EAAP 2009, Session 28 andres.legarra@toulouse.inra.fr

Whole-population relationship matrix including pedigree and markers for genomic selection

Andrés Legarra[1] & Ignacio Aguilar[2,3] & Ignacy Misztal[3] & Dave L Johnson[4]

Rational: use of DYD's or deregressed proofs for genomic selection is problematic

Loss of information

- Bias (selection)
- Tricky covariances and weights, in particular for females (YD's)
- Difficult for many species
- Family information ignored

OBJECTIVE: build up complete relationship matrix

Genomic relationship matrix

- As good as BayesB from experience

If the relationship matrix includes all animals

- Raw records can be used for evaluation
- All animals benefit from increased accuracies from genomic selection
- Avoid selection bias

Conclusion

•Joint relationship matrix **H** gives a unified framework for including all data, pedigree and genotypes

•More elegant, simple and possibly accurate

•Useful for populations where DYDs are inaccurate (sheep, swine…)

Joint relationships

Let u_1 be ungenotyped animals and u_2 be genotyped animals. Let **A** be the numerator relationship matrix.

Let **G** be the *genomic relationship matrix* based on SNP's (VanRaden 2008, Habier et al. 2007)

•**G** contains information on relationships among ancestors and descendants of genotyped animals

•Need to account for **G** in Var ($\mathbf{u}_1$) and in Cov ($\mathbf{u}_1$,$\mathbf{u}_2$)

Genomic information

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

Now, condition ungenotyped animals in genotyped animals by the selection index:

Selection index

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

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

With this, we can create a joint distribution

Joint distribution H

$$\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix} \sim N(0,\mathbf{H}), \mathbf{H} = \begin{bmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{21} & \mathbf{H}_{22}\end{bmatrix}$$

$$\mathbf{H} = \mathbf{A} + \begin{bmatrix}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G}-\mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G}-\mathbf{A}_{22})\\ (\mathbf{G}-\mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G}-\mathbf{A}_{22}\end{bmatrix}$$

Inverse of H

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

Notes:

- **H** = classical relationships + a difference depending on the difference between « numerator » ($\mathbf{A}_{22}$) and « genomic » (**G**) relationship matrices.
- The selection index transmits genomic information in **G** backwards and forwards to all individuals in the pedigree to form **H**
- **H** can be used in genetic evaluations, even of very large size (millions). See Ignacy Misztal (this meeting).

[1]INRA, UR 631, 31326 Castanet-Tolosan, France
andres.legarra@toulouse.inra.fr

[2]INIA, Las Brujas, Uruguay

[3]Animal and Dairy Science, University of Georgia
Athens 30602 USA

[4]LIC, Private Bag 3016, Hamilton, New Zealand