Single Step GBLUP

Andrés Legarra

Toulouse, France

Acknowledgements

- ANR projects Amasgen, Rules&Tools; Apisgene
- Toulouse bioinformatics platform (bioinfo.genotoul.fr)
- GENOMIA funding:

  www.poctefa.eu

Why 2-step procedure

- **y**= μ + **Za** + **e**
  - **y** = data
  - **Z**= incidence matrix of marker effects
  - **a**= marker effect
  - **e**=residuals
- Most often, genotyped animals (bulls) do not have data (trait record)
- Further, most animals with phenotype are not genotyped (e.g. cows)
- This limits practical applications
- Need to get pseudo-data for genotyped animals

Pseudo-data

- So we need pseudo-data
- EBV’s
- DYD’s

Pseudo-data

- EBV's
- The problem with EBV's is that they already share information among individuals
- e.g., a dam EBV is = own yield + parent average + progeny contribution
- But we are including information of the sire in the cow, yet not all SNPs of the sire are in the cow

Pseudo-data

- DYD's avoid *part* of these problems (Van Raden Wiggans 1991)
- DYD = daughter yield deviation
- Record of the daughter, corrected by environmental effects and dam's EBV
- Thus DYD = 0.5 BV sire + mendelian sampling
- E(DYD)=0.5 BV sire
- YD's exist for cows
  - YD = record –environmental effects

Pseudo-data

Problems of DYD’s / YD’s

- YD’s little reliable and subject to preferential treatment
- DYD’s not reliable for many species (sheep, swine)
- Hard to define for some species/traits (maternal effects)
- Extremely complex procedure
- Loss of generality

- Things would be simple if we could add genotypes for all animals

Missing data

Fill-in missing data: data augmentation

- *« data augmentation refers to a scheme of augmenting the observed data so as to make it more easy to analyze »* (Tanner & Wong, 1987)
  - Two flavors: EM and Bayesian (Posterior distributions)
  - For instance: pretending (temporarily) that you know the EBV's simplifies REML -> EMREML
- Augmenting = imputation

Missing data

Fill-in missing data: data augmentation

- Augmenting = imputation
- In both flavors (EM and Bayesian) the procedure needs to « know » that you are *pretending* to know
  - e.g. You don't really know EBVs
  - But you include a *trace(**AC**^{uu})* that informs EMREML about the exactitude of the EBV's

Single Step as a missing data problem

- We can see genotype as a missing data problem (Christensen & Lund, 2010)
- « Genotype » :
  - at the SNPs
  - at multiallelic markers (haplotypes)
  - at the genes/QTLs themselves
- the following derivations are very general

Inferring genotypes

- Genotypes in some individuals can be inferred, but only to some extent
- This is feasible for key individuals (fathers with many progeny genotyped)
- Or by imputing data from parents into an animal genotyped with a SNP chip
  - `Fimpute`, `findhap`, etc
- There is also Gengler's gene content prediction J. Dairy Sci. 91:1652

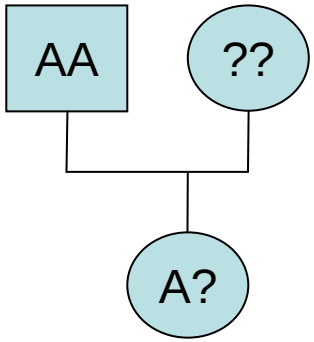
Gengler method

- Linear approximation to the imputation problem
- Assume genotypes are coded as « gene content »: {AA Aa aa} are {0 1 2}
  - **g**= **z**'s not centered in BLUP_SNP
- Gene content can be seen as a quantitative trait
- And we can do BLUP (with pedigree) for this « trait »

$$\hat{\mathbf{z}}_{non\,genotyped} = \left(\mathbf{1} \quad \mathbf{A}_{non\,genotyped,genotyped}\mathbf{A}^{-1}_{genotyped}\right)\begin{pmatrix} \mu_{genotyped} \\ \mathbf{z}_{genotyped} - \mathbf{1}\mu_{genotyped} \end{pmatrix}$$

Example

- Consider a cow daughter of a genotyped bull

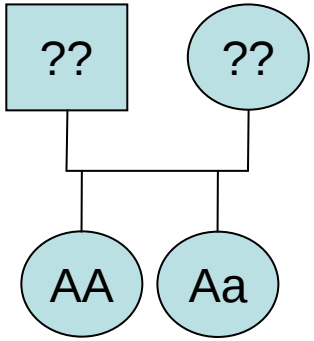


prediction of gene content: $\mathbf{g}_{ij} = 1+p$

from its father

p=freq('a') from the overall population

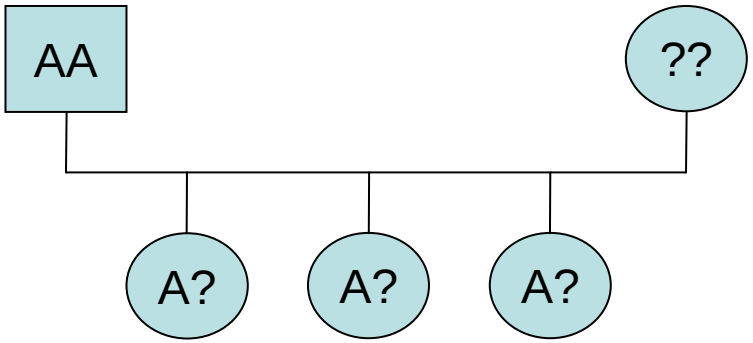
- Consider the parents of two genotyped bulls



prediction of gene content: $\mathbf{g}_{ij} = 1-0.66p$

Example

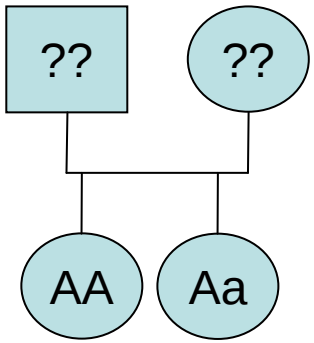
- Consider 3 cows daughters of a genotyped bull



prediction of gene content: $\mathbf{g}_{ij} = 1+p$

WE FORCE THESE COWS TO BE IDENTICAL (CLONES)

- Consider the parents of two genotyped bulls

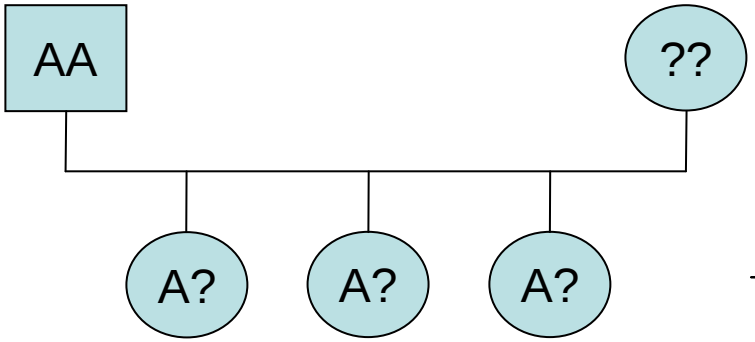


prediction of gene content: $\mathbf{g}_{ij} = 1-0.66p$

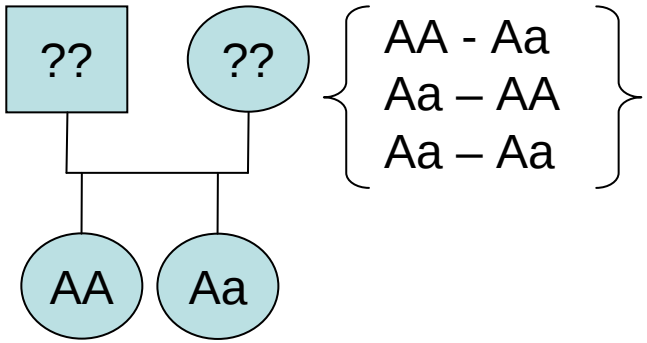
WE FORCE BOTH PARENTS TO BE IDENTICAL (CLONES)

Uncertainty

- In practice the reality is more complex. We can have different « solutions » that are mutually exclusive



1-p cows will be AA
p cows will be Aa

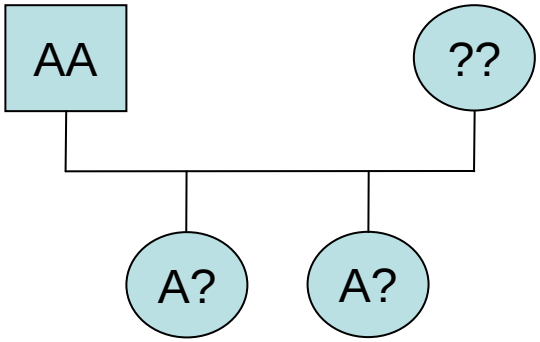


AA - Aa
Aa – AA
Aa – Aa

Some combinations of genotypes are mutually exclusive

Uncertainty

- Consider two cows and *p*. We have four possible states with associated probabilities



genotype1	genotype2	Prob
AA	AA	q^2
Aa	AA	pq
AA	*Aa*	qp
Aa	*Aa*	p^2

We can construct the 4 associated **G** 's and weight them by « Prob », this gives:

$$\mathbf{G} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

if p=0.5

They are not identical by they are similar

$$\mathbf{G} = \begin{pmatrix} 0.83 & 0.33 \\ 0.33 & 0.83 \end{pmatrix}$$

if p=0.4

Imputation

- Things get even more complicated
  - when we consider mendelian incompatibilities
  - markers as linked
- Still, this is doable:
  - **Fimpute, findhapf90, AlphaImpute, PhaseBook**…
  - These are very accurate for close relationships or to impute from low-density to high-density SNPs
- Yet they *all* would impute the cows in the example above as identical cows (or clones)
  - It is impossible in practice to consider all possible genotypes *jointly* (as in the example above)

Linear form of imputation: preliminars

non genotyped

Let $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$ be the pedigree-based relationship matrix,

genotyped

$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}$ can be constructed using Henderson's rules and is used in regular BLUP

REMEMBER !! $\mathbf{A}_{22}^{-1} \neq \mathbf{A}^{22}$

Linear form of imputation

- Let's go back to Gengler's method
- (Christensen and Lund): we can predict gene content of *j* from gene content of *i*
  - And its distribution (uncertainty)

$$\hat{\mathbf{g}}_j = E\left(\mathbf{g}_j \mid \mathbf{g}_i\right) = \mathbf{2}p + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\left(\mathbf{g}_i - \mathbf{2}p\right)$$

$$Var\left(\mathbf{g}_j \mid \mathbf{g}_i\right) = \left(\mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)2pq$$

This is simple selection index machinery (i.e. multivariate normality)

Linear form of imputation

$$\hat{\mathbf{g}}_j = E\left(\mathbf{g}_j \mid \mathbf{g}_i\right) = \mathbf{2}p + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\left(\mathbf{g}_i - \mathbf{2}p\right)$$

$$Var\left(\mathbf{g}_j \mid \mathbf{g}_i\right) = \left(\mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)2pq$$

- This is an approximation: linkage & mendelian rules (incompatibilities) are *not* used
  - But the same approximation is done working with pseudo-data (DYD's)
  - For individuals far away, the linear approximation is very good
- The same expression works for linear functions of gene contents (i.e. breeding values)
  - This is why Legarra et al. (2009) and Christensen & Lund (2010) arrive to the same expression, even if the former did not think in Gengler's method

Linear form of imputation

- 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

Joint distributions

- Using these identities, and summing over all SNPs, we can derive a joint distribution of breeding values
  - In the following, I will put $\sigma^2_u=1$ (to simplify notation)

Unconditional distribution of genetic values of Genotyped individuals (GBLUP)

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

There is an assumption of normality of the distributions which implies no major genes… as in pedigree BLUP

Joint distributions

Unconditional distribution of genetic values of Genotyped individuals (GBLUP)

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

Conditional distribution of Non-Genotyped individuals

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

It is a function of $\mathbf{u}_2$

Joint distributions

Unconditional distribution of genetic values of Genotyped individuals

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

Conditional distribution of Non-Genotyped individuals

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

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

Joint distribution

Joint distributions

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

…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

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

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

$$Var(\mathbf{u}_2) = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$$

because Var(**aX**) = **X**Var(**a**)**X**’

Joint distributions

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

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

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

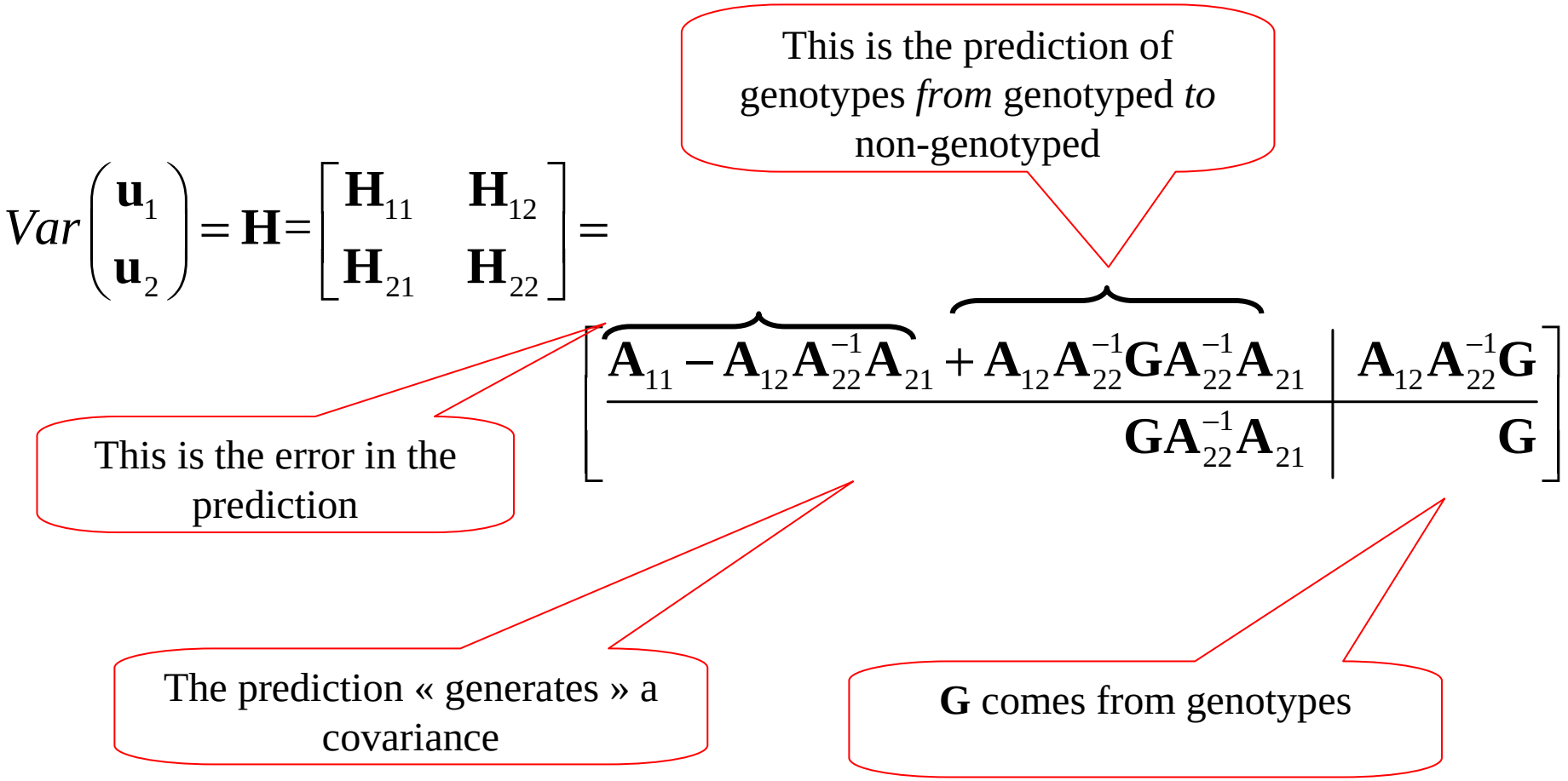
because Cov(**Xa**,**a**) = **X**Var(**a**)

Covariances of all animals

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

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

Covariances of all animals



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

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

- 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…

Single step GBLUP

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

Interpretation of Single step GBLUP

(Aguilar et al., 2010)

- Consider the equation for a genotyped young bull -> one row of $\mathbf{H}^{-1}$

$$-u_s - u_d + 2u_i + \sum_j (g^{ij} - a_{22}^{ij})u_j = 0,$$

$$u_i = \frac{1}{2 + (g^{ii} - a_{22}^{ii})}\left(2\,PA + g^{ii}\,GP - a_{22}^{ii} PP_{22} \right),$$

Parent average

Genomic Predictions from genotyped individuals

Pedigree Predictions from genotyped individuals

Interpretation of Single step GBLUP

(Lawlor, WCGALP)

Braxton, the top genomic tested young bull, better illustrates how certain individuals contribute more or less. …

For seven of his (half) sibs, the genomic relationship with Braxton ranges from 0.55 to 0.93.

Maternal brothers out of **Barbie**			Sires of the bulls **Mates of Barbie**			**Sons** of Cinderella	
	$a^{ij} - g^{ij}$			$a^{ij} - g^{ij}$			$a^{ij} - g^{ij}$
Baltimor	+2.56		Shottle **	+0.84		Cade	+0.54
Bedford	+2.28		Mac	-0.27		Clinto	+0.27
Birk	+1.53		Toystory	-0.97		Capture	+0.09
Blackjack	+1.23		Dundee	-0.81		Casino	+0.47
Bossman	+1.10		Mac	-0.27		Chancellor	+0.72
Chilton	+0.89		Champion	-1.46		Casey	+0.63
Bonzai	+0.53		Dundee	-0.81		Carriage	+0.36

** Sire of Braxton

Large differences

The differences come because we ascertain « realized » instead of « expected » relationhips

example

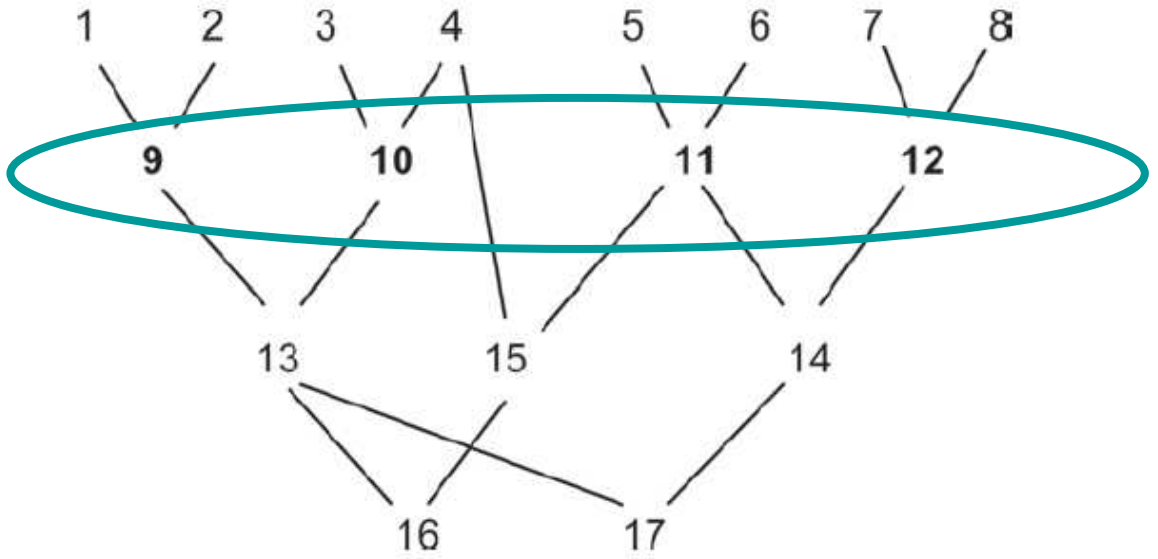


Figure 1. Example pedigree. Genotyped animals are in bold.

example

Table 1. Numerator relationship matrix **A** for the pedigree in Figure 1[1]

1.00								0.50				0.25			0.13	0.13
	1.00							0.50				0.25			0.13	0.13
		1.00							0.50			0.25			0.13	0.13
			1.00						0.50			0.25		0.50	0.38	0.13
				1.00						0.50			0.25	0.25	0.13	0.13
					1.00					0.50			0.25	0.25	0.13	0.13
						1.00					0.50		0.25			0.13
							1.00				0.50		0.25			0.13
0.50	0.50							**1.00**				0.50			0.25	0.25
		0.50	0.50						**1.00**			0.50		0.25	0.38	0.25
				0.50	0.50					**1.00**			0.50	0.50	0.25	0.25
						0.50	0.50				**1.00**		0.50			0.25
0.25	0.25	0.25	0.25					**0.50**	**0.50**			1.00		0.13	0.56	0.50
				0.25	0.25	0.25	0.25			0.50	0.50		1.00	0.25	0.13	0.50
			0.50	0.25	0.25				0.25	0.50		0.13	0.25	1.00	0.56	0.19
0.13	0.13	0.13	0.38	0.13	0.13			0.25	0.38	0.25		0.56	0.13	0.56	1.06	0.34
0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.25	0.25	0.25	0.25	0.50	0.50	0.19	0.34	1.00

[1]Cells with 0 are empty to show the pattern. Coefficients for genotyped animals are in bold. Matrix $\mathbf{A}_g$ is obtained by setting the out-of-diagonal coefficients of genotyped animals to 0.7.

This is the regular relationship matrix. Assume now that animals 9 to 12 have a genomic relationship of 0.7

example

Table 3. Modified relationship matrix **H** including genomic information for genotyped animals and all relatives for the pedigree in Figure 1[1]

1.00		0.18	0.18	0.18	0.18	0.18	0.18	0.50	0.35	0.35	0.35	0.43	0.35	0.26	0.34	0.39
	1.00	0.18	0.18	0.18	0.18	0.18	0.18	0.50	0.35	0.35	0.35	0.43	0.35	0.26	0.34	0.39
0.18	0.18	1.00		0.18	0.18	0.18	0.18	0.35	0.50	0.35	0.35	0.43	0.35	0.18	0.30	0.39
0.18	0.18		1.00	0.18	0.18	0.18	0.18	0.35	0.50	0.35	0.35	0.43	0.35	0.68	0.55	0.39
0.18	0.18	0.18	0.18	1.00		0.18	0.18	0.35	0.35	0.50	0.35	0.35	0.43	0.34	0.34	0.39
0.18	0.18	0.18	0.18		1.00	0.18	0.18	0.35	0.35	0.50	0.35	0.35	0.43	0.34	0.34	0.39
0.18	0.18	0.18	0.18	0.18	0.18	1.00		0.35	0.35	0.35	0.50	0.35	0.43	0.26	0.31	0.39
0.18	0.18	0.18	0.18	0.18	0.18		1.00	0.35	0.35	0.35	0.50	0.35	0.43	0.26	0.31	0.39
0.50	0.50	0.35	0.35	0.35	0.35	0.35	0.35	**1.00**	**0.70**	**0.70**	**0.70**	0.85	0.70	0.53	0.69	0.78
0.35	0.35	0.50	0.50	0.35	0.35	0.35	0.35	**0.70**	**1.00**	**0.70**	**0.70**	0.85	0.70	0.60	0.73	0.78
0.35	0.35	0.35	0.35	0.50	0.50	0.35	0.35	**0.70**	**0.70**	**1.00**	**0.70**	0.70	0.85	0.68	0.69	0.78
0.35	0.35	0.35	0.35	0.35	0.35	0.50	0.50	**0.70**	**0.70**	**0.70**	**1.00**	0.70	0.85	0.53	0.61	0.78
0.43	0.43	0.43	0.43	0.35	0.35	0.35	0.35	**0.85**	**0.85**	**0.70**	**0.70**	1.35	0.70	0.56	0.96	1.03
0.35	0.35	0.35	0.35	0.43	0.43	0.43	0.43	0.70	0.70	0.85	0.85	0.70	1.35	0.60	0.65	1.03
0.26	0.26	0.18	0.68	0.34	0.34	0.26	0.26	0.53	0.60	0.68	0.53	0.56	0.60	1.18	0.87	0.58
0.34	0.34	0.30	0.55	0.34	0.34	0.31	0.31	0.69	0.73	0.69	0.61	0.96	0.65	0.87	1.41	0.80
0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.78	0.78	0.78	0.78	1.03	1.03	0.58	0.80	1.53

[1]Cells with 0 are empty to show the pattern. Coefficients for genotyped animals are in bold.

This parents now are related

G

This guy now is inbred

Computing stuff

- Need to compute $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$, is a challenge.
  - perhaps only in dairy cattle?
- But see Ignacio talk
- Future strategies (Legarra & Ducrocq, JDS, in press)

- Unsymmetric Single Step

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{W}_1 & \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{W}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{W}_1\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{W}_1'\mathbf{R}^{-1}\mathbf{W}_1 + \mathbf{A}^{11}\sigma_u^{-2} & \mathbf{W}_1'\mathbf{R}^{-1}\mathbf{W}_2 + \mathbf{A}^{12}\sigma_u^{-2} & \mathbf{0} & \mathbf{0} \\ \mathbf{W}_2\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{W}_2'\mathbf{R}^{-1}\mathbf{W}_1 + \mathbf{A}^{21}\sigma_u^{-2} & \mathbf{W}_2'\mathbf{R}^{-1}\mathbf{W}_2 + \mathbf{A}^{22}\sigma_u^{-2} & \mathbf{I}\sigma_u^{-2} & -\mathbf{I}\sigma_u^{-2} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_u^{-2} & \mathbf{A}_{22}\sigma_u^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_u^{-2} & \mathbf{0} & \mathbf{G}\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \\ \hat{\boldsymbol{\varphi}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}_2'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

This can be computed efficiently

- Iterative Single Step

1) Solve
$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}_1'\mathbf{W}_1 & \mathbf{X}_2'\mathbf{W}_2 \\ \mathbf{W}_1'\mathbf{X}_1 & \mathbf{W}_1'\mathbf{W}_1 + \alpha_u \mathbf{A}^{11} & \alpha_u \mathbf{A}^{12} \\ \mathbf{W}_2'\mathbf{X}_2 & \alpha_u \mathbf{A}^{21} & \mathbf{W}_2'\mathbf{W}_2 + \alpha_u \mathbf{A}^{22} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}_1'\mathbf{y}_1 \\ \mathbf{W}_2'\mathbf{y}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \alpha_u \hat{\boldsymbol{\varphi}} - \alpha_u \hat{\boldsymbol{\gamma}} \end{bmatrix} \quad (9)$$

for $\mathbf{b}$, $\mathbf{u}_1$ and $\mathbf{u}_2$

2) Solve $\mathbf{A}_{22}\,\hat{\boldsymbol{\varphi}} = \hat{\mathbf{u}}_2$ and $\mathbf{G}\,\hat{\boldsymbol{\gamma}} = \hat{\mathbf{u}}_2$ for $\hat{\boldsymbol{\varphi}}$ *and* $\hat{\boldsymbol{\gamma}}$ (10)

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix}^t = \omega \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix}^* + (1-\omega) \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix}^{t-1}$$

$$\hat{\boldsymbol{\varphi}}^t = \omega \hat{\boldsymbol{\varphi}}^* + (1-\omega)\hat{\boldsymbol{\varphi}}^{t-1} \quad \text{and} \quad \hat{\boldsymbol{\gamma}}^t = \omega \hat{\boldsymbol{\gamma}}^* + (1-\omega)\hat{\boldsymbol{\gamma}}^{t-1}$$

1) regular BLUP
2) genomic predictions based on EBVs
3) avoid double counting based on A_{22}

Problems of SSGBLUP?

- Assumption p($\mathbf{u}_2$)=N(**0**,**G**)
  - Not **0** mean if there is selection (« tuning » solves it: see Vitezica later)
- Same genetic variance in genotyped and ungenotyped animals
  - solved with « tuning »
- Non normality (i.e. major genes)
  - Can solved using **G**=**ZDZ**' with « weights » for SNP (Legarra et al., 2011; Zhang et al., 2010; see GWAS later)
- Assumption that « **A** » is fair. This is false if:
  - pedigree is incorrect
  - distant relationships are too different from reality (Hill & Weir 2010)
- Genetic groups / several breeds
  - Need to modify **H** to include them (work in progress)