

Challenges and Opportunities in Genetic Evaluations Incorporating Genomics

Stephen D. Kachman

Department of Statistics
University of Nebraska–Lincoln

June 3, 2011

- Where we stand
- Tradeoffs
- Challenges and opportunities

Where we stand

Challenges
and
Opportunities

Introduction

Where we
stand

Genotypes
Genotypes to
EPDs

Tradeoffs

Challenges
and
Opportunities

- Genomic information is being incorporated into national cattle evaluations
- Dairy approach with a standardized panel
- Beef approach with heterogeneous MBVs

- Moving towards 50K SNP panels being the standard

- Moving towards 50K SNP panels being the standard
- 800K SNP panels

- Moving towards 50K SNP panels being the standard
- 800K SNP panels
- Sequence data
- Data handling is a growing issue

Separate training population approach

- Training
 - Genotypes and Phenotypes
 - Single type of panel
 - Phenotypes are often based on EPDs
 - Bayes[A- Ω] to estimate SNP (or haplotype) effects

Separate training population approach

- Training
 - Genotypes and Phenotypes
 - Single type of panel
 - Phenotypes are often based on EPDs
 - Bayes[A- Ω] to estimate SNP (or haplotype) effects
- Incorporate into the NCE
 - Selection Index (Dairy)
 - MBV $\rightarrow$ Correlated Trait (Beef)

Converting from genotypes to EPDs

Challenges
and
Opportunities

Introduction

Where we
stand
Genotypes

Genotypes to
EPDs

Tradeoffs

Challenges
and
Opportunities

Combined training/evaluation approach

- Genotypes and Phenotypes
- Standardize to a single type of panel
- Genomic relationship matrix

Combined training/evaluation approach

- Genotypes and Phenotypes
- Standardize to a single type of panel
 - Imputation from smaller to larger panels
 - Prior to training/evaluation
- Genomic relationship matrix

Combined training/evaluation approach

- Genotypes and Phenotypes
- Standardize to a single type of panel
 - Imputation from smaller to larger panels
 - Prior to training/evaluation
- Genomic relationship matrix
 - Mixture of genotyped and ungenotyped animals
 - In principle other types of models could be used

Separate training population approach

- Define the type of training data you will work with
 - Every animal will be genotyped with a specific panel
 - Every genotyped animal will have a phenotype
- Manageable training data set sizes
- Work with sophisticated models which are computationally intensive in the training phase

Separate training population approach

- Evaluation population is different from your training population
 - Separation can be horizontal (different subpopulations of cattle)

Separate training population approach

- Evaluation population is different from your training population
 - Separation can be horizontal (different subpopulations of cattle)
 - Separation will be vertical (different generations of cattle)
 - Animals for which genomic information is most useful are the ones without phenotypic information

Separate training population approach

- Evaluation population is different from your training population
 - Separation can be horizontal (different subpopulations of cattle)
 - Separation will be vertical (different generations of cattle)
 - Animals for which genomic information is most useful are the ones without phenotypic information
 - Allow MBVs to be trained for different panels or populations
 - Proliferation of MBVs

Combined training/evaluation approach

- Train on the entire data set
 - Use the most current set of phenotypic data
- Simultaneously manage all of the genotypic and phenotypic data
 - Updating a database with a 50K/800K/Sequence genotype for an animal
 - Running an evaluation using the raw genotypes
- Computational considerations will drive the type of models that can be used

Combined training/evaluation approach

- Using the evaluation population for training
 - Reduces the horizontal separation
 - Training is weighted towards animals with the most phenotypic information
 - Genomic information is most useful for animals with the least phenotypic information
 - Vertical separation remains



③ Evaluations to produce genomic enhanced EPDs based on first generation MBVs

- First Phase
 - ① Training on phenotype based EPDs
 - ② Produce first generation MBVs
 - ③ Evaluations to produce genomic enhanced EPDs based on first generation MBVs
- Second Phase
 - Current EPDs now contain genomic information
 - Train on genomic enhanced EPDs

Challenges and opportunities

Challenges and Opportunities

Introduction

Where we stand

Tradeoffs

Challenges and Opportunities

- First Phase
 - ① Training on phenotype based EPDs
 - ② Produce first generation MBVs
 - ③ Evaluations to produce genomic enhanced EPDs based on first generation MBVs
- Second Phase
 - Current EPDs now contain genomic information
 - Train on genomic enhanced EPDs
 - Drift to producing MBVs that do good job of predicting the previous MBV?

- First Phase

- 1 Training on phenotype based EPDs
- 2 Produce first generation MBVs
- 3 Evaluations to produce genomic enhanced EPDs based on first generation MBVs

- Second Phase

- Current EPDs now contain genomic information
- Train on genomic enhanced EPDs
Drift to producing MBVs that do good job of predicting the previous MBV?
- Need for phenotype based EPDs for training as we move forward

- Correlated Trait Models
 - Large number of traits
 - More equations
 - Slower convergence

- Correlated Trait Models
 - Large number of traits
 - More equations
 - Slower convergence
 - Large genetic correlation matrix
 - High genetic correlations between MBVs
 - Estimation of a large number of parameters
 - Genetic covariance may not be well behaved

- Correlated Trait Models
 - Large number of traits
 - More equations
 - Slower convergence
 - Large genetic correlation matrix
 - High genetic correlations between MBVs
 - Estimation of a large number of parameters
 - Genetic covariance may not be well behaved
 - Heritabilities close to 1
 - Software has been tuned for moderate heritabilities

MBV Heritability close to 1

Challenges
and
Opportunities

Introduction

Where we
stand

Tradeoffs

Challenges
and
Opportunities

$$\begin{pmatrix} G^{11} & G^{12} \\ G^{21} & \mathbf{R}_2^{-1} + G^{22} \end{pmatrix} \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{R}_2^{-1} \mathbf{1}_{y_2} \end{pmatrix}$$

- Poorly conditioned matrix
- Convergence rates are sensitive to how the model is parameterized

All models are wrong, but some are useful

Challenges
and
Opportunities

Introduction

Where we
stand

Tradeoffs

Challenges
and
Opportunities

- Linkage relationships
 - Change between breeds
 - Change between subpopulations within a breed
 - Change over generations

All models are wrong, but some are useful

Challenges
and
Opportunities

Introduction

Where we
stand

Tradeoffs

Challenges
and
Opportunities

- Linkage relationships
 - Change between breeds
 - Change between subpopulations within a breed
 - Change over generations
 - Same genotypes
 - Same prediction
 - Same accuracy

All models are wrong, but some are useful

Challenges
and
Opportunities

Introduction

Where we
stand

Tradeoffs

Challenges
and
Opportunities

- Assume that effects don't change
 - depend on the genetic background
 - depends on the environment

All models are wrong, but some are useful

Challenges
and
Opportunities

Introduction

Where we
stand

Tradeoffs

Challenges
and
Opportunities

- Assume that effects don't change
 - depend on the genetic background
 - depends on the environment
- Assume genotypes are complete
 - An imputed genotype is equivalent to a complete genotype
 - Some animals may have genotypes that are more informative than others

- Currently implemented approaches for incorporating genomic information separate training and evaluation
 - Computational considerations

- Currently implemented approaches for incorporating genomic information separate training and evaluation
 - Computational considerations
 - Similar to the separation of parameter estimation and prediction of EPDs

- Currently implemented approaches for incorporating genomic information separate training and evaluation
 - Computational considerations
 - Similar to the separation of parameter estimation and prediction of EPDs
- Crossbred animal performance

- Currently implemented approaches for incorporating genomic information separate training and evaluation
 - Computational considerations
 - Similar to the separation of parameter estimation and prediction of EPDs
- Crossbred animal performance
- Need for phenotypes won't go away