SNP-effects depend on the genotype coding

Aniek Bouwman, Ben Hayes and Mario Calus
SNP-effects from genomic prediction

SNP-effects of interest for:

- DGV of young selection candidates
- GWAS
- “Local DGV”
Genotype coding (0, 1, 2)

Centred: \[ z_{ij} = x_{ij} - 2p_i \]

Centred & scaled: \[ w_{ij} = \frac{(x_{ij}-2p_i)}{\sqrt{2p_i(1-p_i)}} \]
Influence of genotype coding

Scaling has:

- no effect on DGV (Stranden & Christensen, 2011)
- effect on SNP-effects (de los Campos et al., 2013)

With scaling low MAF SNP have

- less shrinkage to the mean
- thus larger SNP-effects
Objectives

Show that difference in SNP-effect is due to shrinkage
Show factors influencing the differences in SNP-effects

1. Real data example
2. Theoretical framework
1. Real data

5500 bulls with 700k genotypes

GREML
GRM:

- VR1 (centred: $z_{ij} = x_{ij} - 2p_i$)
- VR2 (centred & scaled: $w_{ij} = \frac{(x_{ij} - 2p_i)}{\sqrt{2p_i(1-p_i)}}$)

Back-solved SNP-effects

VanRaden, 2008
Scaling (VR2) vs no scaling (VR1)

DGV

- Correlation: 0.9988
- Regr coefficient: 1.0011 (VR2 on VR1)

Scaling has no effect on DGV

SNP effects

- Correlation: 0.27

Scaling has an effect on SNP-effects
SNP effect vs MAF

Less shrinkage to the mean
SNP with same MAF grouped

Correlation SNP-effects VR1 & VR2

Regression of SNP-effects from VR2 on VR1

Difference in SNP-effects mainly due to shrinkage
2. Theoretical framework

Derived the theoretical ratio between SNP-effects based on VR1 and VR2

Allele frequency distributions

![Graphs showing allele frequency distributions for two different conditions.](image-url)
Derivation ratio: uniform distribution

\[ \hat{\alpha}_{VR2,j} = \left( \frac{h^2}{1 - h^2 n} + \frac{1}{3} N \right) \left( \frac{h^2}{1 - h^2 n + N} \right) \hat{\alpha}_{VR1,j} \]
More power

\[ h^2 = 0.2 \]

**SNP-effects** from VR1 and VR2 become more similar with increasing power.

- 1,000,000 N markers
  - □ 50,000
  - ▲ 30,000,000

**Uniform**

**U-shape**

- Uniform shape
- SNP-effects from VR1 and VR2 become more similar with increasing power
Conclusions

Genotype coding (scaling) influences SNP-effects

Differences mainly caused by shrinkage

Large differences for low MAF variants

Differences in SNP-effect influenced by power of data

Be aware when using, reporting and comparing SNP-effects