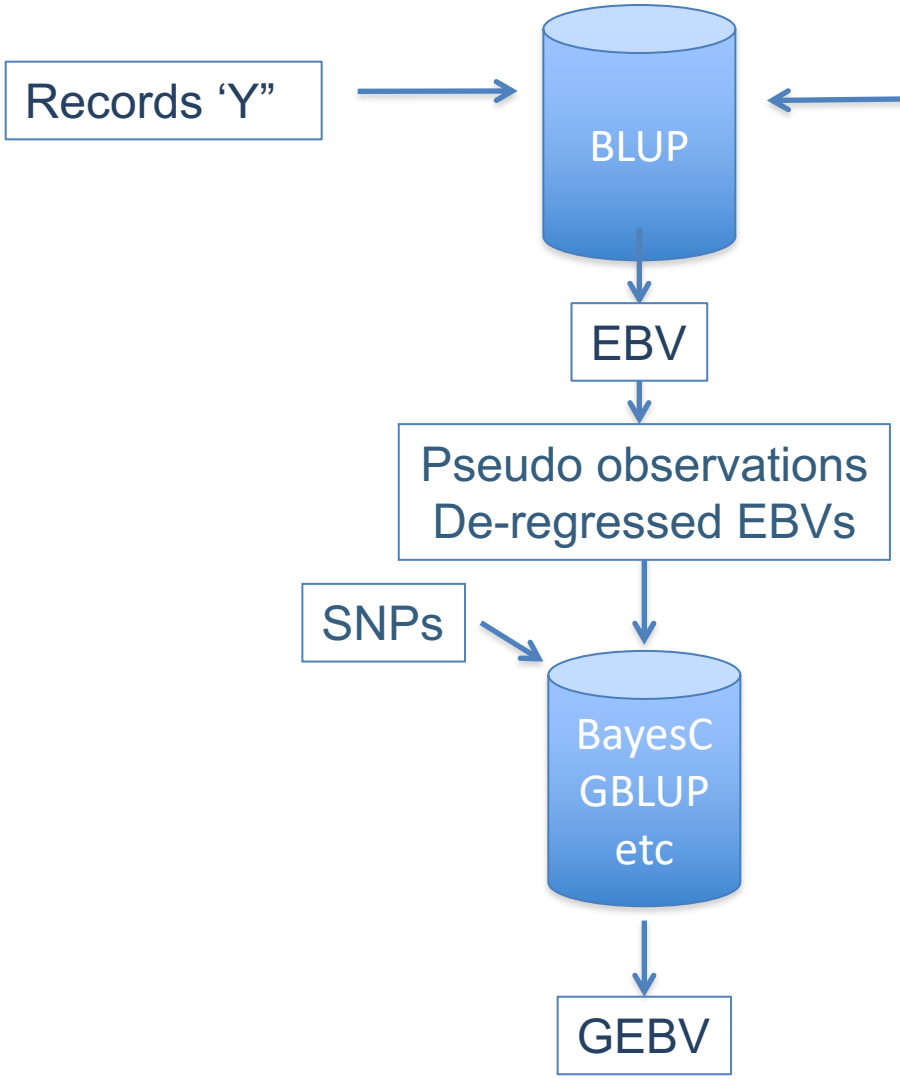
Single Step GBLUP

Multiple-step Genomic evaluation



Not all animals are genotyped

Genomic evaluation

- Estimate effect of all SNPs in the genome:

$$\boldsymbol{y} = \boldsymbol{Za} + \boldsymbol{e}$$

$$\begin{pmatrix} 2 \\ 12 \\ 8 \\ 6.2 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$$

Records

Genotypes

Marker effects

BayesA, BayesB, BLUP-SNP, etc

Genomic evaluation

- What about records?

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

$$\begin{pmatrix} 2 \\ 12 \\ 8 \\ 6.2 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$$

Records

Genotypes

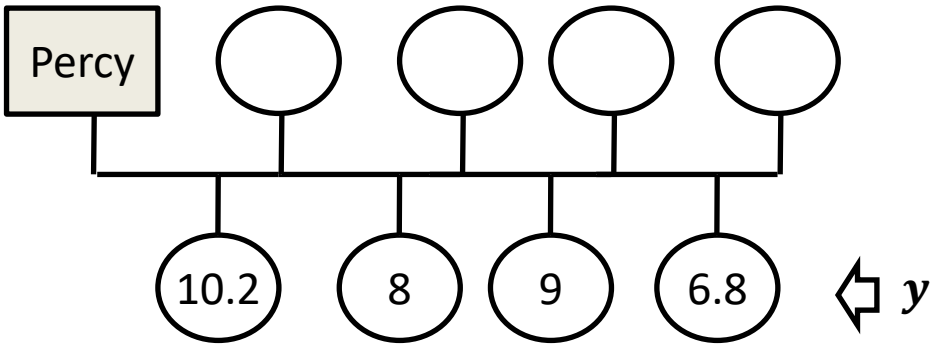
Marker effects

Genomic evaluation

- What about records?
- We genotype key animals (breeding males and maybe females)
  - They may *not* have phenotype on their own
  - They also have progenies who could have phenotype but could not have genotype
- "Project" family phenotypes on genotyped animals
  - Deregressed Proofs, DYD, etc.
  - Let's call this "DP"
- More easy said than done

What about records ?

- Typical in dairy cattle: the male is “assigned” the performance of the daughters
- Similar to a sire model

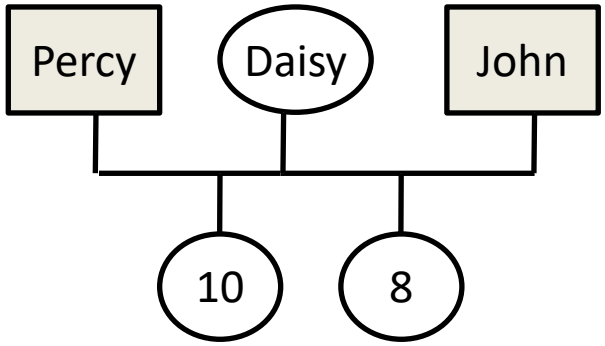


$$DP_{Percy} = \bar{y} = 8.5$$

- But to achieve more accuracy and to avoid selection bias, we need to correct for the dams’ EBV and for the fixed effects
  - This is what we do in DEREGRESSION
- And corrections contain errors which pass on to deregressed proofs

What about records ?

- Assume that Daisy EBV is overestimated as 6.4 (true BV is 4)



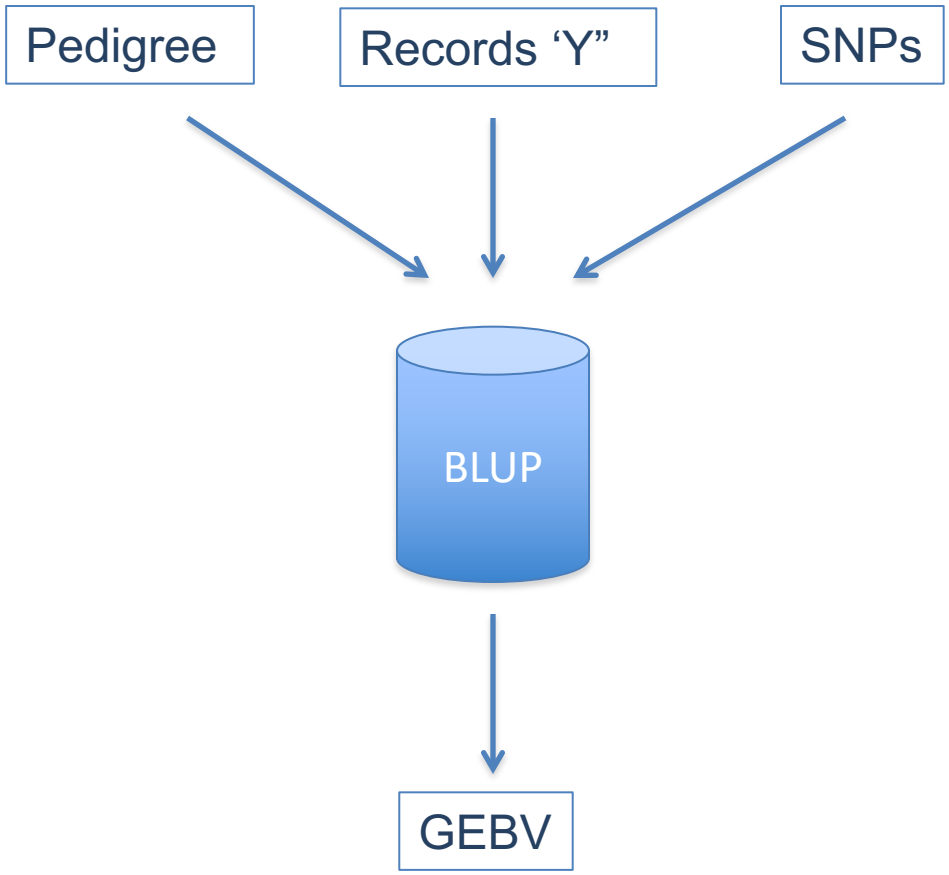
$$DP_{Percy} = 10 - 3.2 = 6.8$$
$$DP_{John} = 8 - 3.2 = 4.8$$

- Now both Percy and John are biased downwards !!
- Sometimes Daisy will be biased upwards and sometimes downwards
- Thus, the deregressed proofs of Percy and John will have a residual covariance
- This covariance is always ignored in practice

- The same problem exists when we correct by effects such as herd

Single-Step Genomic Evaluation

Aguilar et al., 2010



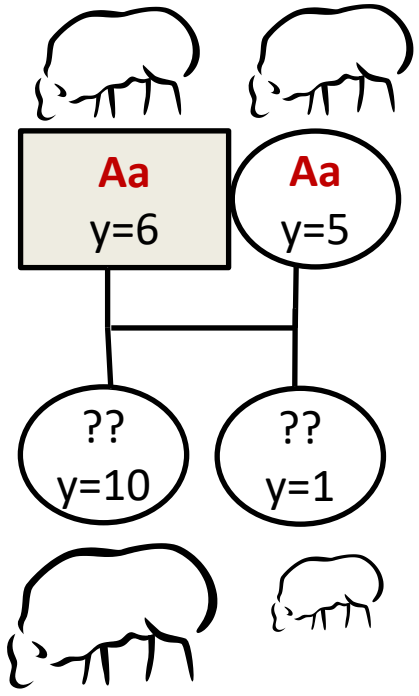
Expand information

- We can do ONE evaluation if we “augment” information generating either
  - genotypes for all animals
  - **G** matrix for all animals

- Imputing algorithms (Beagle, Fimpute, AlphaImpute, etc.) are conceived to impute from low to high density
- For nongenotyped animals, they *may* give a point estimate of the genotype
- Why is this bad?

Problem with point estimates of genotypes

- Imagine a major gene

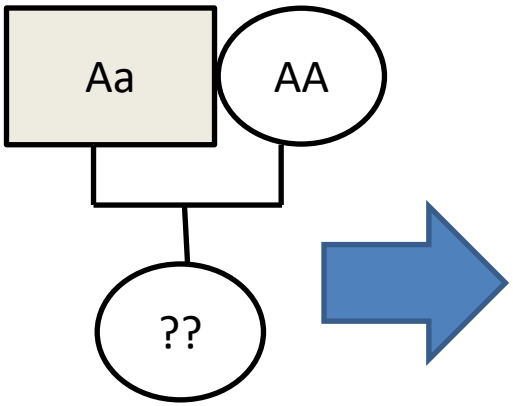


- Point estimate of genotype of the descendants: “Aa”
- Clearly, based on $\boldsymbol{y}$ there is Mendelian segregation where one descendant received “AA” and the other “aa”
- There is variation of true genotype around the point estimate of the genotype
- If we do not consider this variation we consider the offspring as identical twins

Augmenting genotypes

- Gengler et al. (2007) conceived an algebraic way to deal with these point estimates
- Christensen & Lund (2010) showed how to take the variation into account

- Genotype of descendants = half their parents + Mendelian sampling



AA with probability ½
Aa with probability ½

Prediction of Genotype = $\frac{3}{2}"A" + \frac{1}{2}"a"$

Variance(Genotype)=$\frac{1}{4}"A" + \frac{1}{4}"a"$

Missing data

Fill-in missing data: data augmentation

- Augmenting = adding genotypes
- But we need to account for the fact that these are « guesses »

Single Step as a missing data problem

- We can see genotype as a missing data problem (Christensen & Lund, 2010)
- Use the prediction and the distribution of the prediction (if not the procedure does not work)

non genotyped

$$\text{Let} \quad \mathbf{A} = \overbrace{\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}}^{\text{genotyped}}$$

Inferring genotypes

- There is Gengler's gene content prediction J. Dairy Sci. 91:1652
  - Linear approximation to the imputation problem
  - This method can be applied to any member of a pedigr

- Using centered gene content:

$$\hat{\mathbf{Z}}_1 = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{Z}_2$$

- Christensen and Lund realized that

$$Var(\hat{\boldsymbol{Z}}_1|\boldsymbol{Z}_2) = (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})\boldsymbol{V}$$

where $\boldsymbol{V}$ contains $2p_kq_k$ in the diagonal

Inferring genotypes

- Instead of working with individual SNP effects, we will define
  - **u=Za**
  - i.e., the genetic value is the sum of SNP effects
  - We're not really interested in **a** themselves but in **u** (we know from GBLUP that we can jump from one to the other)
  - Moreover, we're interested in the distribution of **u**'s, so that we can compute their covariances and put them into the MME

Christensen & Lund key idea:

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_2 \\ \boldsymbol{u}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{Z}_2 \\ \boldsymbol{Z}_1 \end{pmatrix} \boldsymbol{a}$$

Breeding values

SNP effects

1= « non genotyped »
2= « genotyped »

Re-create GBLUP…

Chistensen & Lund use $Var(A) = E\big(Var(A|B)\big) + Var\big(E(A|B)\big)$ to consider the prediction of the genotype and its variance

$$\mathrm{Var}(\boldsymbol{u}) = \begin{pmatrix} \boldsymbol{Z}_2 \\ \widehat{\boldsymbol{Z}}_1 \end{pmatrix} Var(\boldsymbol{a}) (\boldsymbol{Z}_2' \quad \widehat{\boldsymbol{Z}}_1) + \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & Var(\widehat{\boldsymbol{Z}}_1) \end{pmatrix} Var(\boldsymbol{a})$$

$E(\boldsymbol{Z}_1|\boldsymbol{Z}_2)$

$1/2\Sigma p_i q_i$

$Var(\boldsymbol{Z}_1|\boldsymbol{Z}_2)$

Using Gengler's results

Resulting in:

Covariances of all animals

Legarra et al. 2009; Aguilar et al., 2010; Christensen & Lund, 2010

$$Var\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix}=\mathbf{H}=\begin{bmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{bmatrix}=\begin{bmatrix}\overbrace{\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}+\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}}^{\text{non genotyped}} & \underbrace{\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}}_{\text{genotyped}}\\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G}\end{bmatrix}$$

Let $\mathbf{A}=\begin{bmatrix}\overbrace{\mathbf{A}_{11}}^{\text{non genotyped}} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix}$

Covariances of all animals

$$Var\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix} = \mathbf{H} = \begin{bmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{bmatrix} = \left[\begin{array}{c|c} \overbrace{\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}}^{\text{This is the error in the prediction}} + \overbrace{\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}}^{\text{This is the variance of prediction of genotypes } \textit{from} \text{ genotyped } \textit{to} \text{ non-genotyped}} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \hline \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{array}\right]$$

The prediction « generates » a covariance

G comes from genotypes

$$Var\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix}=\mathbf{H}=\begin{bmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{bmatrix}=$$

$$\begin{bmatrix}\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}+\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G}\end{bmatrix}$$

- Incredibly: $\mathbf{H}^{-1}$ is very simple:

$$\mathbf{H}^{-1}=\mathbf{A}^{-1}+\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\end{bmatrix}$$

…and avoiding « double counting »

Inverse of the regular pedigree relationship matrix

Correcting for genomic relationships…

- **Things would be simple if we had genomic relationships for everyone (Legarra et al., 2009)**
- Things would be simple if we could add genotypes for all animals (Christensen et al., 2010)

Overall modification

- Look at **A** as a « prior » relationship and to **G** as an « observed » relationship
  - **G** is observed for some individuals only, whose « a priori » relationship matrix was $\mathbf{A}_{22}$
- Try to construct a « posterior » relationship matrix

Joint distributions

Unconditional distribution of genetic values of Genotyped individuals

$$p(\mathbf{u}_2) = N(\mathbf{0}, \mathbf{G}) \text{ and}$$

After seeing their genotypes !

Conditional distribution of Non-Genotyped individuals

$$p(\mathbf{u}_1 | \mathbf{u}_2) = N(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2, \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})$$

Because they have no genotypes, this depends only on pedigree

$$p(\mathbf{u}_1, \mathbf{u}_2) = p(\mathbf{u}_2)\, p(\mathbf{u}_1 | \mathbf{u}_2)$$

Joint distribution

Joint distributions

$$p(\mathbf{u}_1, \mathbf{u}_2) = p(\mathbf{u}_1, | \mathbf{u}_2) p(\mathbf{u}_2)$$

prediction of non genotyped from genotyped

"Genomic" relationships

$$= p(\mathbf{u}_1 | \mathbf{u}_2) p(\mathbf{u}_2)$$

$$\propto \exp[-0.5(\mathbf{u}_1 - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2)'\mathbf{A}^{11}(\mathbf{u}_1 - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2)]\exp[-0.5\mathbf{u}_2'\mathbf{G}^{-1}\mathbf{u}_2]$$

$$= \exp\left(-0.5 \begin{bmatrix} \mathbf{u}_1' & \mathbf{u}_2' \end{bmatrix} \begin{bmatrix} \mathbf{A}^{11} & -\mathbf{A}^{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}^{11} & \mathbf{G}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}^{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \right)$$

$$= \exp\left(-0.5 \begin{bmatrix} \mathbf{u}_1' & \mathbf{u}_2' \end{bmatrix} \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{G}^{-1} + \mathbf{A}^{22} - \mathbf{A}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} \right).$$

…for those inclined to algebra

Joint distributions

$$p(\mathbf{u}_2) = N(\mathbf{0}, \mathbf{G})$$

$$p(\mathbf{u}_1 | \mathbf{u}_2) = N\left(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2, \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)$$

Joint distributions

$$p(\mathbf{u}_2) = N(\mathbf{0}, \mathbf{G}) \quad \Rightarrow \quad Var(\mathbf{u}_2) = \mathbf{G}$$

$$p(\mathbf{u}_1 | \mathbf{u}_2) = N\left(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2, \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)$$

Joint distributions

$$p(\mathbf{u}_2) = N(\mathbf{0}, \mathbf{G}) \quad \Rightarrow \quad Var(\mathbf{u}_2) = \mathbf{G}$$

$$p(\mathbf{u}_1 | \mathbf{u}_2) = N\left(\underbrace{\mathbf{A}_{12}\mathbf{A}_{22}^{-1}}\mathbf{u}_2, \underbrace{\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}} \right)$$

$$Var(\mathbf{u}_1) = \overbrace{\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}} + \overbrace{\mathbf{A}_{12}\mathbf{A}_{22}^{-1}}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$$

because Var(**Xt**) = **X**Var(**t**)**X**'

Joint distributions

$$p(\mathbf{u}_2) = N(\mathbf{0}, \mathbf{G}) \quad \Rightarrow \quad Var(\mathbf{u}_2) = \mathbf{G}$$

$$p(\mathbf{u}_1 | \mathbf{u}_2) = N(\underbrace{\mathbf{A}_{12}\mathbf{A}_{22}^{-1}}\mathbf{u}_2, \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})$$

$$Cov(\mathbf{u}_1, \mathbf{u}_2) = \overbrace{\mathbf{A}_{12}\mathbf{A}_{22}^{-1}}\mathbf{G}$$

because Cov(**Xt**,**t**) = **X**Var(**t**)

Covariances of all animals

Legarra et al. 2009; Aguilar et al., 2010; Christensen & Lund, 2010

$$Var\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix}=\mathbf{H}=\begin{bmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{bmatrix}=\begin{bmatrix}\overbrace{\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}+\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}}^{\text{non genotyped}} & \underbrace{\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}}_{\text{genotyped}}\\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G}\end{bmatrix}$$

$$Var\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix}=\mathbf{H}=\begin{bmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{bmatrix}=$$

$$\begin{bmatrix}\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}+\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G}\end{bmatrix}$$

- Incredibly: $\mathbf{H}^{-1}$ is very simple:

$$\mathbf{H}^{-1}=\mathbf{A}^{-1}+\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\end{bmatrix}$$

…and avoiding « double counting »

Inverse of the regular pedigree relationship matrix

Correcting for genomic relationships…

Understanding H matrix

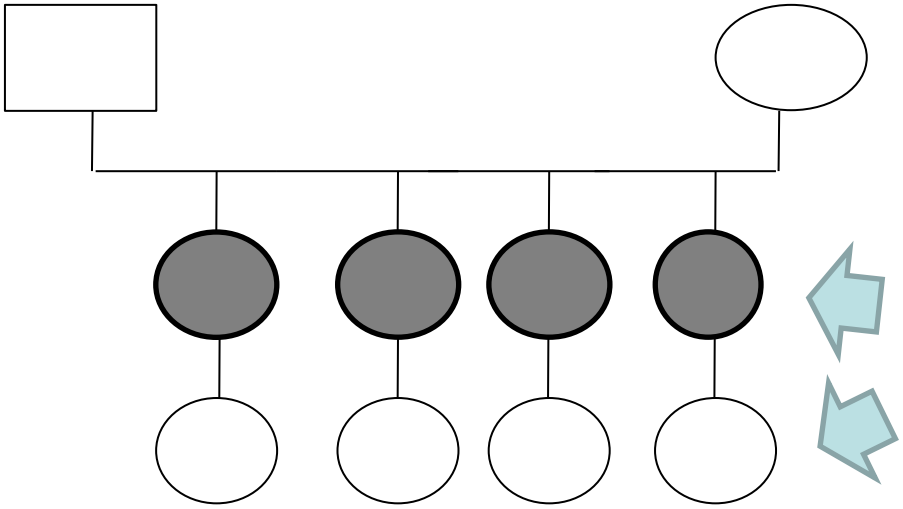
- It is a projection of **G** matrix on the rest of individuals "so that" **G** matrix makes sense
  - e.g. parents of two animals related in **G** should be related in **A**

- It is a Bayesian updating of the pedigree matrix based on new information from genotypes
- The approximation of multivariate normality is good because we have *many* markers

- Typically
  - $\mathbf{A}^{-1}$ in the millions but extremely sparse
  - **G** and $\mathbf{A}_{22}$ in the thousands
  - Leads to a very efficient method of genomic evaluation:
    - **Single Step GBLUP**

Examples on H matrix

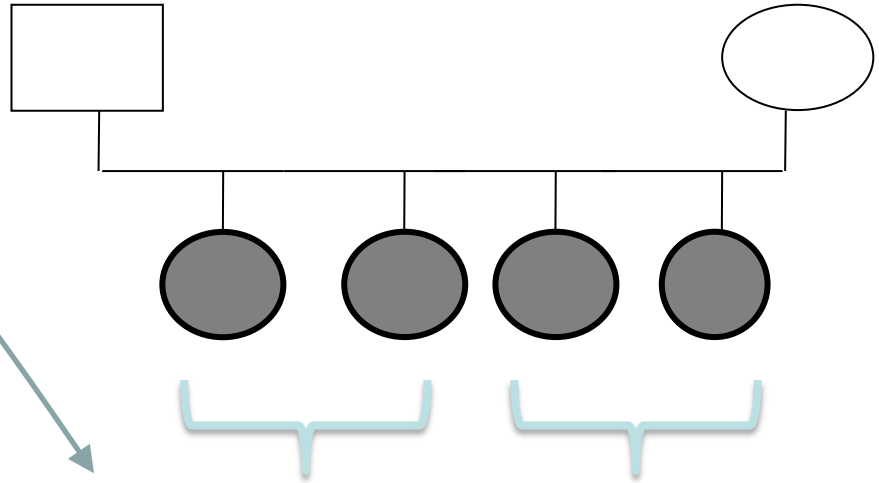
- Consider 4 full-sibs with one progeny each



- With pedigree, sibs are related by 0.5
- their offspring are cousins with a relationship of 0.125

- The 0.5 assumes infinite unlinked loci, with actual genomes relationship varies: 0.5 ± 0.05

- Pedigree; grey is genotyped



G=

1.0	0.7	0.3	0.3
0.7	1.0	0.3	0.3
0.3	0.3	1.0	0.7
0.3	0.3	0.7	1.0

More related than average

More related than average

Less related than average

How is this included in **A**?

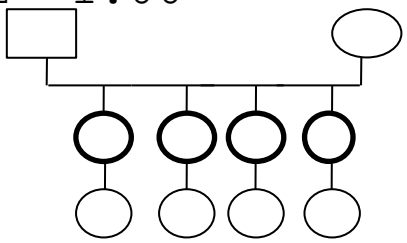
Classical A (pedigree)

1.00	0.00	0.50	0.50	0.50	0.50	0.25	0.25	0.25	0.25
0.00	1.00	0.50	0.50	0.50	0.50	0.25	0.25	0.25	0.25
0.50	0.50	1.00	0.50	0.50	0.50	0.50	0.25	0.25	0.25
0.50	0.50	0.50	1.00	0.50	0.50	0.25	0.50	0.25	0.25
0.50	0.50	0.50	0.50	1.00	0.50	0.25	0.25	0.50	0.25
0.50	0.50	0.50	0.50	0.50	1.00	0.25	0.25	0.25	0.50
0.25	0.25	0.50	0.25	0.25	0.25	1.00	0.12	0.12	0.12
0.25	0.25	0.25	0.50	0.25	0.25	0.12	1.00	0.12	0.12
0.25	0.25	0.25	0.25	0.50	0.25	0.12	0.12	1.00	0.12
0.25	0.25	0.25	0.25	0.25	0.50	0.12	0.12	0.12	1.00

Full-sibs is 0.50

Cousins is 0.125

Uncle-nephew is 0.25



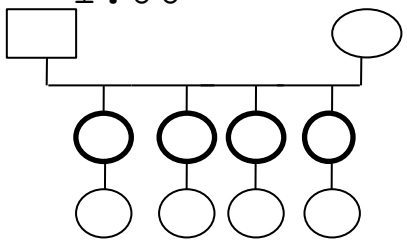
H (pedigree + markers)

1.00	0.00	0.50	0.50	0.50	0.50	0.25	0.25	0.25	0.25
0.00	1.00	0.50	0.50	0.50	0.50	0.25	0.25	0.25	0.25
0.50	0.50	1.00	0.70	0.30	0.30	0.50	0.35	0.15	0.15
0.50	0.50	0.70	1.00	0.30	0.30	0.35	0.50	0.15	0.15
0.50	0.50	0.30	0.30	1.00	0.70	0.15	0.15	0.50	0.35
0.50	0.50	0.30	0.30	0.70	1.00	0.15	0.15	0.35	0.50
0.25	0.25	0.50	0.35	0.15	0.15	1.00	0.17	0.07	0.07
0.25	0.25	0.35	0.50	0.15	0.15	0.17	1.00	0.07	0.07
0.25	0.25	0.15	0.15	0.50	0.35	0.07	0.07	1.00	0.17
0.25	0.25	0.15	0.15	0.35	0.50	0.07	0.07	0.17	1.00

Full-sibs is 0.70 – 0.30

Uncle-nephew is 0.35-0.15

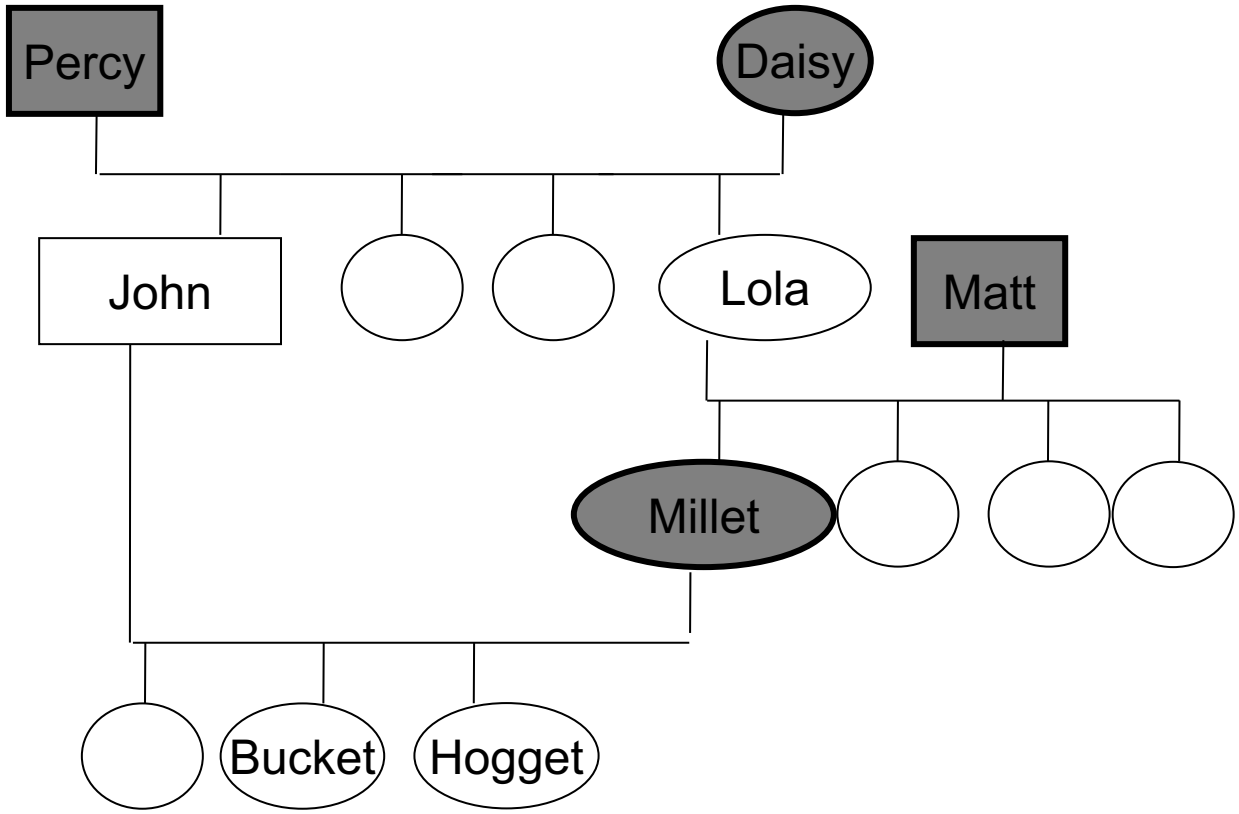
Cousins is 0.17 - 0.07



We have extended genomic relationships to all the pedigree

More complex example

- Pedigree; grey is genotyped

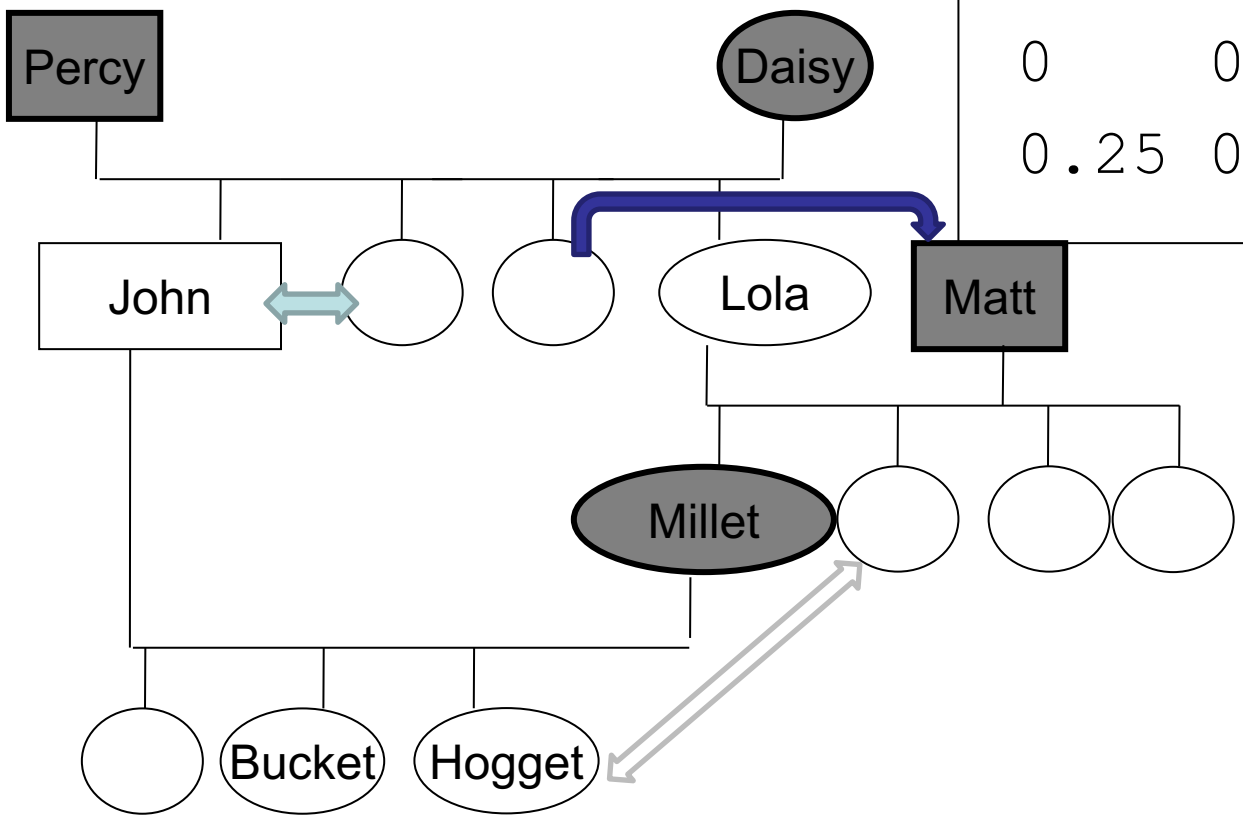


More complex example

- Before genotyping

$\mathbf{A}_{22}=$

1.00	0	0	0.25
0	1	0	0.25
0	0	1	0.50
0.25	0.25	0.50	1.00

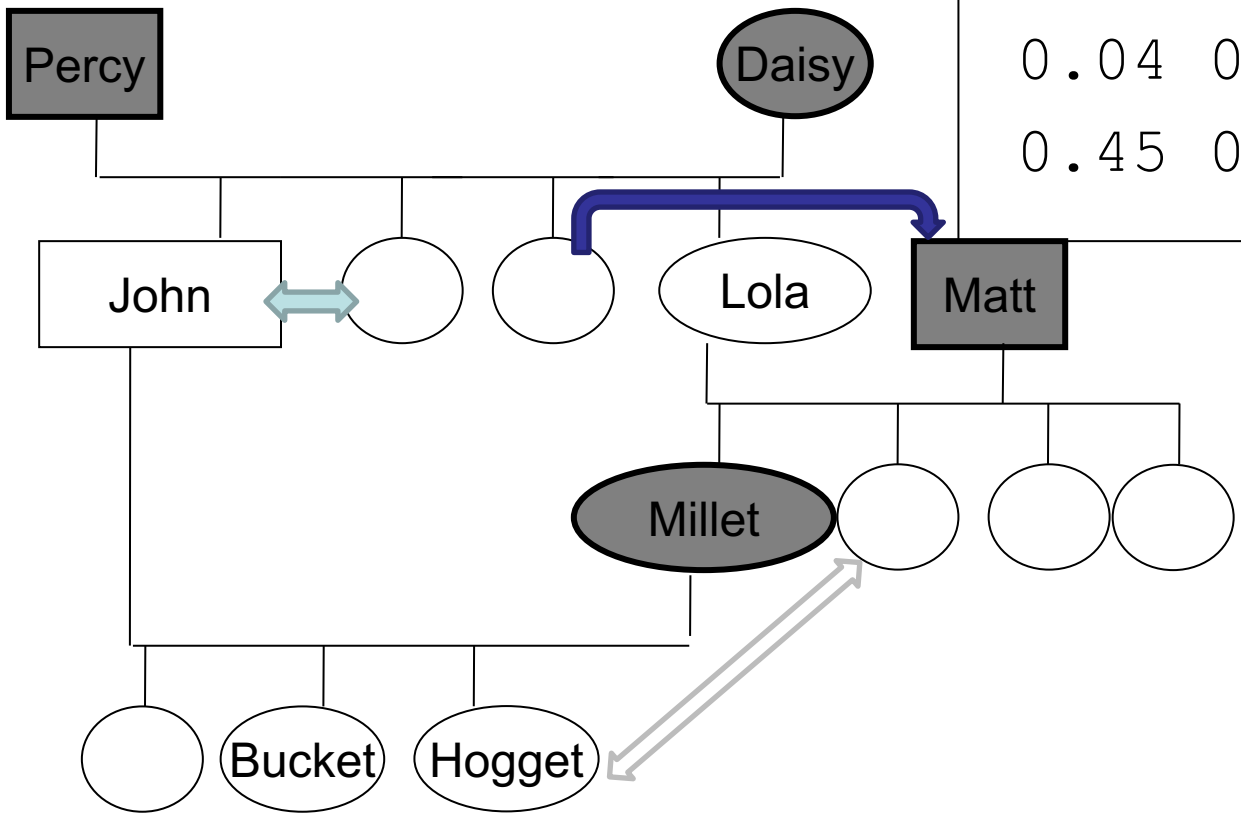


More complex example

- After genotyping

G=

1.13	0.08	0.04	0.45
0.08	0.91	0.14	0.32
0.04	0.14	1.12	0.62
0.45	0.32	0.62	1.10



Classical A (pedigree)

1.00	0.00	0.00	0.50	0.50	0.50	0.5	0.25	0.25	0.25	0.25	0.38	0.38	0.38
0.00	1.00	0.00	0.50	0.50	0.50	0.5	0.25	0.25	0.25	0.25	0.38	0.38	0.38
0.00	0.00	1.00	0.00	0.00	0.00	0.0	0.50	0.50	0.50	0.50	0.25	0.25	0.25
0.50	0.50	0.00	1.00	0.50	0.50	0.5	0.25	0.25	0.25	0.25	0.62	0.62	0.62
0.50	0.50	0.00	0.50	1.00	0.50	0.5	0.25	0.25	0.25	0.25	0.38	0.38	0.38
0.50	0.50	0.00	0.50	0.50	1.00	0.5	0.25	0.25	0.25	0.25	0.38	0.38	0.38
0.50	0.50	0.00	0.50	0.50	0.50	1.0	0.50	0.50	0.50	0.50	0.50	0.50	0.50
0.25	0.25	0.50	0.25	0.25	0.25	0.5	1.00	0.50	0.50	0.50	0.62	0.62	0.62
0.25	0.25	0.50	0.25	0.25	0.25	0.5	0.50	1.00	0.50	0.50	0.38	0.38	0.38
0.25	0.25	0.50	0.25	0.25	0.25	0.5	0.50	0.50	1.00	0.50	0.38	0.38	0.38
0.25	0.25	0.50	0.25	0.25	0.25	0.5	0.50	0.50	0.50	1.00	0.38	0.38	0.38
0.38	0.38	0.25	0.62	0.38	0.38	0.5	0.62	0.38	0.38	0.38	1.12	0.62	0.62
0.38	0.38	0.25	0.62	0.38	0.38	0.5	0.62	0.38	0.38	0.38	0.62	1.12	0.62
0.38	0.38	0.25	0.62	0.38	0.38	0.5	0.62	0.38	0.38	0.38	0.62	0.62	1.12

Full-sibs is 0.50

Unrelated is 0

Uncle-nephew is 0.38

H (pedigree + markers)

1.13	0.08	0.04	0.60	0.60	0.60	0.65	0.45	0.34	0.34	0.34	0.52	0.52	0.52
0.08	0.91	0.14	0.50	0.50	0.50	0.50	0.32	0.32	0.32	0.32	0.41	0.41	0.41
0.04	0.14	1.12	0.09	0.09	0.09	0.09	0.62	0.61	0.61	0.61	0.35	0.35	0.35
0.60	0.50	0.09	1.05	0.55	0.55	0.58	0.38	0.33	0.33	0.33	0.72	0.72	0.72
0.60	0.50	0.09	0.55	1.05	0.55	0.58	0.38	0.33	0.33	0.33	0.47	0.47	0.47
0.60	0.50	0.09	0.55	0.55	1.05	0.58	0.38	0.33	0.33	0.33	0.47	0.47	0.47
0.65	0.50	0.09	0.58	0.58	0.58	1.09	0.62	0.59	0.59	0.59	0.60	0.60	0.60
0.45	0.32	0.62	0.38	0.38	0.38	0.62	1.10	0.62	0.62	0.62	0.74	0.74	0.74
0.34	0.32	0.61	0.33	0.33	0.33	0.59	0.62	1.10	0.60	0.60	0.48	0.48	0.48
0.34	0.32	0.61	0.33	0.33	0.33	0.59	0.62	0.60	1.10	0.60	0.48	0.48	0.48
0.34	0.32	0.61	0.33	0.33	0.33	0.59	0.62	0.60	0.60	1.10	0.48	0.48	0.48
0.52	0.41	0.35	0.72	0.47	0.47	0.60	0.74	0.48	0.48	0.48	1.23	0.73	0.73
0.52	0.41	0.35	0.72	0.47	0.47	0.60	0.74	0.48	0.48	0.48	0.73	1.23	0.73
0.52	0.41	0.35	0.72	0.47	0.47	0.60	0.74	0.48	0.48	0.48	0.73	0.73	1.23

Full-sibs is 0.55

Uncle-nephew is 0.48

"Unrelated" is 0.14 ⇨ Because pedigree founders are related in **G**

Some properties of **H**

- Semi-positive definite always
- Positive definite & invertible iff **G** is invertible
- In practice, if **G** is too different (wrong pedigree or genotyping) from $\mathbf{A}_{22}$, this gives lots of numerical problems
- If everyone is genotyped, Single Step is GBLUP
- If no one is genotyped, Single Step is BLUP

H matrix

- H is then a relationship matrix constructed with markers and pedigree
- But Henderson taught us how to use relationship matrices of any kind

Single step GBLUP

Single Step = Your regular BLUP with small modifications

W: incidence matrix of animals on data

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}\mathbf{R}^{-1}\mathbf{W} + \mathbf{H}^{-1}\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

G

A: pedigree relationship matrix

$\mathbf{A}_{22}$: pedigree matrix among genotyped individuals

This **G** could be *any* matrix describing « genomic » covariances of breeding values;

it does not restrict to VanRaden's (2008) GBLUP

Single step GBLUP

- So the Single Step GBLUP is like regular BLUP changing one small submatrix !!!
- It is almost too simple to be true…

Single Step GBLUP

- Easy modification to a general purpose BLUP software
  - Only changes: addition of $\boldsymbol{G}^{-1}$ and $\boldsymbol{A}_{22}^{-1}$
  - Matrices $\boldsymbol{G}^{-1}$ and $\boldsymbol{A}_{22}^{-1}$ can be computed with external tools

- Can fit any model (probit, GxE,…)

- Simple extraction of SNP effects for indirect prediction or (multimarker) GWAS:

$$\hat{\boldsymbol{a}} = \frac{\boldsymbol{Z}'\boldsymbol{G}^{-1}\hat{\boldsymbol{u}}_2}{2\sum pq}$$

- Avoids selection bias due to genomic preselection

  (Patry & Ducrocq, 2011)

Single Step GBLUP

- What models have we fit so far in SSGBLUP?
  - Multiple traits (up to 18 so far)
  - Multiple trait + correlated genetic maternal effects (beef cattle)
  - Random regressions (lactation curves)
  - Threshold (probit) models
  - Horse rankings (Thurstonian model)
- *Anything* that was fit in BLUP can be fit in SSGBLUP, changing **A** to **H**