Compatibility of pedigree-based and marker-based relationships for single-step genomic prediction

Ole F. Christensen

Aarhus University, Center for Quantitative Genetics and Genomics

EAAP 2012, Bratislava

Single-step genetic evaluation

- Combines phenotypes, genomic and pedigree information using a combined relationship matrix (Misztal, Legarra, Aguilar + coworkers, Christensen and Lund).

- Inverse of this matrix

$$H^{-1} = \begin{bmatrix} G^{-1} - A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + A^{-1}$$

- A_{11} and G need to be "compatible".

- Aim here: Provide explanation; Show a possible way to handle it.

The idea behind single-step methods

- Two types of data: phenotypes y and markers m (-1, 0, 1)
- Some animals genotyped (m^{obs}) but others are not (m^{miss}).
- This is a "missing data problem" !
- A model is specified for the "full data": $f(y, m^{obs}, m^{miss})$
- Marginalisation:

$$f(y, m^{obs}) = \int f(y, m^{obs}, m^{miss}) dm^{miss}$$

- $f(y, m^{obs})$ should be used for inference.

The model behind single-step methods

- Phenotypes conditional on all markers m:

$$y = \mu + a + e$$

where $a \sim N(0, \sigma_a^2 G(m))$ with

$$G(m) = \sum_j (m_j - (2\rho_j - 1))(m_j - (2\rho_j - 1))^T / s$$

and ρ_j's are allele frequencies.

- Markers: $m_j \sim N((2\rho_j - 1)1, v_j A)$

The model behind single-step methods

By marginalisation (integrating m^{miss})

- Phenotypes conditional on observed markers m^{obs}:

$$y = \mu + a + e$$

where $\text{Var}(a) = \sigma_a^2 H$ with

$$H^{-1} = \begin{bmatrix} G(m^{obs})^{-1} - A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + A^{-1}$$

and

$$G(m^{obs}) = \sum_j (m_j^{obs} - (2\rho_j - 1))(m_j^{obs} - (2\rho_j - 1))^T / s$$

- Observed markers: $m_j^{obs} \sim N((2\rho_j - 1)1, v_j A_{11})$

Compatibility issue

- $G(m^{obs})$ and A_{11} need to be "compatible"
- Allele frequencies ρ_j and scaling $s = \sum_j v_j$ used to make compatible.
- log-Likelihood for parameter estimation:

$$\ell_{y,m^{obs}}(\sigma_a^2, \sigma_e^2, \rho, s) = \ell_{y|m^{obs}}(\sigma_a^2, \sigma_e^2, \rho, s) + \ell_{m^{obs}}(\rho, v)$$

- Allele frequencies ρ enter into both terms !, but maximising $\ell_{y,m^{obs}}$ numerically is not feasible computationally.
- Various adjustments of $G(m^{obs})$ used in practice.

Alternative approach: adjusting A instead

- Phenotypes conditional on observed markers m^{obs}:

$$y = \mu + a + e$$

where $\text{Var}(a) = \sigma_a^2 \tilde{H}$ with

$$\tilde{H}^{-1} = \begin{bmatrix} \tilde{G} - (\tilde{A}_{11}(\alpha))^{-1} & 0 \\ 0 & 0 \end{bmatrix} + (\tilde{A}(\alpha))^{-1}$$

with

$$\tilde{G} = \sum_j (m_j^{obs})(m_j^{obs})^T / \tilde{s}$$

- Observed markers: $m_j^{obs} \sim N(0, (\tilde{s}/p)\tilde{A}_{11}(\alpha))$

Relationship matrix $\tilde{A}(\alpha)$

- Founders in the pedigree are related (coefficient=α) and inbreed (coefficient=$\alpha/2$).

- $\tilde{A}(\alpha)$ is defined recursively in the usual way.

- Inverse:

$$\tilde{A}^{-1} = (T^{-1})^{\mathrm{T}} \begin{bmatrix} (\tilde{A}_0)^{-1} & 0 \\ 0 & \tilde{D}^{-1} \end{bmatrix} T^{-1}.$$

- Colleau algorithm for computing $A_{11}(\alpha)$ also exist.

- Fast computing procedure still exists !

Compatibility issue when adjusting A

- $\tilde{G} = \sum_j (m_j^{obs})(m_j^{obs})^T / \tilde{s}$ and $\tilde{A}_{11}(\alpha)$ need to be "compatible"

- Two parameters, α and scaling parameter $\tilde{s}$ used to make compatible.

- log-Likelihood for parameter estimation:

$$\ell_{y,m^{obs}}(\sigma_a^2, \sigma_e^2, \alpha, \tilde{s}) = \ell_{y|m^{obs}}(\sigma_a^2, \sigma_e^2, \alpha, \tilde{s}) + \ell_{m^{obs}}(\alpha, \tilde{s})$$

- Parameters α and $\tilde{s}$ enter into both terms, and maximising $\ell_{y,m^{obs}}$ numerically is computationally feasible.

Conclusion

- Compatibility of G and A: The meaning is that certain parameters should be fitted to data (in theory both phenotypes and observed markers).

- An approach where A contains parameter(s) provides an interesting alternative.