Basics of reliability (and EBV) calculation: how to make blocks

MiX99 course on genomic prediction

COURSE DAY 1, March 9th, 2026

Contents

© Luke

- Introduction to reliability calculation
- Approximation in ApaX
- Restrictions and modifications to CLIM
- Test-day model example
- Discussion

© Luke

Introduction

- In this lecture the main focus is on defining within block effects for reliability calculation
- The lecture tomorrow morning covers the details on how to calculate reliabilities

Introduction to approximation of reliability

© Luke

- Accuracy (r) is the correlation between true and estimated breeding value (EBV)
- In dairy cattle evaluation, it is usually expressed as reliability: r^2
- The higher r^2, the higher the likelihood that the EBV is close to the true breeding value and the less the EBV will change in subsequent evaluations
- Calculation of r^2 requires the EBV diagonal elements of the inverse of the coefficient matrix (**C**) of the MME, which are for the random effects the prediction error variances (PEV):

$$r_i^2 = 1 - \left(\mathrm{PEV}_i / (\mathbf{G}_{ii} * \sigma_a^2\right)$$

where $\mathbf{G}_{ii}$ is the diagonal of relationship matrix for individual i, and σ_a^2 is the genetic variance.

- However, for large MME it is not possible to invert **C**

© Luke

Reliability for animal model y =Xb+Za + e

Assume $\mathbf{a} \sim (\mathbf{0}, \mathbf{A}\sigma_a^2)$ and $\mathbf{e} \sim (\mathbf{0}, \mathbf{R})$, then mixed model equations are

$$\begin{bmatrix} \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \mathbf{A}^{-1}\sigma_a^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix}$$

or shortly **Cs** = **r**. Let

$$\mathbf{C^{-1}} = \begin{bmatrix} \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \mathbf{A}^{-1}\sigma_a^{-2} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C^{bb}} & \mathbf{C^{ba}} \\ \mathbf{C^{ab}} & \mathbf{C^{aa}} \end{bmatrix}$$

Reliability of individual i is

$$r_i^2 = 1 - \frac{\mathrm{Var}(\hat{\mathbf{a}}_i - \mathbf{a}_i)}{\mathrm{Var}(\mathbf{a}_i)} = 1 - \frac{\{\mathbf{C^{aa}}\}_{ii}}{\mathbf{g}_{ii}}$$

where $\{\mathbf{C^{aa}}\}_{ii}$ is prediction error variance (PEV) and $\mathbf{g}_{ii}$ is genetic variance. For animal model, $\mathbf{g}_{ii} = (1 + F_i)\sigma_a^2$

where F_i is inbreeding coefficient of individual i.

In practice, we approximate!

Approximation of reliabilities by ApaX99

- Needs elements from the inverse of the coefficient matrix (**C**) of the MME
- $\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \mathbf{A}^{-1}\sigma_a^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix} = \mathbf{Cc} = \mathbf{r}$
- $\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \mathbf{A}^{-1}\sigma_a^{-2} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}^{bb} & \mathbf{C}^{ba} \\ \mathbf{C}^{ab} & \mathbf{C}^{aa} \end{bmatrix}$
- ApaX99 splits the approximation into two steps:
  - 1st step: approximate the amount of information due to records
  - 2nd step: approximate the amount of information due to pedigree information
- There are different options for the 2nd step
- NOTE: phenotype values do not affect to reliability!

© Luke

1st step: approximate the amount of information due to non-genetic effects

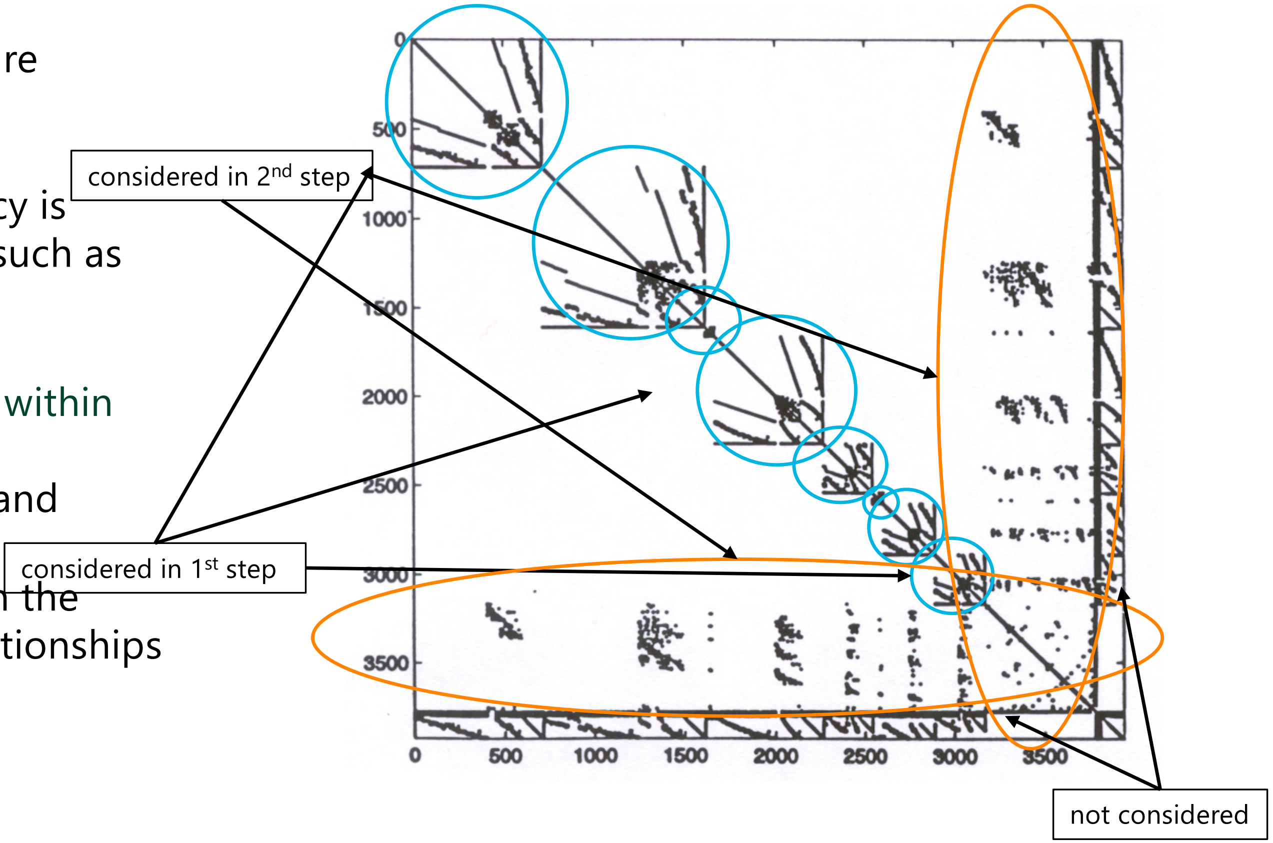
- In approximation **some of the effects** are absorbed to genetic effects

- Effects having lowest estimation accuracy is accounted: contemporary group effect such as herd-year

- contemporary group effect should be a within block effect to save memory
- → see WITHINBLOCKORDER command

- All within block effects are considered in the approximation, except those due to relationships

- Structure of **C** when MME is in block order



considered in 2nd step

considered in 1st step

not considered

© Luke

2nd step: approximate the amount of information due to pedigree

- Different approaches possible:
- Interbull EDC and reliabilities (for sires) ← reliabilities needed for Interbull MACE
- Misztal & Wiggans approach ← reliabilities for all animals
- Jamrozik & Schaeffer approach ← reliabilities for all animals
- Tier & Meyer approach ← reliabilities for all animals, accounts for multiple traits

- approach for the single-step GBLUP by Misztal approach ← uses the genomic relationship matrix (there are better ones)

© Luke

Restrictions and modifications for setting up the MME

- A CLIM file has to be prepared for setting up the MME by *mix99i*
- However, the CLIM file may differ from that one used for the breeding value estimation

- ***Necessary considerations***
- Block and relationship code needs to be defined → DATASORT
- Block code in pedigree as fourth column!
- Only effects defined within blocks are considered → WITHINBLOCKORDER
- Animal genetic effect has to be the first effect within blocks
- Only one effect absorbed approximately (more on this in the next lecture)
- Inbreeding is ignored
- Genetic groups are ignored

Blocks and parallel apax99 (apax99p)

- Parallel (MPI) solver, **mix99p**, and **apax99p** distributes data and equations to multiple processes
- Distribution of equations to processes does not affect to **EBV**s
- However, apax99p absorbs non-genetic effects within process, and it does not absorb equations from different processes at the same time
    - This can have effect to reliabilities!
- Good ordering and renumbering of blocks with `partition_blocks` is especially important in these situations
- If animals change their blocks, specify the block they have the most of their data to pedigree block

© Luke

Test-day model example

© Luke

Test-day model example

- Lets consider test-day model defined in CLIM for EBV (only relevant statements shown):

```
DATASORT BLOCK=HERD PEDIGREECODE=ANI
MODEL
MILK(1)=HET(het|HCL) REC(rec|HCL) YM SEASON(t2 t3 t4 t5 t6 |SCLASS)  DCC &
        HY  HSLP(t2 | HERD)  HTD  PE(t9:25|ANI)@1  GENETIC(t200:215|ANI)@1

RANDOM HTD PE GENETIC
WITHINBLOCKORDER GENETIC PE HTD HSLP HY
PRECON              b     b   b  b    b  b
```

- Effects in the top row are common to all animals
- **Effects** in the bottom row are contemporary group effects
- Cgroup effects are specified as within block effects in `WITHINBLOCKORDER` statement
- Block code in pedigree as fourth column!

Test-day model modified for reliability calculation

© Luke

- Model definition for reliability calculation does not need across block effecs, inbreeding, or preconditioner:

```
DATASORT BLOCK=HERD PEDIGREECODE=ANI
MODEL
MILK(1)= HY  HSLP(t2 | HERD)  HTD  PE(t9:25|ANI)@1  GENETIC(t200:215|ANI)@1

RANDOM GENETIC PE HTD
WITHINBLOCKORDER GENETIC PE HTD HSLP HY
```

- This speeds up running mix99i

© Luke

Conclusion

- ApaX splits the approximation into two steps:
  - 1st step: approximate the amount of information due to records
  - 2nd step: approximate the amount of information due to pedigree information
- Defining within block effects is important
- CLIM file for R2 calculation can be simplified
- Inbreeding is ignored
- Genetic groups and MFs are ignored

© Luke

Time for discussion

luke.fi

© Luonnonvarakeskus
© Natural Resources Institute Finland
© Naturresursinstitutet