array}{c} \text{?} X' R^{-1} y \text{?} \\ \text{?} \text{?} \text{?} \\ \text{?} Z' R^{-1} y \text{?} \end{array} \\ \begin{array}{c} \text{?} \\ \text{?} \\ \text{?} \end{array} Z' R^{-1} X & Z' R^{-1} Z + \mathbf{G}^{-1} & & \end{array}$$

For a univariate evaluation, with an additive genetic effect

Assume iid (independent and identically distributed) residuals

$$\square \mathbf{R} = \mathbf{I} \sigma_e^2$$

Var(additive genetic effect) =  $\mathbf{G} = \mathbf{A}$  ( $\mathbf{A}$  = relationship matrix)

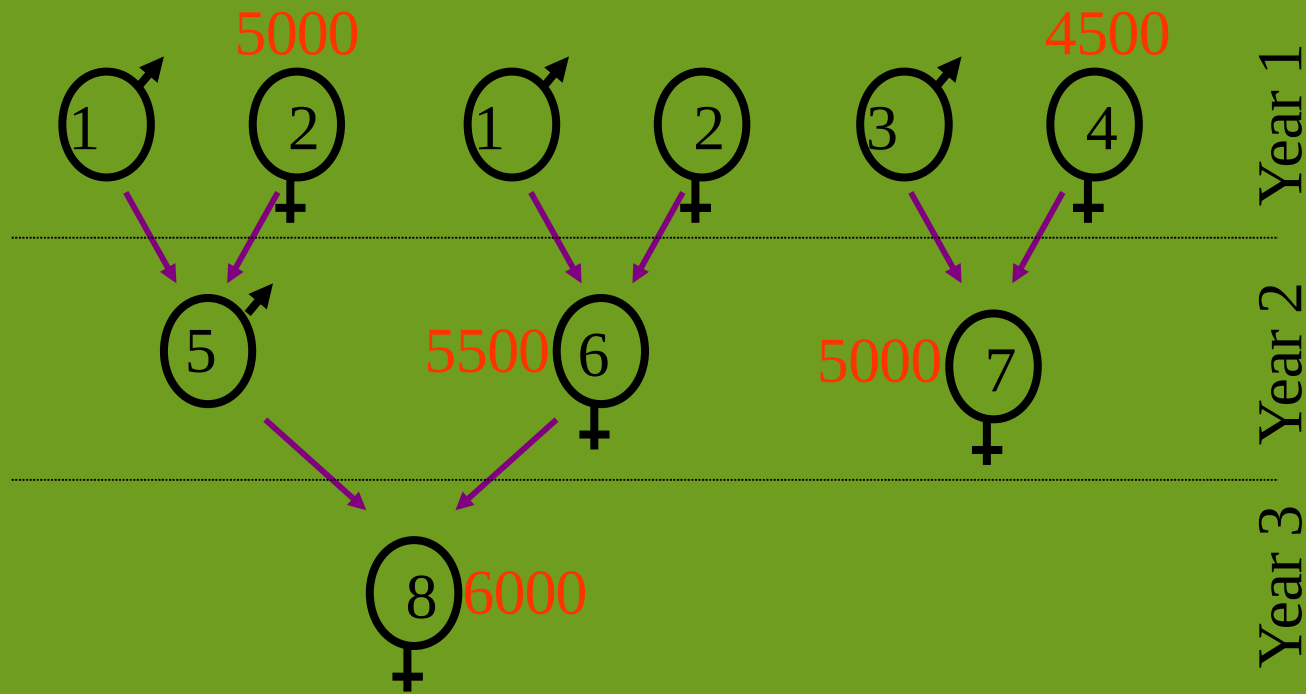
Multiplying both sides by  $\sigma_e^2$

$$\begin{array}{ccccc} \begin{array}{c} \text{?} \\ \text{?} \\ \text{?} \end{array} X' X & X' Z & \begin{array}{c} \text{?} \beta \text{?} \\ \text{?} \text{?} \text{?} \\ \text{?} \alpha \text{?} \end{array} & \begin{array}{c} \text{?} X' y \text{?} \\ \text{?} \text{?} \text{?} \\ \text{?} Z' y \text{?} \end{array} \\ \begin{array}{c} \text{?} \\ \text{?} \\ \text{?} \end{array} Z' X & Z' Z + \alpha \mathbf{A}^{-1} & & \end{array}$$

$$\text{where } \alpha = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2}$$

Note: all this can also be obtained with a Bayesian approach

# Example



$$y_i = \beta_j + a_i + e_i$$

# Mixed Model Equations: Construction

- for the small example:

$$\begin{array}{l}
 \mathbf{X}'\mathbf{X} = \begin{bmatrix} ? & 0 & 0 & ? \\ ? & 2 & 0 & ? \\ ? & 0 & 1 & ? \end{bmatrix} \quad \mathbf{X}'\mathbf{Z} = \begin{bmatrix} ? & 1 & 0 & 1 & 0 & 0 & 0 & 0 & ? \\ ? & 0 & 0 & 0 & 0 & 1 & 1 & 0 & ? \\ ? & 0 & 0 & 0 & 0 & 0 & 0 & 1 & ? \end{bmatrix} \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} ? & 000 + 4500 & ? \\ ? & 5500 + 5000 & ? \\ ? & 6000 & ? \end{bmatrix} \\
 \mathbf{Z}'\mathbf{X} = \begin{bmatrix} ? & 0 & 0 & ? \\ ? & 0 & 0 & ? \\ ? & 0 & 0 & ? \\ ? & 0 & 0 & ? \\ ? & 0 & 0 & ? \\ ? & 0 & 0 & ? \\ ? & 1 & 0 & ? \\ ? & 0 & 1 & ? \\ ? & 0 & 1 & ? \end{bmatrix} \quad \mathbf{Z}'\mathbf{Z} = \begin{bmatrix} ? & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ? \\ ? & 1 & 0 & 0 & 0 & 0 & 0 & 0 & ? \\ ? & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ? \\ ? & 0 & 0 & 1 & 0 & 0 & 0 & 0 & ? \\ ? & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ? \\ ? & 0 & 0 & 0 & 0 & 1 & 0 & 0 & ? \\ ? & 0 & 0 & 0 & 0 & 0 & 1 & 0 & ? \\ ? & 0 & 0 & 0 & 0 & 0 & 0 & 1 & ? \\ ? & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ? \end{bmatrix} \quad \mathbf{Z}'\mathbf{y} = \begin{bmatrix} ? & 0 & ? \\ ? & 5000 & ? \\ ? & 0 & ? \\ ? & 4500 & ? \\ ? & 0 & ? \\ ? & 5500 & ? \\ ? & 5000 & ? \\ ? & 6000 & ? \end{bmatrix}
 \end{array}$$

$m_{ij}$  = number of records influenced by both level  $i$  and level  $j$

$h_i$  = sum of records influenced by level  $i$

# Mixed Model Equations: Construction

- Read a **data file**: animal n<sup>o</sup>, year effect, record

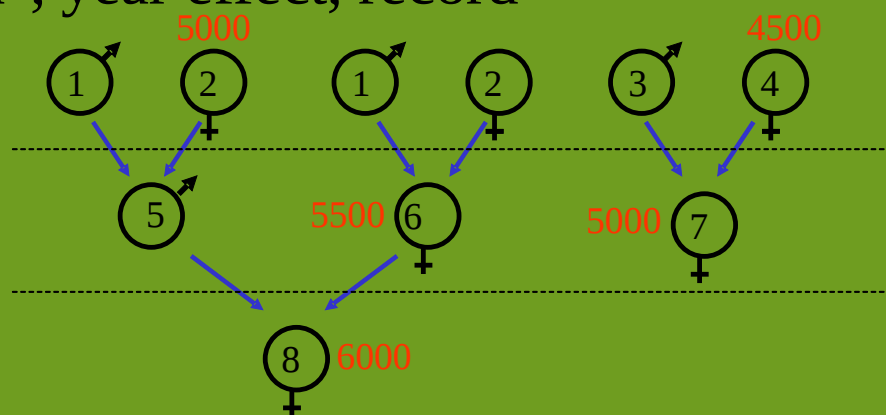
2 1 5000

4 1 4500

6 2 5500

7 2 5000

8 3 6000

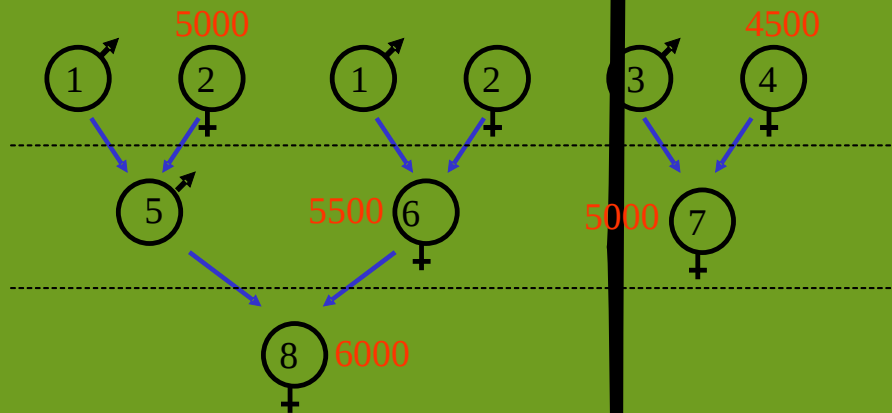


# A : the Numerator Relationship Matrix

2 simple rules:

$$g_{ii} = (1 + F_i) = 1 + \frac{1}{2}g_{sd}$$

$$g_{ij} = \frac{1}{2}(g_{is} + g_{id})$$



	1	2	3	4	1x2	1x2	3x4	5x6
1	1	0	0	0	0.5	0.5	0	0.5
2	0	1	0	0	0.5	0.5	0	0.5
3	0	0	1	0	0	0	0.5	0
4	0	0	0	1	0	0	0.5	0
5	0.5	0.5	0	0	1	0.5	0	0.75
6	0.5	0.5	0	0	0.5	1	0	0.75
7	0	0	0.5	0.5	0	0	1	0
8	0.5	0.5	0	0	0.75	0.75	0	1 + 0.25

# Henderson's rules to construct **A-1** (1976)



# Henderson's rules to construct A-1 (1976)





# Henderson's rules : application

- Read the **pedigree file**: animal n°, sire n°, dam n°

1 0 0	$\mathbf{A}^{-1} =$	$\begin{matrix} ? & 1 & 1 & 1 & 1 \\ ? & + & - & + & - \\ ? & 2 & 2 & 2 & 2 \\ ? & 1 & 1 & 1 & 1 \\ ? & - & + & - & 1 + - + - \\ ? & 2 & 2 & 2 & 2 \end{matrix}$	0	0	-1	-1	0	0	?	
2 0 0		0	0	-1	-1	0	0	?		
3 0 0		$1 + \frac{1}{2}$	$\frac{1}{2}$	0	0	-1	0	?		
4 0 0		$\frac{1}{2}$	$1 + \frac{1}{2}$	0	0	-1	0	?		
5 1 2		$\frac{1}{2}$	$\frac{1}{2}$	$2 + \frac{1}{2}$	$\frac{1}{2}$	0	-1	?		
6 1 2		-1	-1	0	0	$\frac{1}{2}$	$2 + \frac{1}{2}$	0	-1	?
7 3 4		-1	-1	0	0	$\frac{1}{2}$	$2 + \frac{1}{2}$	0	-1	?
8 5 6		0	0	-1	-1	0	0	2	0	?
		0	0	0	0	-1	-1	0	2	?

# Mixed Model Equations: construction

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \alpha A^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ a \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0+2\alpha & 0+\alpha & 0 & 0-\alpha & 0-\alpha & 0 \\ 1 & 0 & 0 & 0+\alpha & 1+2\alpha & 0 & 0-\alpha & 0-\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0+1.5\alpha & 0+0.5\alpha & 0 & 0-\alpha \\ 1 & 0 & 0 & 0 & 0 & 0+0.5\alpha & 1+1.5\alpha & 0 & 0-\alpha \\ 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0+2.5\alpha & 0+0.5\alpha & 0 & 0-\alpha \\ 0 & 1 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0+0.5\alpha & 1+2.5\alpha & 0 & 0-\alpha \\ 0 & 1 & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 1+2\alpha & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 1+2\alpha \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \hline a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{pmatrix} = \begin{pmatrix} 5000+4500 \\ 5500+5000 \\ 6000 \\ \hline 0 \\ 5000 \\ 0 \\ 4500 \\ 0 \\ 5500 \\ 5000 \\ 6000 \end{pmatrix}$$

# Mixed Model Equations: solutions

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{pmatrix} = \begin{pmatrix} 4750 \\ 5250 \\ 5930 \\ 28 \\ 83 \\ -28 \\ -83 \\ 56 \\ 83 \\ -83 \\ 70 \end{pmatrix}$$

Sum for  
« base animals »  
=0

Mean = 18.6

Mean = 70

genetic trend  
/ generation

$$y_8 = 6000 = 5930 + 70 \quad \text{connection through pedigree}$$



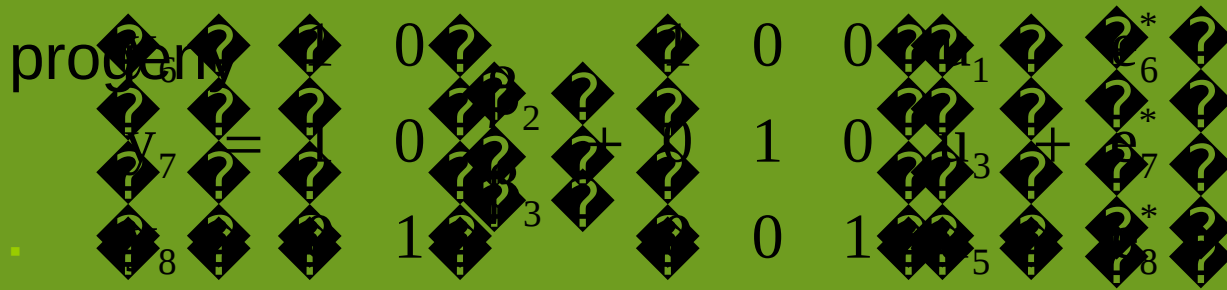
When the **base population** (= population with unknown parents) is **not related, not inbred and not selected** and when **all the data used for selection and planned matings** are **included** in the BLUP evaluation, the use in the mixed model equations of the complete relationship matrix **A** and the additive genetic variance  $\sigma_a^2$  of the base population **accounts for changes in expected value and variance of additive genetic values related to selection, genetic drift, inbreeding and planned matings** (Sorensen et Kennedy, 1982).



- animal model
- fixed and random effects are estimated together
- all pedigree information is used
- all data on which selection was based are used  
(including on correlated traits) □ multi-trait evaluation
- needs  $\sigma_a^2$  in the base (not selected, not related, not inbred)  
population
- ➔ **Multi-trait BLUP animal model** (t traits x (na animals + nf fixed effects))
- ➔ **All other models are simplified versions, with extra hypotheses**



- $y_i = \beta_j + a_i + e_i$      $y_i = \beta_j + u_s + e_i^*$
- records from dams 2 and 4 can't be used (unknown sire)
- assume each dam unselected, unrelated with at most one

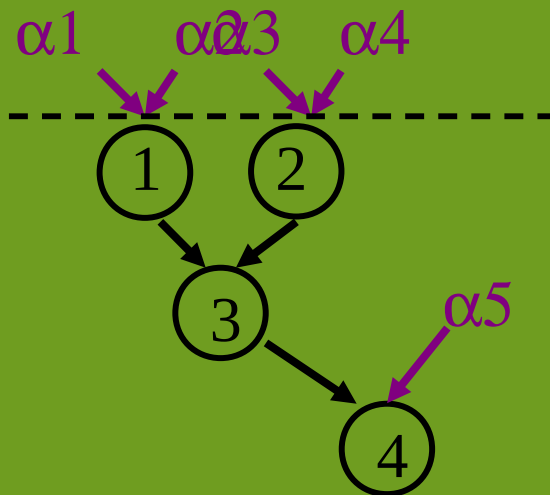


$$\mathbf{G} = \text{var}(\mathbf{u}) = \frac{1}{4} \sigma_a^2 \mathbf{A} \quad \text{with} \quad \mathbf{A}^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & -\frac{2}{3} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{4}{3} \\ \frac{2}{3} & 0 & -\frac{1}{3} \end{bmatrix} \quad \text{and} \quad \alpha = \frac{\sigma_{e^*}^2}{\sigma_u^2} = \frac{\sigma_e^2 + \frac{3}{4} \sigma_a^2}{\frac{1}{4} \sigma_a^2}$$



# BLUP : Groups of unknown parents

## Example:



## Definition of unknown parent groups

by sex of progeny and year of birth:

- $\alpha_1, \alpha_2$  = parents of males of generation 1
- $\alpha_3, \alpha_4$  = parents of females of generation 1
- $\alpha_5$  = parent of females of generation >1

$$Q_g = \begin{matrix} \begin{matrix} \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \end{matrix} & \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.5 & 0.5 & 0 \\ 0.25 & 0.25 & 0.5 \end{matrix} & \begin{matrix} \begin{matrix} \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \end{matrix} \\ \begin{matrix} \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \end{matrix} \\ \begin{matrix} \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \end{matrix} \\ \begin{matrix} \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \\ \diamond ? \end{matrix} \end{matrix}$$



# Groups of unknown parents

$$\square \quad y = X\beta + Z a + e$$

$$= X\beta + Z (Qg + a^*) + e$$

with  $a^* = a - Qg$

$$= \begin{bmatrix} X & ZQ \end{bmatrix} \begin{bmatrix} \beta \\ a^* \end{bmatrix} + e$$

$g$  is a particular fixed effect

➤ Mixed Model Equations (after some transformations):

$$\begin{bmatrix} X'X & 0 & X'Z \\ 0 & Q'A^{-1}Q & -\alpha Q'A^{-1} \\ Z'X & -\alpha A^{-1}Q & Z'Z + \alpha A^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ g \\ Qg + a^* \end{bmatrix} = \begin{bmatrix} X'y \\ 0 \\ Z'y \end{bmatrix}$$

=  $a$

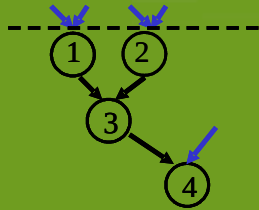
Interest: MME more sparse, easier to construct.  $A$  obtained directly

# Quaas' rules to construct

$$\begin{array}{ccc} \alpha Q' A^{-1} Q & -\alpha Q' A^{-1} & \\ \alpha A^{-1} Q & A^{-1} & \end{array}$$

- read the pedigree file: i , s , d  
(animal, sire or group of unknown sire, dam or group of unknown dam)
- Determine  $d_i^{-1}$  function of number of known parents
  - add  $d_i^{-1}$  to  $A(i,i)$
  - add  $-0.5 d_i^{-1}$  to  $A(i,s)$ ,  $A(s,i)$ ,  $A(i,d)$ ,  $A(d,i)$
  - add  $0.25 d_i^{-1}$  to  $A(s,s)$ ,  $A(s,d)$ ,  $A(d,s)$ ,  $A(d,d)$

# Example



$$Qg = \begin{bmatrix} ? & 1 & 0 & 0 & ? & 1 & g_1 & g_1 \\ ? & 0 & 1 & 0 & ? & 1 & g_2 & g_2 \\ ? & 0 & 0 & 0 & ? & 0 & 2 & g_2 \\ ? & 0.5 & 0.5 & 0 & ? & 1 & 3 & 3 \\ ? & 0.25 & 0.25 & 0.5 & ? & 4 & 3 & g_3 \end{bmatrix}$$

$$d_i^{-1} = 1$$

$$4/3$$

$$\begin{bmatrix} ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \end{bmatrix} \begin{bmatrix} ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \end{bmatrix} \begin{bmatrix} ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \\ ? & ? & ? & ? & ? & ? & ? & ? \end{bmatrix}$$

pedigree  
file

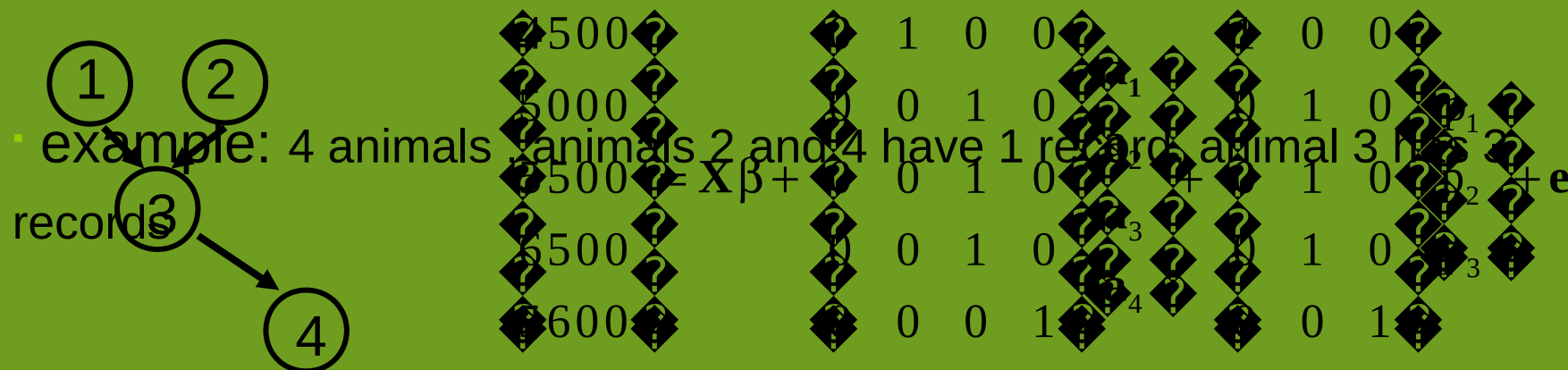
- When we have repeated records on a same animal, all these records

share a same « permanent environment »

= same non (additive) genetic characteristics of the animal, which

are affecting each record in the same way (e.g., consequences of a poor growth rate, or a disease, a lost quarter, etc...)

$$y = Xb + Za + Wp + e$$





- $$\text{Var} \begin{pmatrix} \beta \\ a \\ e \end{pmatrix} = \begin{pmatrix} \sigma_a^2 & 0 & 0 \\ 0 & \text{I}\sigma_w^2 & 0 \\ 0 & 0 & \text{I}\sigma_e^2 \end{pmatrix}$$

- Mixed Model Equations:

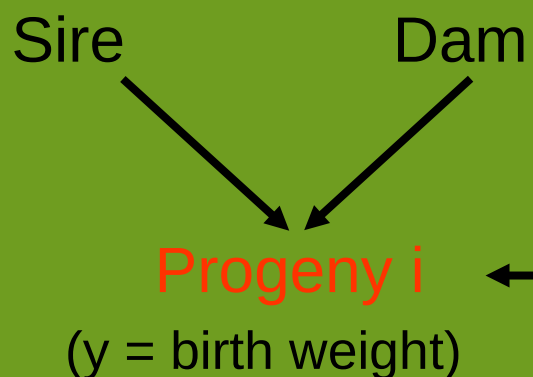
$$\begin{pmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + \alpha_1 A^{-1} & Z'W \\ W'X & W'Z & W'W + \alpha_2 I \end{pmatrix} \begin{pmatrix} \beta \\ a \\ e \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \\ W'y \end{pmatrix}$$

with  $\alpha_1 = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-r}{h^2}$   
 and  $\alpha_2 = \frac{\sigma_e^2}{\sigma_w^2} = \frac{1-r}{r-h^2}$

# Model

- Maternal effect = all (genetic or not genetic) characteristics of the dam which influence the observation of a progeny

- Example: birth weight:



Format = characteristics of the dam  
 Gestation length with immediate effect on  
 Uterine capacity birth weight  
 → Act as an environmental effect on progeny

- direct effect □ animal model (y = record of the progeny)
- maternal effect □ repeatability model (y = record of the dam)



- $y$  = fixed effect affecting the progeny or the dam  
 + direct effect (additive genetic value of the progeny)  
 + maternal effect (additive genetic value + perm. environ. effect of the dam)

➤  $y = X\beta + Ua + Zm + Wp + e$

- Adding columns of 0 to **U** and **Z**, **a** and **m** can be made of same length = they represent the same (all) animals. Then for any pair of related animals  $i, j$  with relationship coefficient  $g_{ij}$  :

$$\begin{aligned} \text{cov}(a_i, a_j) &= g_{ij} \sigma_a^2 \\ \text{cov}(m_i, m_j) &= g_{ij} \sigma_m^2 \\ \text{cov}(a_i, m_j) &= g_{ij} \sigma_{am} \end{aligned} \quad \text{var} \begin{bmatrix} a \\ m \\ p \end{bmatrix} = \begin{bmatrix} \sigma_a^2 A & \sigma_{am} A & 0 \\ \sigma_{ma} A & \sigma_m^2 A & 0 \\ 0 & 0 & \sigma_p^2 I \end{bmatrix}$$



➤ Start with  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{m} + \mathbf{W}\mathbf{p} + \mathbf{e}$

i.e.  $y_i = \mathbf{x}_i' \boldsymbol{\beta} + a_i + m_{\text{dam}} + p_{\text{dam}} + e_i$

➤ Using  $a_i = \frac{1}{2}a_{\text{sire}} + \frac{1}{2}a_{\text{dam}} + \phi_i$

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \frac{1}{2}a_{\text{sire}} + \frac{1}{2}a_{\text{dam}} + m_{\text{dam}} + p_{\text{dam}} + (\phi_i + e_i)$$

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \frac{1}{2}a_{\text{sire}} + \left[ \begin{array}{l} + \frac{1}{4}a_{\text{mgs}} \\ + \frac{1}{2}m_{\text{mgs}} \end{array} \right] + \left[ \begin{array}{l} + \frac{1}{4}a_{\text{mgd}} + \frac{1}{2}\phi_{\text{dam}} \\ + \frac{1}{2}m_{\text{mgd}} + \phi_{\text{dam}}^{\text{m}} \\ + p_{\text{dam}} \end{array} \right] + (\phi_i + e_i)$$

(should have 1 daughter per maternal grand dam)

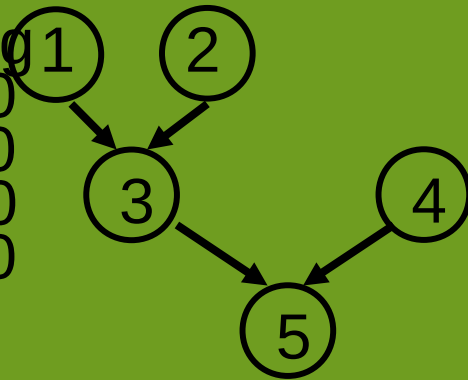
$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_{\text{sire}} + \mathbf{v}_{\text{mgs}} + \mathbf{t}_{\text{dam}} + e_i^*$$





➤ Animal sex contemporary Backfat ADG  
group thickness

1	1	1	22 mm	650g
2	2	1	15	570
3	1	2	19	590
4	2	2	24	700
5	2	2	21	620



➤ model for animal i, trait j :

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{a} + e_{ij}$$







- The hypotheses required to get the « ideal » properties of Animal model BLUP are better fulfilled
  - **Correction of bias related to selection on correlated traits**
- **Better accuracy** □ stronger genetic gain  
(except when genetic and residual correlations between two traits are the same)
- Better data structure (less disconnexion)
- Optimal weights for a multiple trait selection objective = **economic weights** (no need to worry about combining univariate (but correlated) predicted breeding values)



- Estimation of genetic parameters (covariances): difficult, **costly** and needs a lot of data to be accurate
- parameters can be wrong □ sometimes worse than univariate (**but** in univariate analyses, we assume covariances = 0)
- Computing work for  $t$  traits :  **$t^2$  or  $t^3$**  x computing work for 1 trait



- Discrete data: several types
  - Binary (0/1, dead/alive, sick/healthy)
    - Generalized linear models (Threshold models, logistic)
  - Ordered categorical (1/2/3; no help/easy calving/difficult calving)
    - Generalized linear models (Threshold models)
- ... many others



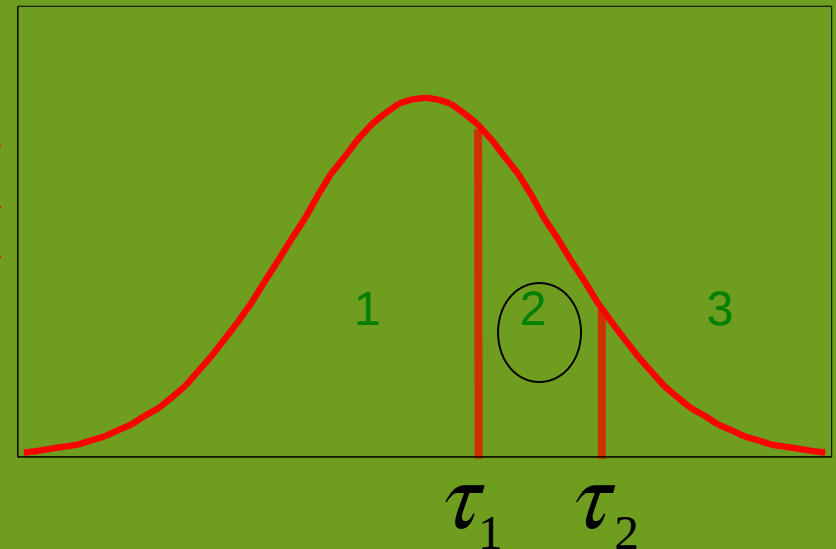
- Calving ease: no help = 1 / easy = 2 / difficult = 3

Assume a non observable underlying variable for animal  $i$

$$\Pi_{ij} = \text{Prob}(\tau_{j-1} < y_i < \tau_j)$$

$$\Pi_{ij} = \Phi\left(\frac{\tau_j - \eta_i}{\sigma_i}\right) - \Phi\left(\frac{\tau_{j-1} - \eta_i}{\sigma_i}\right)$$

$$\eta_i = \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a}$$



- « link function » = F = probit = integral of the normal density



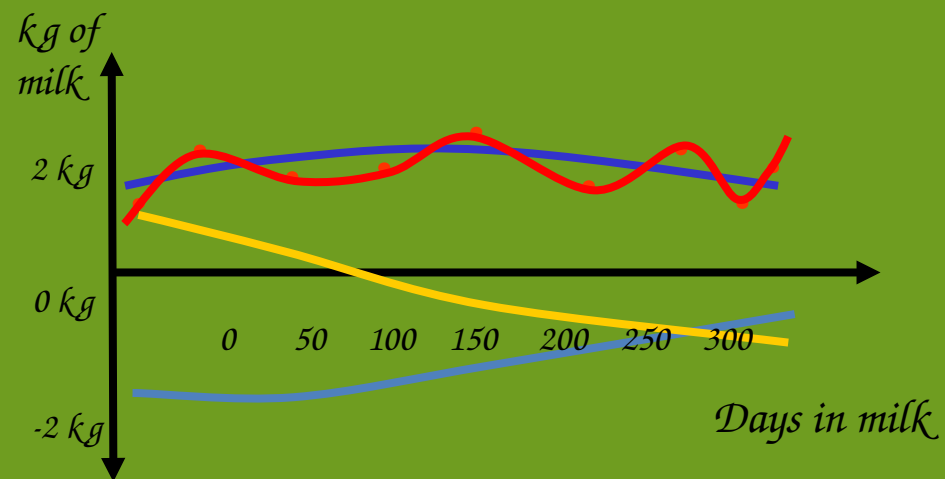
- Example: growth, test day milk production...
- Random regression models + BLUP

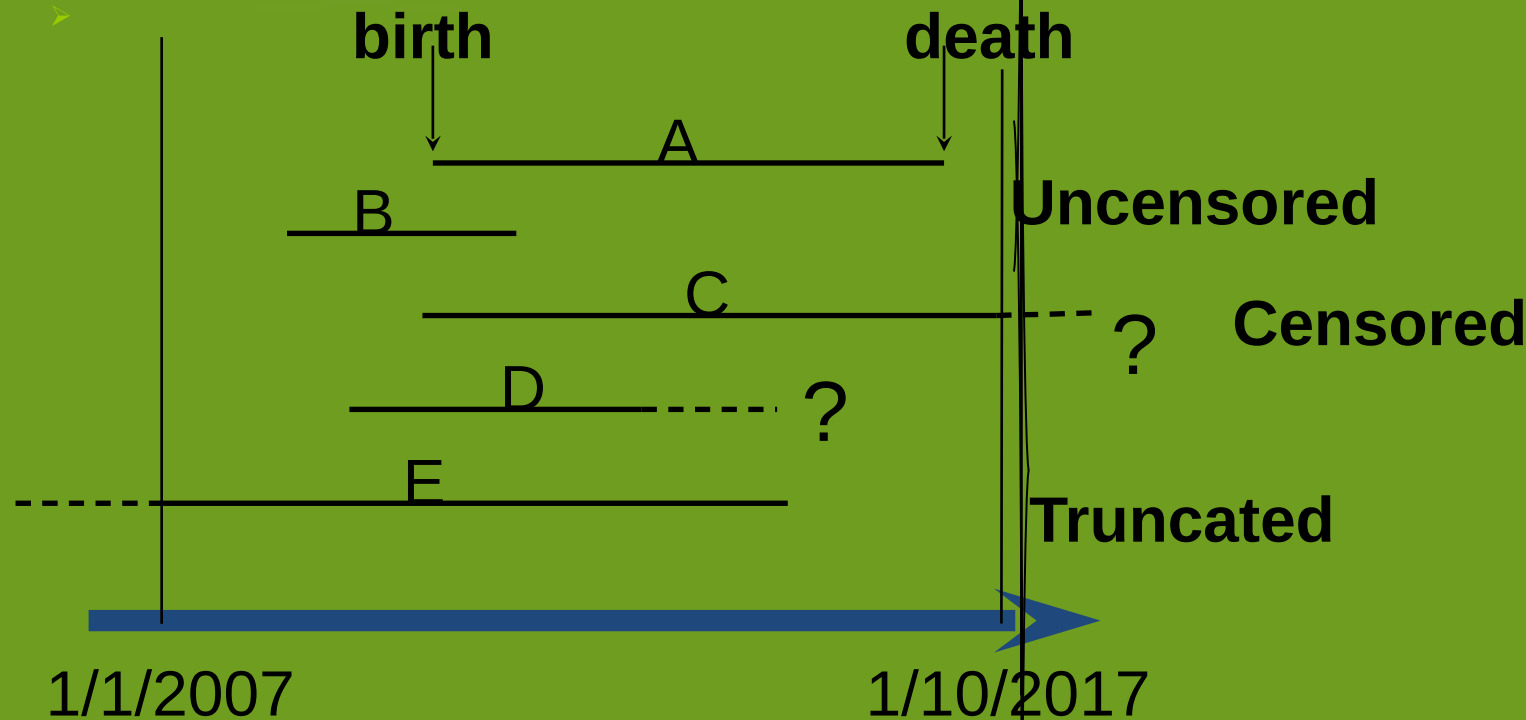
Consider one genetic value « per day » and describe it as a function of time  $t$

$$a(t) = \lambda_0 + \lambda_1 t + \lambda_2 t^2 + \lambda_3 t^3 + \lambda_4 t^4$$

$$a(t) = \alpha_1(t) \textcircled{a_1} + \alpha_2(t) \textcircled{a_2}$$

Mean production      Persistency





- Bayesian approach prior x likelihood
- In the likelihood construction, special treatment of censored records