Using theoretical and realized accuracies to estimate changes in heritabilities and genetic correlations

Ignacy Misztal
University of Georgia
Changes under genomic selection

• Under selection:
  – Heritabilities decline
  – Antagonistic correlations intensify

• Accelerated changes under genomic selection
  – Heritability halved for growth, negative correlations almost doubled (Hidalgo et al, 2019)

• With wrong parameters
  – Selection index inaccurate
  – Genetic gain not as expected
  – Possible deterioration of fitness traits

• How to estimate parameter change for last generation/year with commercial data sets?
How to estimate parameter change by generation?

- If REML/Gibss sampling
  - Base population parameters
  - Impossible computations with too many genotypes
  - Biases with too few generations (Cesarani et al., 2019)

Data sampling in Hidalgo et al.(2019)

Months of computing

Base population parameters
Requirements for new estimation

• All data including genomic
• Parameters by last generation/year
• Reasonable computing cost
Realized and theoretical accuracies

**Realized accuracy** \( acc = \frac{corr (y - Xb, \hat{u})}{h} \)

Legarra et al. (2008)
- \( y - Xb \) - adjusted phenotype
- \( \hat{u} \) - breeding value obtained without that phenotype
- \( h^2 \) - heritability

**Theoretical accuracy** \( acc = \sqrt{\frac{Nh^2}{Nh^2 + Me}} \)

Daetwyler et al. (2008)
- \( N \) – number of genotyped animals with phenotypes
- \( Me \) – number of independent chromosome segments
- \( Me \approx 5k \) (chickens, pigs), 10k (beef), 15k (Holsteins)

Pocrnic et al. (2017)
Derivations of realized accuracies

\[ acc = \text{corr}(u, \hat{u}) = \frac{\text{cov}(u, \hat{u})}{\sqrt{\sigma_u^2 \sigma_{\hat{u}}^2}} \]

\[ y = Xb + u + e, \quad y - Xb = u + e \]

\[ \text{corr}(y - Xb, \hat{u}) = \frac{\text{cov}(u + e, \hat{u})}{\sqrt{\sigma_u^2 \sigma_{\hat{u}}^2}} = \frac{\text{cov}(u, \hat{u})}{\sqrt{\sigma_u^2 \sigma_{\hat{u}}^2}} = \frac{\text{cov}(u, \hat{u})}{\sqrt{\sigma_u^2 \sigma_{\hat{u}}^2}} = \frac{\text{cov}(u, \hat{u})}{\sqrt{\sigma_u^2 \sigma_{\hat{u}}^2}} = \frac{\text{cov}(u, \hat{u})}{\sqrt{\sigma_u^2 \sigma_{\hat{u}}^2}} \]

\[ = \frac{\text{cov}(u, \hat{u})}{\sqrt{\frac{1}{h^2} \sigma_u^2 \sigma_{\hat{u}}^2}} = acc \ h \]
Pig data set

150k records on growth (h^2=0.21)
25k records on fitness (h^2=0.05)

53k genotyped animals

Theoretical accuracy
\[
\sqrt{\frac{53k \times 0.21}{53k \times 0.21 + 5k}} = 0.83
\]

Realized accuracy
0.82

Hollifield et al., 2021

\[
\sqrt{\frac{25k \times 0.05}{25k \times 0.05 + 5k}} = 0.44
\]

0.41
Broiler chicken data set

820k phenotypes for growth  
\( h^2 = 0.3 \)  
150k genotyped

Theoretical accuracy  
\[
\sqrt{\frac{150k \times 0.3}{150k \times 0.3 + 5k}} = 0.94
\]

Realized accuracy  
0.58

Both accuracy same 0.89 if \( h^2 = 0.13 \)

Company was using 0.14! (Breen, 2022)
Formulas for estimating heritability

\[ \hat{h}^2: \left( Nh^2 \right) \left( Nh^2 + M_e \right) = \text{corr} \left( y - Xb, \hat{u} \right) / h \]

\[ \hat{h}^2 = \frac{c^2 + \sqrt{c^4 + 4c^2M_e/N}}{2} \]

\[ c = \text{corr}(y - Xb, \hat{u}) \]

\[ SE(\hat{h}^2) \approx \frac{1}{\sqrt{N_{\text{val}}}} \left[ c + \frac{2c^2 + 4M_e/N}{\sqrt{c^2 + 4M_e/N}} \right] \hat{h}^2 \approx \frac{3c}{\sqrt{N_{\text{val}}}} \]

N – # animals in reference  N_{val} – number of animals in validation
Heritability for milk in Holsteins

Genomic predictions for yield traits in US Holsteins with unknown parent groups

A. Cesarani, Y. Masuda, S. Tsuruta, E. L. Nicolazzi, P. M. VanRaden, D. Lourenco, and I. Misztal

1Department of Animal and Dairy Science, University of Georgia, Athens 30602
2Council on Dairy Cattle Breeding, Bowie, MD 20716
3Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

<table>
<thead>
<tr>
<th># animals with phenotypes and genotypes</th>
<th>580k</th>
</tr>
</thead>
<tbody>
<tr>
<td># animals with validation</td>
<td>381k</td>
</tr>
<tr>
<td>Assumed $M_e$</td>
<td>15k</td>
</tr>
<tr>
<td>Initial $h^2$</td>
<td>0.35</td>
</tr>
<tr>
<td>Predictivity</td>
<td>0.55</td>
</tr>
<tr>
<td>Calculated $h^2$</td>
<td>0.33</td>
</tr>
</tbody>
</table>
Predictivity with large populations

\[
acc = \frac{corr(y - Xb, \hat{u})}{h} < 1
\]

\[
h^2 > [corr(y - Xb, \hat{u})]^2
\]

If \( corr(y - Xb, \hat{u}) = 0.55 \) then \( h^2 > 0.30 \)
How to estimate genetic correlations?

Predictivity for trait i

\[ \text{corr}(y_i - Xb_i, \hat{u}_i) = \text{acc}_i \ h_i \]

What is predictivity from trait i to trait j?

\[ \text{corr}(y_i - Xb_i, \hat{u}_j) = ? \]

\[ \text{corr}(y_i - Xb_i, \hat{u}_j) = \text{acc}_j \ \text{corr}_{ij} \ h_i \]

\[ \text{corr}_{ij} = \frac{\text{corr}(y_i - Xb_i, \hat{u}_j)}{h_i \ \text{acc}_j} \]

\[ \text{SD}(\text{corr}_{ij}) \approx \frac{1}{h_i \ \text{acc}_j \ \sqrt{N_{val}}} \]
# Test by simulation

<table>
<thead>
<tr>
<th></th>
<th>Trait 1</th>
<th>Trait 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Validation population</strong></td>
<td>10k</td>
<td>10k</td>
</tr>
<tr>
<td><strong>Heritability</strong></td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td><strong>Correlation</strong></td>
<td>-0.5</td>
<td></td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td>0.8</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### Estimate

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>corr(_{12})</td>
<td>-0.50±0.06</td>
</tr>
<tr>
<td>corr(_{21})</td>
<td>-0.50±0.03</td>
</tr>
</tbody>
</table>
Procedure

• Select reference population, number of genotyped N > 10,000
• Select validation population, number of genotyped N_{val} > 5000
• Calculate predictivities within and across traits
• Recalculate heritabilities
  \[ \hat{h}^2 = \frac{c^2 + \sqrt{c^4 + 4c^2 M_e/N}}{2} \]
• Recalculate accuracies
  \[ acc = corr(y - Xb, \hat{u})/h \]
• Calculate genetic correlations
  \[ corr_{ij} = \frac{corr(y_i - Xb_i, \hat{u}_j)}{h_i \ acc_j} \pm \frac{1}{h_i \ acc_j \ \sqrt{N}} \]
  – Select correlation with smaller SE
Conclusions

- Estimate of heritability from theoretical and realized accuracies
- Estimate of genetic correlations by predictivity across traits
- Any data size and data slice
- Low standard errors with large reference and validation populations