Genomic evaluation for both purebred and crossbred performance

Ole F. Christensen

Aarhus University, Center for Quantitative Genetics and Genomics

GenSAP meeting 2014 Korsør, day2

Motivation (pigs)

• Genomic selection has been implemented for purebreds.

- Crossbreeding is predominant in the production.
- Ideally, genomic evaluation should include records on crossbreds, and breeding values for crossbred performance should be presented.
- Aim here: develop methods
 - Two-breed terminal crossbreding
 - Single-step method (pedigree and genomic)

Outline

- Genomic model for purebreds and crossbreds.
- Pedigree-based model for purebreds and crossbreds.
- Single-step method for purebreds and crossbreds
- Future work

Genomic model for purebreds and crossbreds

Purebreds

$$\begin{split} \mathbf{y}_{\mathcal{A}} &= \mu_{\mathcal{A}} \mathbf{1} + \mathbf{m}_{\mathbf{A}} \boldsymbol{\beta}_{\mathcal{A}} + \mathbf{e}_{\mathcal{A}}, \\ \mathbf{y}_{\mathcal{B}} &= \mu_{\mathcal{B}} \mathbf{1} + \mathbf{m}_{\mathbf{B}} \boldsymbol{\beta}_{\mathcal{B}} + \mathbf{e}_{\mathcal{B}}, \end{split}$$
where $\mathbf{m}_{ij}^{\mathcal{A}} = 0, 1, 2$ for genotypes 11, 12, 22 (same for $\mathbf{m}_{ij}^{\mathcal{B}}$).

$$\mathbf{y}_{\mathcal{A}\mathcal{B}} = \mu_{\mathcal{A}\mathcal{B}}\mathbf{1} + \mathbf{q}_{\mathcal{A}}\boldsymbol{\beta}_{\mathcal{A}}^{c} + \mathbf{q}_{\mathcal{B}}\boldsymbol{\beta}_{\mathcal{B}}^{c} + \mathbf{e}_{\mathcal{A}\mathcal{B}}$$

where $\mathbf{q}_{ij}^{\mathcal{A}} = 0, 1$ for haplotypes 1, 2 (same for $\mathbf{q}_{ij}^{\mathcal{B}}$).

Genotypes phased to breed of origin.

Genomic model for purebreds and crossbreds

$$\begin{aligned} \mathbf{y}_{\mathcal{A}} &= \mu_{\mathcal{A}} \mathbf{1} + \mathbf{a}_{\mathcal{A}} + \mathbf{e}_{\mathcal{A}}, \\ \mathbf{y}_{\mathcal{B}} &= \mu_{\mathcal{B}} \mathbf{1} + \mathbf{a}_{\mathcal{B}} + \mathbf{e}_{\mathcal{B}}, \\ \mathbf{y}_{\mathcal{A}\mathcal{B}} &= \mu_{\mathcal{A}\mathcal{B}} \mathbf{1} + \mathbf{c}_{\mathcal{A}\mathcal{B}}^{\mathcal{A}} + \mathbf{c}_{\mathcal{A}\mathcal{B}}^{\mathcal{B}} + \mathbf{e}_{\mathcal{A}\mathcal{B}} \end{aligned}$$

► BVs for purebred performance: $\mathbf{a}_{\mathcal{A}} = \mathbf{m}_{\mathbf{A}} \boldsymbol{\beta}_{\mathcal{A}}$, $\mathbf{a}_{\mathcal{B}} = \mathbf{m}_{\mathbf{B}} \boldsymbol{\beta}_{\mathcal{B}}$

• Partial genetic effects:
$$\mathbf{c}_{\mathcal{A}\mathcal{B}}^{\mathcal{A}} = \mathbf{q}_{\mathcal{A}}\beta_{\mathcal{A}}^{c}$$
, $\mathbf{c}_{\mathcal{A}\mathcal{B}}^{\mathcal{B}} = \mathbf{q}_{\mathbf{B}}\beta_{\mathcal{B}}^{c}$.

► BVs for crossbred performance: $\mathbf{c}_{\mathcal{A}} = \mathbf{m}_{\mathcal{A}} \beta_{\mathcal{A}}^{c}$, $\mathbf{c}_{\mathcal{B}} = \mathbf{m}_{\mathcal{B}} \beta_{\mathcal{B}}^{c}$

 Equiv. model: 2 Genomic breed specific "partial" relationship matrices, G^(A), G^(B).

Genomic breed \mathcal{A} partial relationship matrix

$$\mathbf{G}^{(\mathcal{A})} = \begin{bmatrix} \mathbf{G}^{(\mathcal{A})}_{\mathcal{A},\mathcal{A}} & \mathbf{G}^{(\mathcal{A})}_{\mathcal{A},\mathcal{A}\mathcal{B}} \\ \mathbf{G}^{(\mathcal{A})}_{\mathcal{A}\mathcal{B},\mathcal{A}} & \mathbf{G}^{(\mathcal{A})}_{\mathcal{A}\mathcal{B},\mathcal{A}\mathcal{B}} \end{bmatrix}$$

$$\begin{split} \mathbf{G}_{\mathcal{A},\mathcal{A}}^{(\mathcal{A})} &= \frac{(\mathbf{m}^{\mathcal{A}} - 2\bar{\rho}^{\mathcal{A}}\mathbf{1}^{\mathrm{T}})(\mathbf{m}^{\mathcal{A}} - 2\bar{\rho}^{\mathcal{A}}\mathbf{1}^{\mathrm{T}})^{\mathrm{T}}}{s^{(\mathcal{A})}}, \\ \mathbf{G}_{\mathcal{A},\mathcal{A}\mathcal{B}}^{(\mathcal{A})} &= \frac{(\mathbf{m}^{\mathcal{A}} - 2\bar{\rho}^{\mathcal{A}} - \mathbf{1}^{\mathrm{T}})(\mathbf{q}^{\mathcal{A}} - \bar{\rho}^{\mathcal{A}}\mathbf{1}^{\mathrm{T}})^{\mathrm{T}}}{s^{(\mathcal{A})}}, \\ \mathbf{G}_{\mathcal{A}\mathcal{B},\mathcal{A}\mathcal{B}}^{(\mathcal{A})} &= \frac{(\mathbf{q}^{\mathcal{A}} - \bar{\rho}^{\mathcal{A}}\mathbf{1}^{\mathrm{T}})(\mathbf{q}^{\mathcal{A}} - \bar{\rho}^{\mathcal{A}}\mathbf{1}^{\mathrm{T}})^{\mathrm{T}}}{s^{(\mathcal{A})}}. \end{split}$$

- ▶ For breed \mathcal{A} : $\mathbf{m}_{ij}^{\mathcal{A}} = 0, 1, 2$ for genotypes 11, 12, 22
- For crossbreds: $\mathbf{q}_{ii}^{\mathcal{A}} = 0, 1$ for haplotypes 1, 2.
- Allele frequency: $\bar{\rho}_j$.

Pedigree-based model

(Wei and van der Werf, 1994)

$$\begin{split} \mathbf{y}_{\mathcal{A}} &= \mu_{\mathcal{A}} \mathbf{1} + \mathbf{Z}_{\mathcal{A}} \mathbf{a}_{\mathcal{A}} + \mathbf{e}_{\mathcal{A}}, \\ \mathbf{y}_{\mathcal{B}} &= \mu_{\mathcal{B}} \mathbf{1} + \mathbf{Z}_{\mathcal{B}} \mathbf{a}_{\mathcal{B}} + \mathbf{e}_{\mathcal{B}}, \\ \mathbf{y}_{\mathcal{A}\mathcal{B}} &= \mu_{\mathcal{A}\mathcal{B}} \mathbf{1} + \mathbf{c}_{\mathcal{A}\mathcal{B}}^{\mathcal{A}} + \mathbf{c}_{\mathcal{A}\mathcal{B}}^{\mathcal{B}} + \mathbf{e}_{\mathcal{A}\mathcal{B}}, \end{split}$$

where

- BVs for purebred performance: $\mathbf{a}_{\mathcal{A}}$, $\mathbf{a}_{\mathcal{B}}$.
- Breed of origin specific genetic effects: $\mathbf{c}_{\mathcal{AB}}^{\mathcal{A}}$, $\mathbf{c}_{\mathcal{AB}}^{\mathcal{B}}$,
- BVs for crossbred performance: $\mathbf{c}_{\mathcal{A}}$, $\mathbf{c}_{\mathcal{B}}$.

Pedigree-based model

- ► BVs for crossbred performance c_A are correlated with BVs for purebred performance a_A (genetic correlation) and with partial genetic effects c^(A)_{AB} (partial pedigree relationships).
- Supports a breeding goal with both purebred and crossbred performance.
- Allows different genetic variances in two breeds.
- ► c_A, c_B captures dominance effects, different genetic background, G × E

Genetic covariances

Breed
$$\mathcal{A}$$
:

$$\operatorname{Var} \begin{bmatrix} \mathbf{a}_{\mathcal{A}}^{(\mathcal{A})} \\ \star \\ \mathbf{c}_{\mathcal{A}} \\ \mathbf{c}_{\mathcal{AB}}^{(\mathcal{A})} \end{bmatrix} = \mathbf{\Sigma}_{g}^{(\mathcal{A})} \bigotimes \mathbf{A}^{(\mathcal{A})}$$

where $\mathbf{A}^{(\mathcal{A})}$ is partial relationship matrix for breed \mathcal{A} (Garcia-Cortes and Toro, 2006).

• Breed \mathcal{B} is similar.

• Breed A and breed B specific genetic effects are independent.

Partial relationship matrix $\mathbf{A}^{(\mathcal{A})}$

Recursions (Garcia-Cortes, 1996) :

$$\mathbf{A}_{i,i'}^{(\mathcal{A})} = (\mathbf{A}_{f(i),i'}^{(\mathcal{A})} + \mathbf{A}_{m(i),i'}^{(\mathcal{A})})/2$$
$$\mathbf{A}_{i,i}^{(\mathcal{A})} = f_i^{\mathcal{A}} + (\mathbf{A}_{f(i),i'}^{(\mathcal{A})} + \mathbf{A}_{m(i),i'}^{(\mathcal{A})})/2$$

where $f_i^{\mathcal{A}} = 1, 0, 1/2$ for breed \mathcal{A} , breed \mathcal{B} , and crossbred \mathcal{AB} .

- Ignore breed B animals in relationship matrix.
- Sparse inverse can be computed directly.
- Submatrix can be computed using Colleau (2002) algorithm.

Combined relationship matrix

(pedigree and genomic)

Combined partial relationship matrix:

$$(\mathbf{H}^{(\mathcal{A})})^{-1} = \begin{bmatrix} (\mathbf{G}^{(\mathcal{A})})^{-1} - (\mathbf{A}_{11}^{(\mathcal{A})})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + (\mathbf{A}^{(\mathcal{A})})^{-1}, \quad (1)$$

► Sparse inverse (**A**^(A))⁻¹ computed directly.

Single-step method

- Partial relationship matrices based on pedigree are replaced by combined partial relationship matrices.
- ► Can be fitted using standard software for multi-trait genetic evaluation using (H^(A))⁻¹ and (H^(B))⁻¹ as input.
- Compatibility of partial G and A: Linear adjustment of partial G towards partial A.
- Assumption: marker-data for crossbred animals can be phased to breed of origin.
- Method is available!

Acknowledgements

- ► Co-authors: Per Madsen, Bjarne Nielsen, Guosheng Su
- Project: "Developing genomic selection for a pig breeding system based on crossbreeding"
- Funding: through "Green Development and Demonstration Programme" (GUDP) by Danish Ministry of Food, Agriculture and Fisheries, Pig Research Centre and Aarhus University.

Future work

 PhD student Tao Xiang will use the model on fertility trait in Landrace (L), Yorkshire (Y) and crossbred (LY).

Questions of interest to investigate:

- Genetic correlation between purebred and crossbred performance.
- Accuracy of breeding values for crossbred performance.
- Extension to three-breed terminal crossbreeding is possible.

Alternative: relationships across breeds

Alternative is to assume that base individuals in A and B are related (and inbreed) and breeds are related. Relationships in base populations determined by

$$\Gamma = \left[\begin{array}{cc} \gamma_{\mathcal{A}} & \gamma_{\mathcal{A},\mathcal{B}} \\ \gamma_{\mathcal{A},\mathcal{B}} & \gamma_{\mathcal{B}} \end{array} \right].$$

- Additive relationship matrix A(Γ)
- ▶ Marker-based relationship matrix **G**_{0.5} (all allele freq are 0.5)
- Estimate Γ from G_{0.5}
- Combined relationship matrix $(\mathbf{H})^{-1} = \begin{bmatrix} (\mathbf{G}_{0.5})^{-1} - (\mathbf{A}(\Gamma)_{11})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + (\mathbf{A}(\Gamma))^{-1}.$

Relationships across breeds

Inverse of $\mathbf{A}(\Gamma)$ computing procedure:

 \blacktriangleright Add $\mathcal{A},\ \mathcal{B}$ as two "metafounders" to the pedigree.

$$\mathbf{A}(\Gamma)^{-1} = (\mathbf{T}^{-1})^{\mathrm{T}} \begin{bmatrix} (\Gamma)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \mathbf{T}^{-1}$$

► Read Γ^{-1} .

Remaining part of procedure is as usual.

Legarra, Christensen, Vitezica, Aguilar, Misztal (2014). *Genetics* submitted.

Funding: GenSAP