Binary traits: the speed-accuracy trade-off

Genomic prediction of a binary trait using a threshold model

Jan ten Napel, Ismo Stranden, Matti Taskinen, Jérémie Vandenplas, Roel Veerkamp & Kaarina Matilainen

Background

- Many traits have a continuous phenotypic distribution

- Categorical traits have a discontinuous distribution of phenotypes

  - Binary traits (0/1, all-or-none, absent-present)

  - Ordered categorical traits (e.g. small, medium, large)

  - Unordered categorical traits (e.g red, yellow, blue)

- This has implications for genetic evaluation!

Linear models are incorrect for categorical traits

- Common linear methods to estimate genetic merit

  - Assume a continuous phenotypic distribution

- Assumptions are violated for categorical traits

  - Residual variance can be very heterogeneous

    - Extreme category problem: all observations in a fixed-effect class have the same score

  - Genetic and residual effects not independent

  - Heritability and variance components on observed scale depend on prevalence

Alternative approach: threshold model

- Presume unobserved continuous trait underlying the observed categorical trait: liability

- Threshold for every change in category

- Liability is unobserved so

  - Unknown mean and variance

  - Either mean or a threshold set to zero

  - Residual variance set to 1
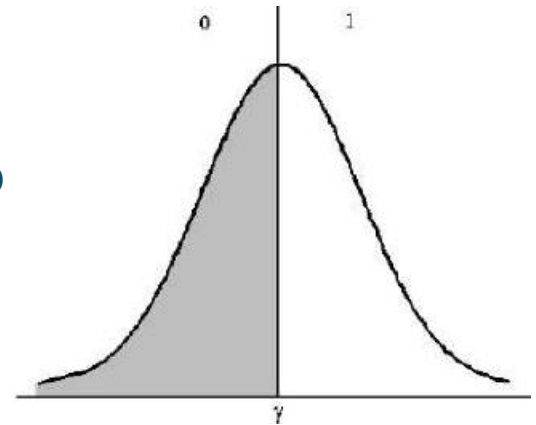
Heritability underlying scale is 0.20
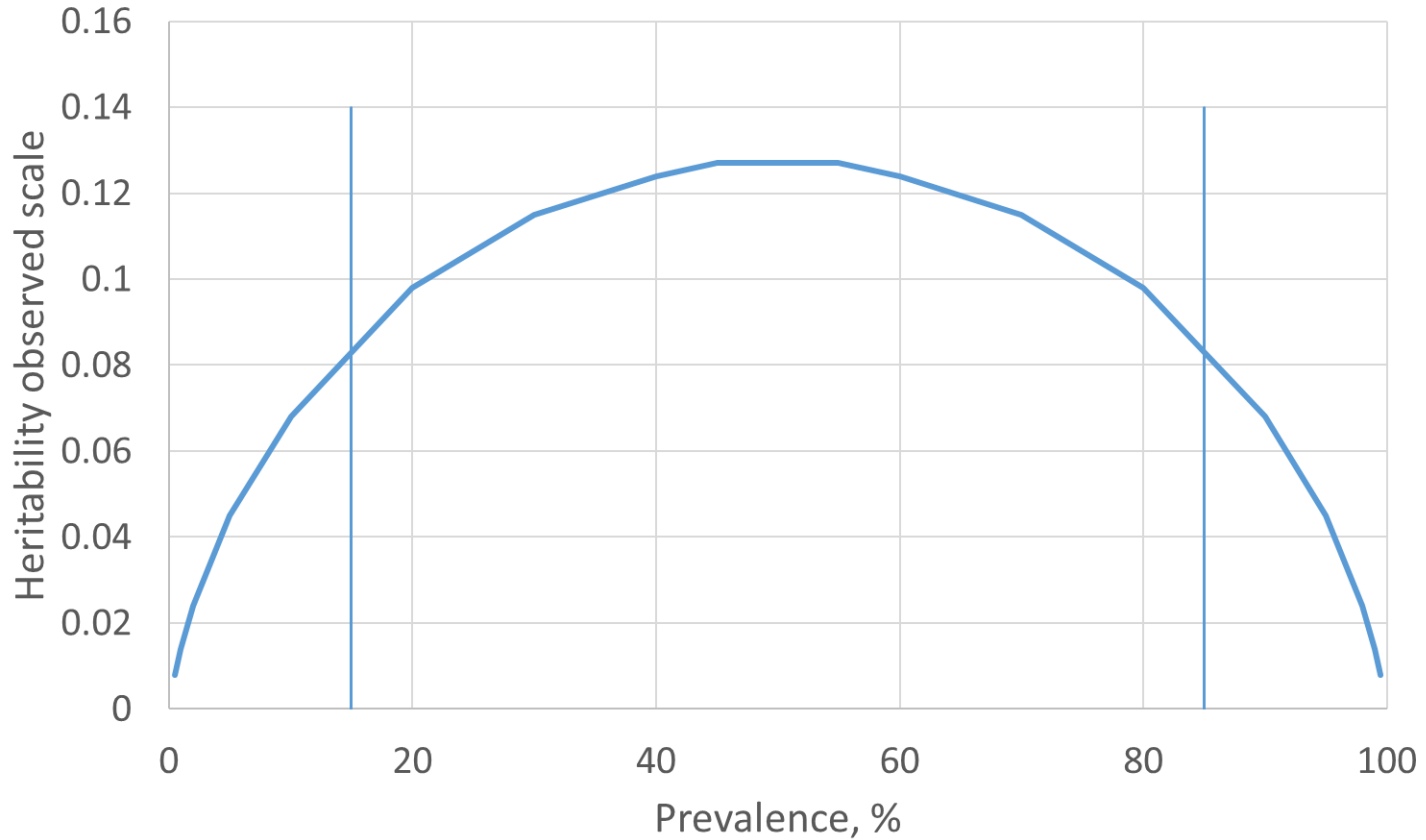
Prediction of effects

- Prediction of genetic effects (EBVs)

  - Threshold model on liability scale

  - Linear model on observed scale

- Consensus in literature (analyses based on pedigree)

  - At intermediate range of prevalence (15-85%)

    - Rank correlation of EBV very high

    - Linear model much faster

- Is the benefit of using a threshold model affected by the use of genomic information?

Validation study - aim

- Compare realised reliability for various ways to evaluate a binary trait

- Models
  - Threshold model with Gibbs sampling
  - Threshold model with iterative solving (MiXBLUP)
  - Linear model with iterative solving (MiXBLUP)

- Genetic relationships
  - Pedigree information
  - Pedigree + genomic information

Validation study - simulation

- Population simulated using QMSim
    - 10 generations of 1,080 individuals out of 15 sires and 90 dams (N=10,905)
    - 19 chromosomes, with 19,000 SNP and 760 QTL
- Trait
    - True breeding value on liability scale (h^2=0.20)
    - Liability is simulated phenotype, affected by sex and herd
    - Observed phenotype is derived binary trait with prevalence of 15%

Validation study - design

- Generation 1-9 have phenotypes (training population)

- Generation 10 (validation population)

- Generation 7-10 are genotyped

- Realised reliability is squared correlation between EBV and TBV of animals in generation 10

Validation study: realised reliability

Method	Pedigree		Pedigree+genomic	
Threshold – Gibbs	19.1		22.0	
Threshold – solving	19.0		21.8	
	+0.8			+1.7
Linear - solving	18.2	+1.9	20.1	

- Both threshold models give virtually the same result
- Benefit threshold model increases with use of genomics (+1.7 vs +0.8)
- Benefit threshold model similar to benefit using genomics (+1.7 vs +1.9)

Validation study: run time

- **Run-time**
  - Substantially longer for Gibbs sampling than iterative solving (~60x)
  - Only slightly longer for threshold model than linear model (~1.5-2x)

Wrap-up

- The benefit of using a threshold model in this example
    - Is comparable to the benefit of using genomic information
    - Is higher when using genomic information than pedigree information only

Genomic breeding values for binary traits?
At least consider a threshold model!

Acknowledgments

Breed4Food is dedicated to be the leading research consortium in animal breeding, genetics and genomics enabling the Breed4Food partners to breed better products to benefit society's needs.

World Congress on
Genetics Applied
to Livestock Production

Rotterdam | The Netherlands
3 - 8 July 2022

WCGALP 2022

QUALITY, INCLUSIVE,
ATTRACTIVE AND MODERN.
WITH A DUTCH TOUCH.

www.wcgalp2022.com

IMPORTANT DATES

12 January 2022: Deadline abstract submission
1 March   2022: Deadline early bird registration