The evolution of genetic evaluation: Fisher to Visscher

Rohan Fernando

September, 2018

Iowa State University

The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edin (1918) 53:399-433.

Two-way Factor Experiment

Data:

$$
\begin{array}{cc}
A_1 & B_1 \\
A_1 & B_2 \\
A_2 & B_1 \\
A_2 & B_2 \\
\vdots &
\end{array}
$$

Two-way Factor Experiment

Data:

$$
\begin{array}{cc}
A_1 & B_1 \\
A_1 & B_2 \\
A_2 & B_1 \\
A_2 & B_2 \\
& \vdots
\end{array}
$$

Model:

$$y_{ijk} = \mu + A_i + B_j + e_{ijk}$$

Two-way Factor Experiment

In Matrix notation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Least Squares:

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

Model for multifactorial inheritance:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Q}\alpha + \mathbf{e}$$

The correlation between relatives on the supposition of Mendelian inheritance

Model for multifactorial inheritance:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Q}\alpha + \mathbf{e}$$

But, $\mathbf{Q}$ (matrix of genotype covariates) is not observed!

The correlation between relatives on the supposition of Mendelian inheritance

Model for multifactorial inheritance:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Q}\alpha + \mathbf{e}$$

But, $\mathbf{Q}$ (matrix of genotype covariates) is not observed!

Relatives are expected to have similar genotypes:

The correlation between relatives on the supposition of Mendelian inheritance

Model for multifactorial inheritance:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\boldsymbol{\alpha} + \mathbf{e}$$

But, $\mathbf{Q}$ (matrix of genotype covariates) is not observed!

Relatives are expected to have similar genotypes:

So, covariances of genotypic values, $\mathbf{a} = \mathbf{Q}\boldsymbol{\alpha}$, can be expressed as

$$\mathrm{Var}(\mathbf{a}) = \mathbf{A}\sigma_a^2$$

The correlation between relatives on the supposition of Mendelian inheritance

Model for multifactorial inheritance:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\boldsymbol{\alpha} + \mathbf{e}$$

But, $\mathbf{Q}$ (matrix of genotype covariates) is not observed!

Relatives are expected to have similar genotypes:

So, covariances of genotypic values, $\mathbf{a} = \mathbf{Q}\boldsymbol{\alpha}$, can be expressed as

$$\text{Var}(\mathbf{a}) = \mathbf{A}\sigma_a^2$$

Fisher also considered non-additive inheritance and assortative mating.

Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

BLUP: Henderson's MME

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Paradigm Shift

Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:181929.

Genetic Evaluation: Genomic

Model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{M}\alpha + \mathbf{e}$$

Model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{M}\alpha + \mathbf{e}$$

Often, $p > n$

Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\alpha} + \mathbf{e}$$

Often, $p > n$

Prior information (Meuwissen et al. 2001): $\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}\sigma_\alpha^2)$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{M} \\ \mathbf{M}'\mathbf{X} & \mathbf{M}'\mathbf{M} + \mathbf{I}\frac{\sigma_e^2}{\sigma_\alpha^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{M}'\mathbf{y} \end{bmatrix}$$

Model:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_p \end{bmatrix} \beta + \begin{bmatrix} \mathbf{M} \\ \mathbf{M}_p \end{bmatrix} \alpha + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_p \end{bmatrix}$$

Least Squares with Additional Information

Model:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_p \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{M} \\ \mathbf{M}_p \end{bmatrix} \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_p \end{bmatrix}$$

Least Squares:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X}\frac{1}{\sigma_e^2} + \mathbf{X}'_p\mathbf{X}_p\frac{1}{\sigma_\alpha^2} & \mathbf{X}'\mathbf{M}\frac{1}{\sigma_e^2} + \mathbf{X}'_p\mathbf{M}_p\frac{1}{\sigma_\alpha^2} \\ \mathbf{M}'\mathbf{X}\frac{1}{\sigma_e^2} + \mathbf{M}'_p\mathbf{X}_p\frac{1}{\sigma_\alpha^2} & \mathbf{M}'\mathbf{M}\frac{1}{\sigma_e^2} + \mathbf{M}'_p\mathbf{M}_p\frac{1}{\sigma_\alpha^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y}\frac{1}{\sigma_e^2} + \mathbf{X}'_p\mathbf{y}_p\frac{1}{\sigma_\alpha^2} \\ \mathbf{M}'\mathbf{y}\frac{1}{\sigma_e^2} + \mathbf{M}'_p\mathbf{y}_p\frac{1}{\sigma_\alpha^2} \end{bmatrix}$$

Least Squares with Additional Information

Model:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_p \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{M} \\ \mathbf{M}_p \end{bmatrix} \boldsymbol{\alpha} + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_p \end{bmatrix}$$

Least Squares:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X}\frac{1}{\sigma_e^2} + \mathbf{X}_p'\mathbf{X}_p\frac{1}{\sigma_\alpha^2} & \mathbf{X}'\mathbf{M}\frac{1}{\sigma_e^2} + \mathbf{X}_p'\mathbf{M}_p\frac{1}{\sigma_\alpha^2} \\ \mathbf{M}'\mathbf{X}\frac{1}{\sigma_e^2} + \mathbf{M}_p'\mathbf{X}_p\frac{1}{\sigma_\alpha^2} & \mathbf{M}'\mathbf{M}\frac{1}{\sigma_e^2} + \mathbf{M}_p'\mathbf{M}_p\frac{1}{\sigma_\alpha^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y}\frac{1}{\sigma_e^2} + \mathbf{X}_p'\mathbf{y}_p\frac{1}{\sigma_\alpha^2} \\ \mathbf{M}'\mathbf{y}\frac{1}{\sigma_e^2} + \mathbf{M}_p'\mathbf{y}_p\frac{1}{\sigma_\alpha^2} \end{bmatrix}$$

Assumption of $\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}\sigma_\alpha^2)$ equivalent to: $\mathbf{X}_p = \mathbf{0}$, $\mathbf{M}_p = \mathbf{I}$, and $\mathbf{y}_p = \mathbf{0}$,

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'M} \\ \mathbf{M'X} & \mathbf{M'M} + \mathbf{I}\frac{\sigma_e^2}{\sigma_\alpha^2} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{M'y} \end{bmatrix}$$

Identical to the MME under the assumption: $\alpha \sim N(\mathbf{0}, \mathbf{I}\sigma_\alpha^2)$

Relationships Between Individuals: Pedigree

Consider data from halfsib groups

Sire model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}_s + \epsilon$$

Relationships Between Individuals: Pedigree

Consider data from halfsib groups

Sire model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}_s + \boldsymbol{\epsilon}$$

BLUP:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{I}\lambda_s \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}}_s \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$\lambda_s = \frac{\sigma_\epsilon^2}{\sigma_s^2}, \quad \sigma_s^2 = \sigma_a^2/4, \quad \text{and} \quad \sigma_\epsilon^2 = 3/4\sigma_a^2 + \sigma_e^2$

Relationships Between Individuals: Pedigree

Breeding value model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \mathbf{e}$$

Relationships Between Individuals: Pedigree

Breeding value model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{a} + \mathbf{e}$$

$$\mathbf{a} = \mathbf{Z}\mathbf{u}_s + \mathbf{u}_r$$

$$\text{Var}(\mathbf{a}) = \mathbf{Z}\mathbf{Z}'\frac{1}{4}\sigma_a^2 + \mathbf{I}\frac{3}{4}\sigma_a^2 = \mathbf{A}\sigma_a^2$$

BLUP:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Can show:

$$\hat{\mathbf{a}} = \mathbf{Z}\hat{\mathbf{u}}_s + (\mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\mathbf{u}}_s)/(1 + \frac{\sigma_e^2}{3/4\sigma_a^2})$$

Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\alpha} + \mathbf{e}$$
$$= \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \mathbf{e}$$

where $\mathbf{a} = \mathbf{M}\boldsymbol{\alpha}$, and $\text{Var}(\mathbf{a}) = \mathbf{M}\mathbf{M}'\sigma_\alpha^2 = \mathbf{G}\sigma_\alpha^2$ (Habier et al. 2007).

Model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\alpha} + \mathbf{e}$$
$$= \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \mathbf{e}$$

where $\mathbf{a} = \mathbf{M}\boldsymbol{\alpha}$, and $\text{Var}(\mathbf{a}) = \mathbf{M}\mathbf{M}'\sigma_\alpha^2 = \mathbf{G}\sigma_\alpha^2$ (Habier et al. 2007). BLUP:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}' \\ \mathbf{X} & \mathbf{I} + \mathbf{G}^{-1}\frac{\sigma_e^2}{\sigma_\alpha^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Linkage Equilibrium: $\Pr(Q_1|A_1) = \Pr(Q_1|A_2)$: marker information has to be combined with pedigree (Chevalet, 1984)

Linkage Equilibrium: $\Pr(Q_1|A_1) = \Pr(Q_1|A_2)$: marker information has to be combined with pedigree (Chevalet, 1984)

Linkage Disequilibrium: $\Pr(Q_1|A_1) \neq \Pr(Q_1|A_2)$: marker information can be used without pedigree (Haley and Knott, 1992)

Relationships Between Loci: Linkage

Linkage Equilibrium: $\Pr(Q_1|A_1) = \Pr(Q_1|A_2)$: marker information has to be combined with pedigree (Chevalet, 1984)

Linkage Disequilibrium: $\Pr(Q_1|A_1) \neq \Pr(Q_1|A_2)$: marker information can be used without pedigree (Haley and Knott, 1992)

When multiple regression is used, markers provide linkage information (Habier et al. 2013)

Consider a genomic segment of 1 cM

Linkage Signal from Multiple Regression

Consider a genomic segment of 1 cM

Conditional only on pedigree, additive relationship between halfsibs is 0.25.

Linkage Signal from Multiple Regression

Consider a genomic segment of 1 cM

Conditional only on pedigree, additive relationship between halfsibs is 0.25.
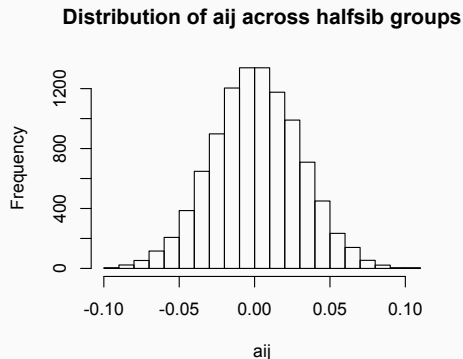
The relationship computed from genotypes has a bimodal distribution that is centered at 0.25.

Distribution of aij within halfsib groups
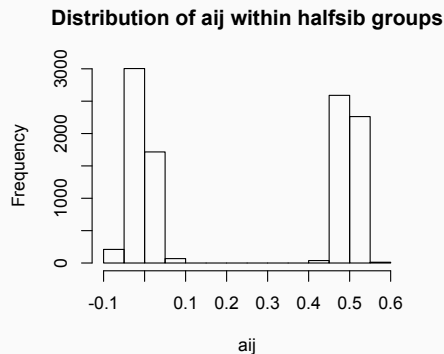
Figure 1: *Distribution of genomic relationship coefficients within a halfsib group of size 100, computed using 1000 SNP marker genotypes in a 1 cM interval.*

Distribution of aij across halfsib groups

Figure 2: *Distribution of genomic relationship coefficients between two halfsib groups of size 100, computed using 1000 SNP marker genotypes in a 1 cM interval.*
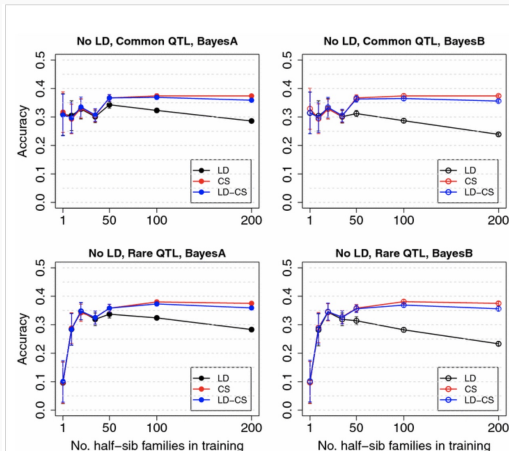
Fig. 2

Mean accuracy with different numbers of half-sib families in training with no historical LD. LD, the LD model; CS, the CS model; LD-CS, the LD-CS model. *Top panel*, the Common QTL scenario, *Bottom panel*, the Rare QTL scenario. *Left panel*, BayesA, *right panel*, BayesB

- Fisher could not observe $\mathbf{Q}$, but took advantage of expected similarity between rows of $\mathbf{Q}$.

Summary

- Fisher could not observe **Q**, but took advantage of expected similarity between rows of **Q**.
- In genomic evaluation, can observe **M**, but $p > n$
  - treat marker effects as random
  - relationships between and within individuals

Summary

- Fisher could not observe $\mathbf{Q}$, but took advantage of expected similarity between rows of $\mathbf{Q}$.
- In genomic evaluation, can observe $\mathbf{M}$, but $p > n$
  - treat marker effects as random
  - relationships between and within individuals
- Explicit modeling of linkage can improve accuracy