Approximating reliabilities

Timo Pitkänen

MiX99 course: test-day models and single step genomic prediction

COURSE DAY 1, April 10th, 2025

Contents

Calculation of reliabilities for BLUP test-day model (step 1 of single step pipeline)

- Set up model
- Define breeding values of which reliabilitis are calculated
- Apax instruction file
- Solutions

Calculation of single step reliabilities

- 3 steps explained

BLUP reliabilities for 9 trait test-day model

Model and definition of breeding value of interest

```
TITLE   MT random regression TDM

DATAFILE  lact3.dat
INTEGER   HERD ANI LACT HTD HC2Y C2Y AGE DCC DDRY DIM ResClass
REAL      MILK PROTEIN FAT
MISSING   -64.0

PEDFILE  TD9.ped
PEDIGREE G  am
TABLEFILE LG3_W004.cov
TABLEINDEX DIM
RESIDUAL ResClass
RESIDFILE lact3.res
PARFILE   lact3.par

RANDOM   HTD  PE G
TMPDIR  tmpMiX

TRAITGROUP LACTMODEL
MILK(1)    =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
PROTEIN(1) =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
FAT(1)     =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
MILK(2)    =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
PROTEIN(2) =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
FAT(2)     =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
MILK(3)    =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
PROTEIN(3) =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
FAT(3)     =  LACcurve(t1 t2 t3 t4) HTD PE(1 t2 t3 t5|ANI)  G(1 t2 t3 t5|ANI)
DATASORT  BLOCK=HERD PEDIGREECODE=ANI
WITHINBLOCKORDER G PE HTD
PRECON           b  b  d    d
```

- Model has milk, protein and fat for 3 lactations.
- The aim is to estimate R2 for 305d yield
- For R2 calculation, 305d yield is defined as sum of breeding values for test-days :
  15 45 75 105 135 165 195 225 255 285
- Choosing test-days for every 30 days assumes the observations are done once a month

Theory 1 trait

- Breeding value for DIM $\boldsymbol{d}$, trait $\boldsymbol{t}$, animal $\boldsymbol{a}$ modelled as:

$$BV_{atd} = \sum_{i=1}^{4} \boldsymbol{c_i(d)} \boldsymbol{u_{ai}}$$

$$= \underbrace{[\boldsymbol{c_1(d)}\ \boldsymbol{c_2(d)}\ \boldsymbol{c_3(d)}\ \boldsymbol{c_4(d)}]}_{\text{Legendre + Wilmink covariates for DIM d}} \underbrace{\begin{bmatrix} \boldsymbol{u_{a1}} \\ \boldsymbol{u_{a2}} \\ \boldsymbol{u_{a3}} \\ \boldsymbol{u_{a4}} \end{bmatrix}}_{\text{Random genetic regression coefficients for animal a}} = \boldsymbol{c(d)'u_a}$$

- 305d breeding value is

$$\mathrm{BV}_{305\mathrm{a}} = \sum_{d \in dims} \boldsymbol{c(d)'u_a} = \boldsymbol{c'_{305}u_a},$$

$$\boldsymbol{c_{305}} = \sum_{\boldsymbol{d \in dims}} \begin{bmatrix} \boldsymbol{c_1(d)} \\ \boldsymbol{c_2(d)} \\ \boldsymbol{c_3(d)} \\ \boldsymbol{c_4(d)} \end{bmatrix}$$

- Variance of 305d BV

$$\mathrm{var}(\mathrm{BV}_{305\mathrm{a}}) = c'_{305} G_t\, c_{305}$$

$$G_t = \mathrm{var}(\boldsymbol{u_a})$$

Theory 9 traits

- Assume $\boldsymbol{u_a}$ contains all 36 random regression coefficients for genetic effect ordered in trait order
- Then 305d BVs for 9 traits can be calculated as

$$\boldsymbol{BV_9} = \underbrace{\begin{bmatrix} \boldsymbol{c'_{305}} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{c'_{305}} \end{bmatrix}_{\boldsymbol{35\times 9}}}\quad \boldsymbol{u_a} = \boldsymbol{C_9 u_a}$$

Block diagonal matrix, row *n* picks coefficients for *n*th trait

- Variance of $\boldsymbol{BV_9}$

$$\boldsymbol{G_9} = \boldsymbol{var(BV_9)} = \begin{bmatrix} \boldsymbol{c'_{305}} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{c'_{305}} \end{bmatrix} \boldsymbol{G} \begin{bmatrix} \boldsymbol{c'_{305}} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{c'_{305}} \end{bmatrix}'$$

Theory, reliability calculation

- First, prediction error covariance (**PEC**) is calculated for $\boldsymbol{u_a}$ **based on test-day model**
- PEC for $\boldsymbol{C_9 u_a}$ is obtained as

$$\boldsymbol{pec}(\boldsymbol{C_9 u_a}) = \boldsymbol{C_9 pec}(\boldsymbol{u_a})\boldsymbol{C_9'}$$

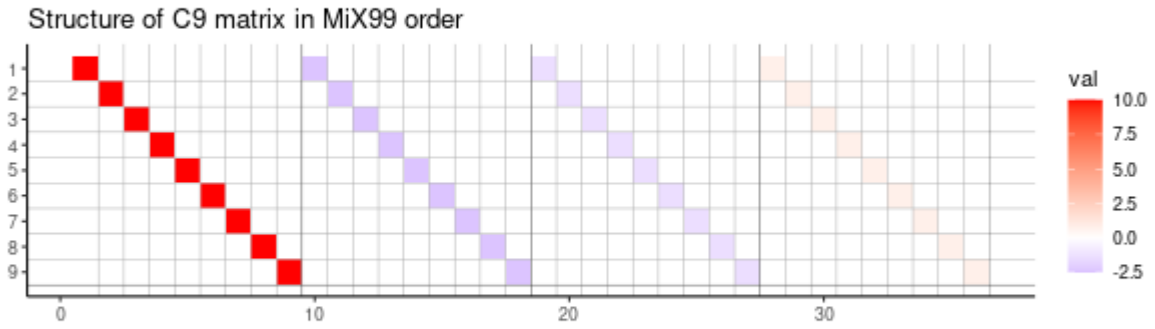
- R2 is calculated as (elementwise)

$$R_9^2 = 1 - diag(\boldsymbol{pec}(\boldsymbol{C_9 u_a}))./diag(\boldsymbol{G_9})$$

Luke

Calculation example for genetic covariance matrix of 9 trait 305d breeding values in R

```
> # Selected dims
> DIMS <- seq(15, by = 30, length.out = 10)
> DIMS
 [1]  15  45  75 105 135 165 195 225 255 285
> #Sum over selected dims, drop additional columns
> c305 <- colSums(subset(covtable[, -5], DIM %in%  DIMS))[-1]
> c305
      t1       t2       t3       t5
10.00000 -2.50440 -1.38713  0.78536
> # create C9 matrix (this assumes MiX99 ordering of effects)
> C9 <- t(c305) %x% diag(9)
> # calculate genetic covariance matrix
> G9   <- C9 %*% G %*% t(C9)
> #name rows and cols and show the matrix
> colnames(G9) <- rownames(G9) <- cnames
> round(G9, 2)
       M1     P1     F1     M2     P2     F2     M3     P3     F3
M1 247.33 156.44 110.53 184.50 111.20  74.26 172.96  96.39  66.23
P1 156.44 124.23  85.72 116.45  92.12  61.40 108.85  80.30  54.54
F1 110.53  85.72  87.73  78.57  63.08  65.92  77.27  57.24  62.24
M2 184.50 116.45  78.57 221.62 140.58  99.78 222.86 135.15 100.88
P2 111.20  92.12  63.08 140.58 110.48  79.54 143.11 106.09  79.68
F2  74.26  61.40  65.92  99.78  79.54  79.05 106.54  79.72  80.78
M3 172.96 108.85  77.27 222.86 143.11 106.54 245.00 151.14 118.56
P3  96.39  80.30  57.24 135.15 106.09  79.72 151.14 111.29  87.36
F3  66.23  54.54  62.24 100.88  79.68  80.78 118.56  87.36  89.43
```



Reliability calculation in MiX99

NOTE

- Residual classes are not used in reliability approximation.
- Instead, it uses residual covariance matrix given in parfile.
- One can use weighted mean of matrices for classes or just one

Calculation has two steps

1. Run preprocessor
   - Requires working clim/mix file
2. Run apax99 (apax99 < options.apx)
   - Requires option file that specifies what to calculate

```
# AccurType
  4
# MaxNonZ
  303750
# StartDIM
  8
# CovarInfo
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
# OrginalDIR
 MiX99_DIR.DIR
# NumBVs
  9
# Weights
1 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 0 1 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 0 1 0 0
0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 1
# H2calc
  1 2
# Jfilter
  3
```

Apax99 instruction file explained, Type of analysis


```
# AccurType
  4
```

Give type of analysis:

0 - Exact reliabilities
1 - Interbull reliabilities
2 - Misztal & Wiggans type reliabilities
3 - Jamrozik et al. type reliabilities
4 - Tier & Meyer type reliabilities
5 - Tier & Meyer improved (beta testing)
6 - Method 4 with PEV information read from file
7 - Univariate collapsed with exact relationships
10 - Interbull EDCs/reliabilities with genotype discount (sire model)
11 - Interbull EDCs/reliabilities with genotype discount (animal model)
20 - single-trait ERC from reversed reliability approximation
40 - multi-trait ERC from reversed reliability approximation

Options: i or o = input or output to file:PEV.bin
m = maternal trait in genetics (in addition to individual effect).
r = calculate maternal trait reliabilities.
l = long output listing.
n = no checking of data release
c = no checking of pedigree loops/cycles
Gn = genomic reliabilities for ssGBLUP:
G0 = no genomic information used
G1 = I.Misztal method Approx1
G2 = I.Misztal method Approx2
p = pedigree file name given for methods 20 and 40
Options i, o, r are not available for methods 0, 4, 5

Not considered to be good methods

>> # AccurType
>> 4

4 - Tier & Meyer type reliabilities

Apax99 instruction file explained, MaxNonZ, Startdim, Covariate info


```
# MaxNonZ
  303750
# StartDIM
  8


# CovarInfo
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
  15 10 30
```

```
Give maximum number of non-zeros in the sparse matrix. The largest herd will determine this value.
 Most often values around 500,000 will suffice.
   >> # MaxNonZ
   >> 303750
```

First DIM in the covariable file

For each trait, the 305d breeding value consists of sum of selected 10 DIMs starting from 15, by 30

Apax99 instruction file explained, MaxNonZ, Startdim, Covariate info


```
# OrginalDIR
 MiX99_DIR.DIR
# NumBVs
9
# Weights
1 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 0 1 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 0 1 0 0
0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 1
```

- Directive file, mix99i creates MiX99_DIR.DIR automatically. It is usually ok.
- For how many breeding values R2 is calculated. This defines how many weights rows will be given next.
- Trait weights for which R2 are calculated. Each column is trait in the model
- In this example R2s is calculated for 305d yields separately for each trait
- For combined indices one might have weights like this (average over lact)

```
# NumBVs
3  # ONLY 3 R2s calculated
# Weights
1 0 0 1 0 0 1 0 0
0 1 0 0 1 0 0 1 0
0 0 1 0 0 1 0 0 1
```

Apax99 instruction file explained,H2 calculation

```
# H2calc
  1 2
```

```
 There are           4  random effects in the model.
 The two last ones are animal genetic and residual.
 Give numbers of the random effects accounted in the
 heritability calculations in addition to the 2 last ones:
 Give 0 if no additional random effects are accounted.
 For example: 2 3 4 would include random effects 2 3 and 4
              in the calculation of heritability.

   >> # H2calc
   >> 1 2
1 2
```

- **H2calc** defines which random effects are accounted in heritability calculation for phenotypic variance
- Genetic and residual effects are automatically included
- In our example model HTD (random effect #1) and PE (random effect #2) are also included in H2 calculation

Apax99 instruction file explained, Jfilter

```
# Random effect number in model    1     2  3   (residual =4)
RANDOM                                  HTD  PE G

#  EFFECT#  IN JFILTER  1     2   3
WITHINBLOCKORDER   G    PE HTD
```

```
Effects that have been defined within the block
are accounted approximately in the calculations.

Number of within block effects:           3

Effect#    Description
          1     Animal genetic effect
          2     Random effect number           2
          3     Random effect number           1

Give effect# ABOVE WHICH EXACT absorption is used.
For example: giving 3 means that only effect 2 is exactly absorbed.
Number should be:           2  to            4
  >> # Jfilter
  >> 3

Approx. absorption of effect            3
Exact absorption of effect 2 to the first effect
```

PE

HTD

JFilter Calculation of reliabilities is programmed such that all within block effects are absorbed to the animal genetic effects. They can be absorbed either exactly or approximately. JFilter describes which effects are approximately absorbed. If the exact absorption is done, the sparse matrix may be filled and exhaust available memory. In addition, the computations may be slow. Therefore, some effects can be only approximately absorbed with no additional fill-in to the sparse matrix. In order to minimize memory use in absorption, ordering of the effects within block should be such that the smallest number is given to an effect with observations from a single animal and larger numbers to the effects with observations from several animals. Effects are ordered as given in the within block ordering. Exact absorption is commonly done for the effects that are within animals such as permanent environmental effects. For example, let's consider a case in which we have three within block effects: 1) genetic, 2) permanent environment, and 3) herd year. If we want to absorb the herd effect approximately, the JFilter value of 3 is given. Thus, only the permanent environment effect (effect 2) is exactly absorbed to the genetic effects. In a situation, in which we want to exactly absorb all other within block effects to the genetic effects (n within block effects), a value of n+1 should be given to the JFilter. In the example given above, JFilter would be 4.

```
# Jfilter
  3
```

```
Effects that have been defined within the block
are accounted approximately in the calculations.

Number of within block effects:          3

Effect#    Description
         1     Animal genetic effect
         2     Random effect number          2
         3     Random effect number          1

Give effect# ABOVE WHICH EXACT absorption is used.
For example: giving 3 means that only effect 2 is exactly absorbed.
Number should be:          2  to           4
  >> # Jfilter
  >> 3

Approx. absorption of effect           3
Exact absorption of effect 2 to the first effect
```

- Number 3 is given
- It means within block effects #1 and #2 are absorpted exactly (GE and PE)
- Within block effect #3 is approximately absorpted (HTD)

Variance component summary


```
==================== Variance component summary ====================

 1.  VC  var.=   0.4133      0.3034      0.2559      0.3335      0.2054      0.1702      0.2860      0.1612      0.1367
 2.  VC  var.=    286.9       148.4       99.26       431.7       196.4       136.2       455.6       192.0       147.3
Genetic  var.=    247.3       124.2       87.73       221.6       110.5       79.05       245.0       111.3       89.43
Residual var.=    11.04       9.062       12.20       16.33       10.96       13.26       16.30       8.844       14.96

Total var.    =    545.7       282.0       199.5       670.0       318.0       228.7       717.1       312.3       251.9
Direct Anim V=    247.3       124.2       87.73       221.6       110.5       79.05       245.0       111.3       89.43

h2            =   0.4532      0.4405      0.4399      0.3308      0.3474      0.3457      0.3416      0.3563      0.3551
lambda        =    7.825       8.080       8.094       11.09       10.51       10.57       10.71       10.23       10.27
```

Diagonal matches to Genetic var.

```
> round(G9, 2)
       M1     P1     F1     M2     P2     F2     M3     P3
M1 247.33 156.44 110.53 184.50 111.20  74.26 172.96  96.39  6
P1 156.44 124.23  85.72 116.45  92.12  61.40 108.85  80.30  5
F1 110.53  85.72  87.73  78.57  63.08  65.92  77.27  57.24  6
M2 184.50 116.45  78.57 221.62 140.58  99.78 222.86 135.15 10
P2 111.20  92.12  63.08 140.58 110.48  79.54 143.11 106.09  7
F2  74.26  61.40  65.92  99.78  79.54  79.05 106.54  79.72  8
M3 172.96 108.85  77.27 222.86 143.11 106.54 245.00 151.14 11
P3  96.39  80.30  57.24 135.15 106.09  79.72 151.14 111.29  8
F3  66.23  54.54  62.24 100.88  79.68  80.78 118.56  87.36  8
```

- Variance component summary shows the variance components for the breeding value of interest
- In this example, for sum of 10 selected test-days
- This table must be checked to see possible errors in model definition

Luke

Output

```
            "PEVani"-file: Solutions for Animal Effects

    Column  |  Description
   ----------------------
      1       Code of Animal
      2       Number of Descendants
      3       Number of Observations
      4       Reliability for BV  1
      5       Reliability for BV  2
      6       Reliability for BV  3
      7       Reliability for BV  4
      8       Reliability for BV  5
      9       Reliability for BV  6
     10       Reliability for BV  7
     11       Reliability for BV  8
     12       Reliability for BV  9

head PEVani
            1        1      0  0.28174      0.28023      0.28146      0.26838      0.26938      0....
            4        1      0  0.34724      0.34561      0.34819      0.33918      0.34085      0....
            7        1      0  0.32960      0.32884      0.32994      0.32732      0.32822      0....
           10        1      0  0.33518      0.33314      0.33388      0.31582      0.31515      0....
           13        1      0  0.30093      0.29969      0.30061      0.28986      0.29031      0....
           16        1      0  0.25641      0.25608      0.25632      0.25372      0.25399      0....
            3        2      0  0.26133      0.26051      0.26102      0.24569      0.24662      0....
           20        1      0  0.30781      0.30682      0.30764      0.29848      0.29971      0....
           23        1      0  0.39414      0.39317      0.39339      0.35402      0.35514      0....
           25        1      0  0.30374      0.30217      0.30269      0.29176      0.29173      0....
```

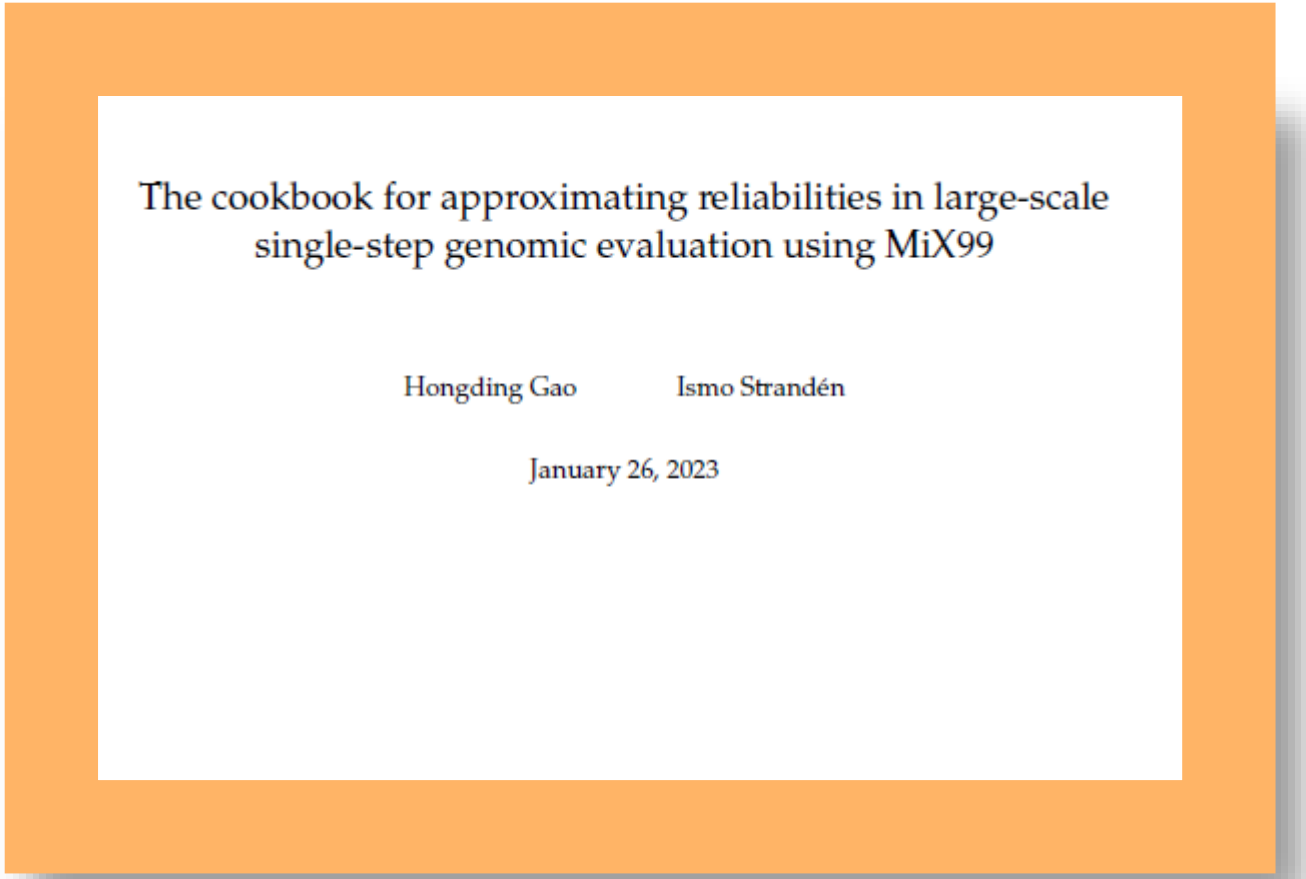
- Reliabilities are written to PEVani file
- Reliabilities are in order of weight matrix:

```
>> # Weights
>> 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 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
```

Genomic reliabilities

Introduction

- Approximating genomic reliabilities is, currently, not so simple
- There is a cookbook about how to approximation can be done with MiX99
- In this presentation an overview of the required steps are explained for getting reliabilities for genotyped animals


The cookbook for approximating reliabilities in large-scale single-step genomic evaluation using MiX99

Hongding Gao Ismo Strandén

January 26, 2023

Steps for approximating genomic reliabilities

- There are 7 steps for getting genomic reliability for all animals in the pedigree
- Steps 1-3 are enough for genotyped animals
- Steps 4-6 are needed if reliabilities are calculated for all animals
- Steps 1-3 are explained in more detail

Step 1. PBLUP reliabilities for all animals

- In the first step PBLUP reliabilities for all animals are calculated for desired traits
- Calculation is explained in the first part of this presentation
- PBLUP Reliabilities for genotyped animals are used in the next step

Step 2. reverse reliability for the genotyped animals

Effective record contribution

- Effective record contribution (ERC) for genotyped animals are obtained using reverse reliability approximation
- It reverses the R2 calculation, from R2 to “count of observations”.
- Reverse R2 approximation takes pedigree and correlated traits in to account when dividing information between animals.
  - => High R2 does not always mean high ERC
- ERC used as a weight in very simple model gives original R2 back!

- In step 2 PBLUP R2 **for genotyped** animals is used as **input**
- The output is **ERC** for **genotyped** animals
- ERC measures how much information each genotyped animal has in PBLUP model

Luke

Step 2. reverse reliability for the genotyped animals, apax99

- ERC calculation can be done with apax99
  - option 20 assumes **independent traits**,
  - option 40 for **correlated traits**,
- Both options also require pedigree
- For both methods, R2 for animals of interest are given as an input!

INPUT files

- Pedigree (animal sire dam)
- R2s (animal r2trait1 r2trait2 …)
- Option 40
  - Genetic covariance matrix of traits
  - Residual covariance matrix of traits

  Needs to be calculated by hand, apax99 output gives only variances
- Option 20
  - Heritabilities for traits

  Can be obtained from Apax99 output

Luke

Step 2. reverse reliability for the genotyped animals, apax99 input file

```
# For OPTION 40
# Reliability method (Accurtype): Reversed reliability approximation
40 p

# Filename of the EBV reliability information (id and r2s)
# Number of rows equal to the number of genotyped animals
# column 1: ID, column 2: r2 of trait 1, r2 of trait 2
ebv_r2.txt

# Pedigree file
pedigree.ped

# Number of ERC values (and r2s):
3

# G0 file (genetic variance-covariance matrix)
G0.txt

# R0 file (residual variance-covariance matrix)
R0.txt

# ERC: tolerc maxit smallest tol_newton maxit_newton (0 default)
1e-6 100 1e-10 1e-4 20

# ERC: tolr2 minit lastit belowit (0 default)
1e-7 10 0.5 10
```

```
35 22 20
22 23 23
20 23 24
```

```
44 18 12
18 48 21
12 21 47
```

- No need to run mix99i before apax!
- Genetic and residual covariance matrices are given as full dense matrices.
- Run ERC calculation

```
apax99 < options.slv
```

Step 2. reverse reliability for the genotyped animals, apax99 output PEVani file

"PEVani"-file: Solutions for Animal Effects

Column	Description
1	Code of Animal
2	Number of Descendants
3	Number or Reliability Observations
4	ERC 1
5	ERC 2
6	ERC 3
7	Reliability for ERC 1
8	Reliability for ERC 2
9	Reliability for ERC 3

head PEVani

13	74	1	1.6701	2.6197	0.97548	0.44502	0.60614	0.45660
15	1	1	0.13560	0.16967	0.77069E-01	0.61127E-01	0.90639E-01	0.62253E-01
19	25	1	0.36228	0.43377	0.20462	0.14817	0.20308	0.14985
26	0	1	0.13885	0.16673	0.78403E-01	0.62499E-01	0.89209E-01	0.63263E-01
41	4	1	0.61178	0.75926	0.34638	0.22704	0.30846	0.22980

ERCs for genotyped animals

R2 calculated from ERC (no pedigree accounted)

$$\alpha = (1 - h^2)/h^2$$

$$R^2 = ERC/(ERC + \alpha)$$

Model reliabilities usually are higher than this value

Step 3 GEBV reliabilities for genotyped animals

- Idea is to use SNBLUP/GBLUP model
- ERC from STEP2 as weight for animal
- `Snp_blup_rel` program is used

3. Theory behind **snp_blup_rel**

SNP BLUP and GBLUP are equivalent models. Assume:

- genotypes in centered and scaled matrix $\mathbf{Z}$
- observations in vector $\mathbf{y}$
- weights (usually ERC/EDC) in diagonal matrix $\mathbf{W}$

VanRaden I: $\mathbf{Z} = (\mathbf{M} - \mathbf{P})/\sqrt{k}$

Denote SNP effects $\mathbf{g}$ and correspondingly DGVs $\mathbf{u} = \mathbf{Zg}$

and let $\text{var}(\mathbf{g}) = \mathbf{D}\sigma_u^2$ and $\text{var}(\mathbf{u}) = \mathbf{ZDZ}'\sigma_u^2$

In GBLUP you solve MME for $\hat{\mathbf{u}}$

$$\begin{bmatrix} \mathbf{1}'\mathbf{W}^{-1}\mathbf{1} & \mathbf{1}'\mathbf{W}^{-1} \\ \mathbf{W}^{-1}\mathbf{1} & \mathbf{W}^{-1} + \lambda(\mathbf{ZDZ}')^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{W}^{-1}\mathbf{y} \\ \mathbf{W}^{-1}\mathbf{y} \end{bmatrix}$$

where $\lambda = \sigma_e^2 / \sigma_u^2$. And consequently

$$PEV(\hat{\mathbf{u}}) = \text{var}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{C}^{u,u}\sigma_e^2$$

where $\mathbf{C}^{u,u}$ is sub-matrix of $\hat{\mathbf{u}}$ in the inverse of the MME coefficient matrix.

Note: The dimension of MME is 1+n,
where n = the number of rows (animals) in the Z matrix.

Luke
NATURAL RESOURCES INSTITUTE FINLAND

Luke

Theory behind **snp_blup_rel**

- All the computations in the program have been optimized and in para-version use multiple cores (when appropriate)

- BUT the program computes direct solutions via inverse:
  it does not iterate solutions
  ➔ the size of the MME is m+1
  – if the number of animals (n) is less than the number of SNPs (m), consider using GBLUP instead of SNP-BLUP

- Main use of **snp_blup_rel** is to compute R^2 and PEV
  – it is computationally efficient to use less SNPs in estimation of R^2 and PEV than used in DGV/GEBV calculations.

  And then scale the R^2 down to resemble the higher marker density.

CAUTION: Note the dimensions of your work.

Scaling R2 down when subset of markers is used

Subset markers

- To predict R^2 of full SNPs based on R^2 of subset SNPs
- Linear relationship between them: $(1 - r^2_{full}) = b * (1 - r^2_{sub})$
- $1 + h^2$ can be used as the b

$$(1 - r^2_{full}) = (1 + h^2) * (1 - r^2_{sub})$$
$$r^2_{full} = (1 + h^2) r^2_{sub} - h^2$$

- For example case:

$$h^2 = 0.432$$
$$r^2_{full} = (1 + 0.432) r^2_{sub} - 0.432$$

- The formula does not take the size of reduction in the account.
- Will produce negative reliabilities for small r^2_{sub} values

Running snp_blup_rel_para

- Below is an example to use every third markers, with the option [−*s* 0 0 3]
- Run it as follows (type *snp_blup_rel_para* to get information for the options/flags)

```
snp_blup_rel_para \
     -nthr 10 -memhigh
     -m PvR1 -c 2pq -h2 0.443 \
     -a base_af.dat -s 0 0 3  -o Id_and_ERC.dat  -wt 2 markers.dat  SNP_rel.out \
     > snp_blup_rel.log
```

10 threads, high memory

Marker data file and output file

```
-a afreq_file : afreq_file is file for allele frequencies by marker,
                allele frequencies must be between 0 and 1.
                Record has format: <marker number> <population allele frequencies>
                Marker numbers are from 1 upwards (id column is column zero).

-s fst last jump : select markers in the genotype file:
                   from fst to last by jump.
                   Default is all markers are used. Zero means default.

-o data_file  : id numbers of animals with observation (input).
                If no data file is given, all individuals have data.

-wt col_num   : column of weights in data_file, negative weights are zeroed.
```

```
-h2    values : heritabilities (as by -straits). NOTE: lambda is calculated to be (1-h2)/h2.
```

Conclusions

- 3 Steps required for approximating reliabilities for genotyped animals was covered.
- A bit of theory of PBLUP and GBLUP reliabilities were explained.
- Calculating pedigree PBLUP reliabilities is perhaps the most challenging part of the pipeline.
- R2 given by snp_blup_rel needs to be scaled down if the subset of markers is used

Questions?