An inversion-free method to compute genomic predictions using an animal model approach

Jørgen Ødegård[1], Per Madsen[2] & Theo Meuwissen[3]
[1]AquaGen AS
[2]Aarhus University
[3]Norwegian University of Life Sciences

One-step genomic evaluation

- Combines the genomic and pedigree-based relationship matrices into a single unified matrix
    - Used in an equivalent animal model
- Phenotypic data from genotyped and ungenotyped animals can be utilized jointly

One-step genomic animal model
(Aguillar et al. 2010, Christensen & Lund 2010)

- Standard Henderson’s mixed model equations (HMME-1)
  - $\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$

    $\lambda = \frac{\sigma_e^2}{\sigma_g^2}$
  - $\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_\omega & \mathbf{G}_\omega \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \\ \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{G}_\omega & \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{G}_\omega \mathbf{A}_{11}^{-1} \mathbf{A}_{12} + \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \end{bmatrix}$
- $\mathbf{G}_{\boldsymbol{\omega}}$ = the genomic relationship matrix (or some function of it)
- The inverse of **H** has a simple structure:
  - $\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{H}^{11} & \mathbf{H}^{12} \\ \mathbf{H}^{21} & \mathbf{H}^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_\omega^{-1} - \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{G}_\omega^{-1} + \mathbf{A}^{0} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}$

    $\mathbf{A}^{0} = \mathbf{A}^{11} - \mathbf{A}_{11}^{-1}$
  - Equations of ungenotyped animals are identical to the pedigree-based equations

Limitations of HMME-1

- Requires $\mathbf{G}_{\omega}^{-1}$
  - Cannot be set up directly
  - May not be of full rank
  - Inversion of big-sized matrices are computationally demanding
    - May eventually be impossible if dimension >100,000
- Requires $\mathbf{A}^{\mathbf{0}}(=\mathbf{A}^{\mathbf{11}}-\mathbf{A}_{\mathbf{11}}^{-\mathbf{1}})$
  - $\mathbf{A}_{\mathbf{11}}^{-\mathbf{1}}$ is hard to compute unless all ancestors are included among the genotyped animals
  - $\mathbf{A}^{\mathbf{0}}$ can be set up directly if ancestors are included among the genotyped

Weighted least squares (WLS) method

- The WLS equations:
- $\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$
- $\hat{\mathbf{g}} = \sigma_g^2\mathbf{H}\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) = \sigma_g^2\mathbf{H}\mathbf{Z}'\hat{\mathbf{s}}$
  - $\mathbf{V} = \mathbf{Z}\mathbf{H}\mathbf{Z}'\sigma_g^2 + \mathbf{I}\sigma_e^2$
  - $\hat{\mathbf{s}} = \mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)$, i.e. $\mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{V}\hat{\mathbf{s}} = \mathbf{y}$
- $\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}\mathbf{H}\mathbf{Z}'\sigma_g^2 \\ \mathbf{X} & \mathbf{Z}\mathbf{H}\mathbf{Z}'\sigma_g^2 + \mathbf{I}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y} \end{bmatrix}$

Alternative Henderson's mixed model equations (HMME-2) (Henderson, 1984)

- $\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{HZ'X} & \mathbf{HZ'Z} + \mathbf{I}\lambda \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{HZ'y} \end{bmatrix}$
- The method does not require the inverse relationship matrix
- The entire **H** must be set up
- Method implemented in the DMU software using both direct and iterative solvers

Hybrid MME (HMME-3)

- The genetic effects are split into genotyped and ungenotyped animals $\hat{\mathbf{g}} = \begin{bmatrix} \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \end{bmatrix}$
    - Equations for genotyped animals are similar to HMME-2
    - Equations for ungenotyped animals are as in HMME-1

- $\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 & \mathbf{X}'\mathbf{Z}_2 \\ \mathbf{G}_\omega \mathbf{Z}_1{}'\mathbf{X} & \mathbf{G}_\omega \mathbf{Z}_1{}'\mathbf{Z}_1 + \left(\mathbf{G}_\omega \mathbf{A}^0 + \mathbf{I}\right)\lambda & \mathbf{G}_\omega \mathbf{A}^{12}\lambda \\ \mathbf{Z}_2{}'\mathbf{X} & \mathbf{A}^{21}\lambda & \mathbf{Z}_2\mathbf{Z}_2 + \mathbf{A}^{22}\lambda \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{g}}_1 \\ \hat{\mathbf{g}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{G}_\omega \mathbf{Z}_1{}'\mathbf{y}_1 \\ \mathbf{Z}_2{}'\mathbf{y}_2 \end{bmatrix}$
    - $\mathbf{A}^0 = \mathbf{A}^{11} - (\mathbf{A}_{11})^{-1}$

- The entire **H** is not needed

Limitations of the one-step approach

- All approaches need $\mathbf{A}_{11}^{-1}$ (or $\mathbf{A}^{0}$)
- Can be set up directly if $\mathbf{H}_{11}$ is expanded with relationships of ungenotyped ancestors

$$\mathbf{H}_{11} = \begin{bmatrix} \mathbf{H}_{00} & \mathbf{H}_{01} \\ \mathbf{H}_{10} & \mathbf{G}_{\boldsymbol{\omega}} \end{bmatrix}$$

How to expand $\mathbf{H}_{11}$?

- Imputation of missing genotypes
    - Not always possible

$$\mathbf{H}_{11} = \begin{bmatrix} \mathbf{G}_{\omega(00)} & \mathbf{G}_{\omega(01)} \\ \mathbf{G}_{\omega(10)} & \mathbf{G}_{\omega(11)} \end{bmatrix}$$

How to expand $\mathbf{H}_{11}$?

- Set up $\mathbf{H}_{11}$ through linkage analysis
  - Pedigree information will be used to incorporate ungenotyped ancestors of genotyped animals
  - Genomic IBD matrix

$$\mathbf{H}_{11} = \begin{bmatrix} \mathbf{G}_{\mathbf{LA(00)}} & \mathbf{G}_{\mathbf{LA(01)}} \\ \mathbf{G}_{\mathbf{LA(10)}} & \mathbf{G}_{\mathbf{LA(11)}} \end{bmatrix}$$

How to expand $\mathbf{H_{11}}$?

- Fill inn ungenotyped ancestors using pedigree relationships
  - Simple, but not necessarily the best

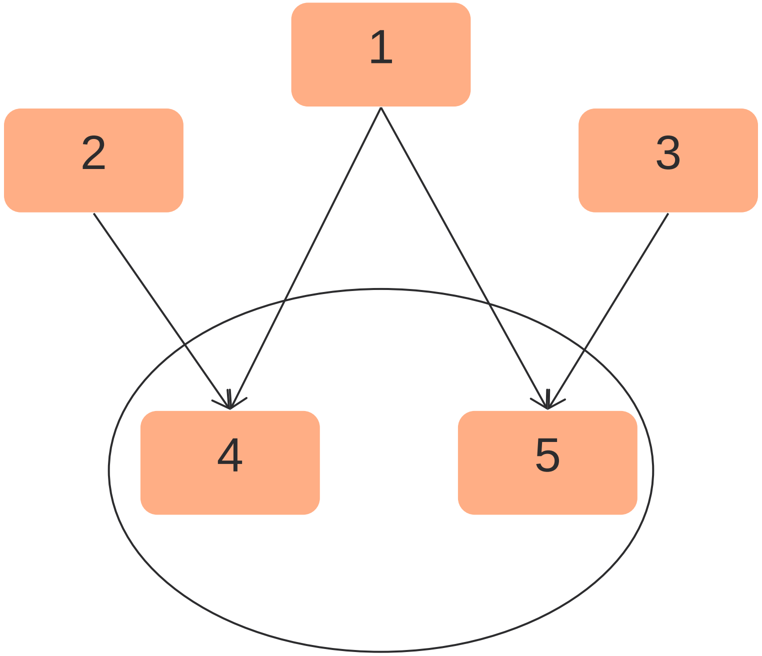
$$\mathbf{H_{11}} = \begin{bmatrix} \mathbf{A_{00}} & \mathbf{A_{01}} \\ \mathbf{A_{10}} & \mathbf{G_{\omega}} \end{bmatrix}$$

Is the one-step approach appropriate?

- Parents genotyped and offspring ungenotyped
  - Infers the most likely relationship structure among ungenotyped offspring
  - Appropriate
- Parents ungenotyped and offspring genotyped
  - The one-step will modify founder relationships such that they explain the genomic relationships
- Co-segregation of loci among relatives will cause deviations from pedigree relationships
  - Occur spontaneously
- The one-step approach explains effects of Mendelian sampling and linkage by relationships between distant founders
  - Unjustified and will likely give bias

Example: Inappropriate one-step analysis

- Assume the following pedigree of two paternal half-sibs



Genotyped

Example: Inappropriate one-step analysis

$$A = \begin{pmatrix} 1 & 0 & 0 & 0.5 & 0.5 \\ 0 & 1 & 0 & 0.5 & 0 \\ 0 & 0 & 1 & 0 & 0.5 \\ 0.5 & 0.5 & 0 & 1 & 0.25 \\ 0.5 & 0 & 0.5 & 0.25 & 1 \end{pmatrix}$$

The paternal half-sibs (4&5) are genotyped:

$$\mathbf{G}_{\omega} = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$$

The slightly increased relationship can be attributed to co-segregation of linked loci. However, the one-step relationship matrix modifies the ancestral relationships

(matrix sorted from oldest to youngest)

$$H = \begin{pmatrix} 1.016 & 0.008 & 0.008 & 0.520 & 0.520 \\ 0.008 & 0.993 & 0.015 & 0.493 & 0.027 \\ 0.008 & 0.015 & 0.993 & 0.027 & 0.493 \\ 0.520 & 0.493 & 0.027 & 1.000 & 0.300 \\ 0.520 & 0.027 & 0.493 & 0.300 & 1.000 \end{pmatrix}$$

AquaGen

Implications

- Solving large-scale MME without inverting $\mathbf{G}_{\omega}$ is possible
- Expanding $\mathbf{G}_{\omega}$ with ungenotyped ancestors may be needed
  - For solving large-scale MME (using HMME-3)
  - For obtaining proper predictions of genetic effects
- Linkage analysis may be used to construct $\mathbf{G}_{\omega}$ including ungenotyped ancestors