Indirect predictions based on SNP effects from GBLUP with increasing number of genotyped animals

Andre Garcia\textsuperscript{1}, Y. Masuda\textsuperscript{1}, S. Miller\textsuperscript{2}, I. Misztal\textsuperscript{1}, D. Lourenco\textsuperscript{1}

\textsuperscript{1} University of Georgia - Animal and dairy science
\textsuperscript{2} Angus Genetics Inc. (AGI) - American Angus Association
Introduction

• Rapid increase of young genotyped animals into evaluations

• GBLUP or ssGBLUP

\[
\begin{bmatrix}
X'X & X'W \\
W'X & W'W + G^{-1}\lambda
\end{bmatrix}
\begin{bmatrix}
\beta \\
u
\end{bmatrix}
= 
\begin{bmatrix}
X'y \\
W'y
\end{bmatrix}
\]

Becomes infeasible to invert directly
Introduction

• Algorithm for Proven and Young (APY)

\[
G = \begin{bmatrix}
G_{cc} & G_{cn} \\
G_{nc} & G_{nn}
\end{bmatrix} \quad c = \text{core}
\]

\[
G_{nc} = \text{non-core}
\]

\[
G_{APY}^{-1} = \begin{bmatrix}
G_{cc}^{-1} & 0 \\
0 & 0
\end{bmatrix} + \begin{bmatrix}
-G_{cc}^{-1}G_{cn} \\
-G_{cc}^{-1}G_{cn}
\end{bmatrix} M_{nn}^{-1}\begin{bmatrix}
-G_{nc}G_{cc}^{-1} & I
\end{bmatrix}
\]

\[
m_{nn,i} = g_{ii} - G_{ic}G_{cc}^{-1}G_{ci}
\]

Misztal et al., 2014
Indirect predictions

Why do we may want to use IP?
  • Interim evaluations
  • Not all genotyped animals are included in evaluations
  • Genomic predictions for non-registered animals

How to calculate IP?

First we need SNP effects
\[ \hat{a} = DZ'G^{-1}\hat{u} \]

Then calculate IP
\[ IP = Z\hat{a} \]

VanRaden 2008
Stranden & Garrick 2009
Wang et al. 2012
What happens with IP when the number of genotyped animals keep increasing?
Data

• American Angus Association
• BW, WW and PWG

![Bar chart showing the number of genotyped animals by year of birth. The chart shows a significant increase from 114,937 in 2013 to 280,506.](image-url)
IP and core animals

• GBLUP with APY to get GEBV

• SNP effects:
  \[ \hat{a} = DZ' G^{-1}_{APY} \hat{u} \]  
  \[ \hat{a}_{core} = DZ' G^{-1}_{core} \hat{u}_{core} \]

• How do we choose core animals?
  All genotyped animals → Subset of core animals

• Correlations
  • cor(GEBV, IP)
  • cor(GEBV, IP_{core})
IP and core animals

- Should reflect the dimensionality of G
- Explain 98-99% of the variation in G
- Often randomly selected

- 99% of variation = 19k core animals

1. Born up to 2013 (core13)

2. Resampled as more genotypes were added (core14 and core15)
Results

\[ \hat{a} = DZ'G^{-1}_{APY}\hat{u} \]

- All genotyped animals for each year class with APY
- Correlations between IP and GEBV ≥ 0.99
- All traits and all core definitions
- GEBV successfully retrieved from IP
Results

Weaning Weight

• Subset of core animals (19k)

1. Born up to 2013

\[ \hat{a}_{\text{core13}} = DZ' G^{-1}_{\text{core13}} \hat{u}_{\text{core13}} \]
Results

Weaning Weight
- Subset of core animals (19k)

2. Born up to 2014

\[
\hat{a}_{\text{core14}} = DZ'G_{\text{core14}}^{-1} \hat{u}_{\text{core14}} \\
\hat{a}_{\text{core15}} = DZ'G_{\text{core15}}^{-1} \hat{u}_{\text{core15}}
\]
How many animals are needed to obtain reliable indirect predictions?
Data up to 2015

Correlation between GEBV and IP

% of variance explained by eigenvalues

98-99% of variance explained

97% 98.5% 99.2% 99.8% 99.9%

% of variance explained by eigenvalues

10000 15000 20000 30000 40000

Number of animals

50% 64% 78% 85% 89% 91%

cor(GEBV, IP)

0.61 0.73 0.85 0.89 0.93 0.94 0.98 0.99 0.98

Weaning Weight

0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Remarks

• Using all genotyped animals to calculate SNP effects:
  • Indirect predictions are robust
  • Independent of the choice of core animals for APY

• Using the subset of core animals:
  • Robust indirect predictions with updated set of core animals

• Using only a small portion of the data:
  • Indirect predictions are less accurate
Acknowledgements