Application of Whole-Genome Prediction Methods for Genome-Wide Association Studies: a Bayesian Approach

R.L. Fernando A. Toosi D.J. Garrick J.C.M. Dekkers

Department of Animal Science
Iowa State University

10^{th} World Congress of Genetics Applied to Livestock Production

Two Approaches

- Bayesian multiple-regression models (BMR)
- Single-marker models (SM)

Two Approaches

- Bayesian multiple-regression models (BMR)
- Single-marker models (SM)

Compare Approaches

	SM	BMR
Model	Simple Regression	Multiple Regression
False Positives (FP)	Genomewise Error Rate	Proportion of FP
Inference	Frequentist	Bayesian

Models

Simple Regression

- QTL may have low LD with all markers in region
- Need to explicitly model population structure

Multiple Regression

- Inference based on genomic windows
- Markers can capture population structure
    - Explicit modeling of structure results in lower power
- Inference of QTL

Models

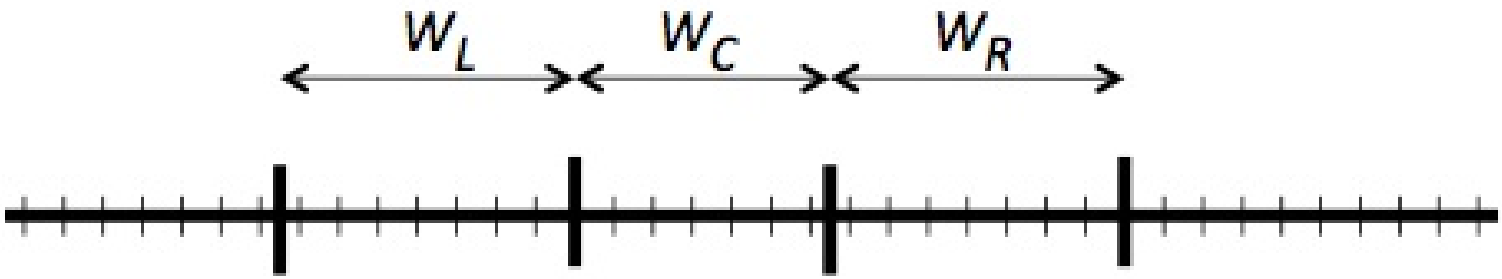
Simple Regression

- QTL may have low LD with all markers in region
- Need to explicitly model population structure

Multiple Regression

- Inference based on genomic windows
- Markers can capture population structure
  - Explicit modeling of structure results in lower power
- Inference of QTL

Composite Genomic Window



Controlling False Positives

Genomewise error rate

- Control probability of one or more false positives among all tests
- Incurs multiple-test penalty

Proportion of false positives

- Control proportion of false positives (PFP)
- Related to FDR
- No multiple-test penalty (Fernando et al., 2004; Stephens and Balding, 2009)

Controlling False Positives

Genomewise error rate

- Control probability of one or more false positives among all tests
- Incurs multiple-test penalty

Proportion of false positives

- Control proportion of false positives (PFP)
- Related to FDR
- No multiple-test penalty (Fernando et al., 2004; Stephens and Balding, 2009)

Definition PFP

- V number of false positives
- R number of positives
- $\text{PFP} = \frac{\text{E}(V)}{\text{E}(R)}$

- $\text{FDR} = \text{E}(\frac{V}{R}|R > 0)\Pr(R > 0)$
- If PFP is γ in each of n independent experiments, the proportion of false positives among significant results across all experiments will converge to γ as n increases.

- In general, the above property does not hold for FDR.

- PFP is a multiple test extension of the posterior type I error rate (PER).

- If PER is γ for a random test, PFP is also γ for the collection of tests.

Definition PFP

- V number of false positives
- R number of positives
- $\text{PFP} = \frac{\text{E}(V)}{\text{E}(R)}$
- $\text{FDR} = \text{E}(\frac{V}{R}|R > 0)\,\text{Pr}(R > 0)$
- If PFP is γ in each of n independent experiments, the proportion of false positives among significant results across all experiments will converge to γ as n increases.
- In general, the above property does not hold for FDR.
- PFP is a multiple test extension of the posterior type I error rate (PER).
- If PER is γ for a random test, PFP is also γ for the collection of tests.

Definition PFP

- V number of false positives
- R number of positives
- $\text{PFP} = \frac{\text{E}(V)}{\text{E}(R)}$

- $\text{FDR} = \text{E}(\frac{V}{R}|R > 0)\,\text{Pr}(R > 0)$
- If PFP is γ in each of n independent experiments, the proportion of false positives among significant results across all experiments will converge to γ as n increases.

- In general, the above property does not hold for FDR.

- PFP is a multiple test extension of the posterior type I error rate (PER).

- If PER is γ for a random test, PFP is also γ for the collection of tests.

Definition PFP

- V number of false positives
- R number of positives
- $\text{PFP} = \frac{\text{E}(V)}{\text{E}(R)}$

- $\text{FDR} = \text{E}(\frac{V}{R}|R > 0)\,\text{Pr}(R > 0)$
- If PFP is γ in each of n independent experiments, the proportion of false positives among significant results across all experiments will converge to γ as n increases.

- In general, the above property does not hold for FDR.

- PFP is a multiple test extension of the posterior type I error rate (PER).

- If PER is γ for a random test, PFP is also γ for the collection of tests.

Definition PFP

- V number of false positives
- R number of positives
- $\text{PFP} = \frac{\text{E}(V)}{\text{E}(R)}$
- $\text{FDR} = \text{E}(\frac{V}{R}|R > 0)\,\text{Pr}(R > 0)$
- If PFP is γ in each of n independent experiments, the proportion of false positives among significant results across all experiments will converge to γ as n increases.
- In general, the above property does not hold for FDR.
- PFP is a multiple test extension of the posterior type I error rate (PER).
- If PER is γ for a random test, PFP is also γ for the collection of tests.

Definition PFP

- V number of false positives
- R number of positives
- $\text{PFP} = \frac{\text{E}(V)}{\text{E}(R)}$
- $\text{FDR} = \text{E}(\frac{V}{R}|R > 0)\,\text{Pr}(R > 0)$
- If PFP is γ in each of n independent experiments, the proportion of false positives among significant results across all experiments will converge to γ as n increases.
- In general, the above property does not hold for FDR.
- PFP is a multiple test extension of the posterior type I error rate (PER).
- If PER is γ for a random test, PFP is also γ for the collection of tests.

Definition of PER

- In the frequentist approach, inference on H_0 is based on the distribution of some test statistic given H_0 is true
- posterior type I error rate (PER) is the conditional probability of H_0 being true given that, based on a statistical test, H_0 has been rejected.

$$\begin{aligned}\text{PER} &= \frac{\Pr(H_0 \text{ is rejected}, H_0 \text{ is true})}{\Pr(H_0 \text{ is rejected}, H_0 \text{ is true}) + \Pr(H_0 \text{ is rejected}, H_0 \text{ is false})} \\ &= \frac{\alpha \Pr(H_0)}{\alpha \Pr(H_0) + (1-\beta)[1-\Pr(H_0)]}\end{aligned}$$

α is the type I error rate, and $(1-\beta)$ is the power of the test

Definition of PER

- In the frequentist approach, inference on H_0 is based on the distribution of some test statistic given H_0 is true
- posterior type I error rate (PER) is the conditional probability of H_0 being true given that, based on a statistical test, H_0 has been rejected.

$$\begin{aligned} \text{PER} &= \frac{\Pr(H_0 \text{ is rejected}, H_0 \text{ is true})}{\Pr(H_0 \text{ is rejected}, H_0 \text{ is true}) + \Pr(H_0 \text{ is rejected}, H_0 \text{ is false})} \\ &= \frac{\alpha \Pr(H_0)}{\alpha \Pr(H_0) + (1-\beta)[1-\Pr(H_0)]} \end{aligned}$$

α is the type I error rate, and $(1-\beta)$ is the power of the test

Definition of PER

- In the frequentist approach, inference on H_0 is based on the distribution of some test statistic given H_0 is true
- posterior type I error rate (PER) is the conditional probability of H_0 being true given that, based on a statistical test, H_0 has been rejected.

$$\begin{aligned} \text{PER} &= \frac{\Pr(H_0 \text{ is rejected}, H_0 \text{ is true})}{\Pr(H_0 \text{ is rejected}, H_0 \text{ is true}) + \Pr(H_0 \text{ is rejected}, H_0 \text{ is false})} \\ &= \frac{\alpha \Pr(H_0)}{\alpha \Pr(H_0) + (1-\beta)[1-\Pr(H_0)]} \end{aligned}$$

α is the type I error rate, and $(1-\beta)$ is the power of the test

Definition of PER

- In the Bayesian approach, inference on H_0 is based on $\Pr(H_0|\mathbf{y})$.
- Typically, $\Pr(H_0|\mathbf{y})$ is estimated by counting the number of MCMC samples where H_0 is true.
- If H_0 is rejected when $\Pr(H_0|\mathbf{y}) < \gamma$, PER $< \gamma$.
- $\Pr(H_0|\mathbf{y})$ is not a frequentist probability.

Definition of PER

- In the Bayesian approach, inference on H_0 is based on $\Pr(H_0|\mathbf{y})$.
- Typically, $\Pr(H_0|\mathbf{y})$ is estimated by counting the number of MCMC samples where H_0 is true.
- If H_0 is rejected when $\Pr(H_0|\mathbf{y}) < \gamma$, PER $< \gamma$.
- $\Pr(H_0|\mathbf{y})$ is not a frequentist probability.

Definition of PER

- In the Bayesian approach, inference on H_0 is based on $\Pr(H_0|\mathbf{y})$.
- Typically, $\Pr(H_0|\mathbf{y})$ is estimated by counting the number of MCMC samples where H_0 is true.
- If H_0 is rejected when $\Pr(H_0|\mathbf{y}) < \gamma$, PER $< \gamma$.
- $\Pr(H_0|\mathbf{y})$ is not a frequentist probability.

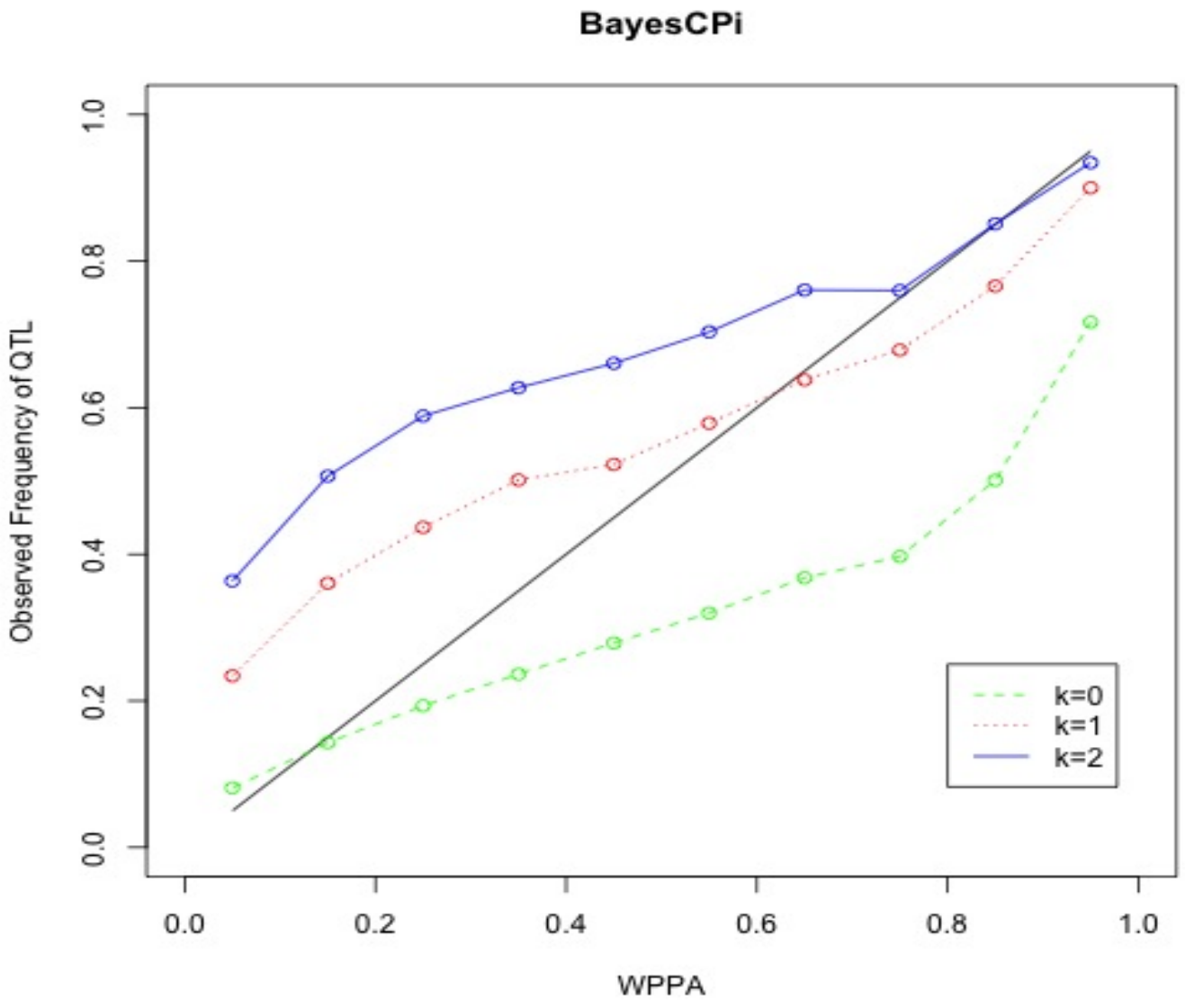
Definition of PER

- In the Bayesian approach, inference on H_0 is based on $\Pr(H_0|\mathbf{y})$.
- Typically, $\Pr(H_0|\mathbf{y})$ is estimated by counting the number of MCMC samples where H_0 is true.
- If H_0 is rejected when $\Pr(H_0|\mathbf{y}) < \gamma$, PER $< \gamma$.
- $\Pr(H_0|\mathbf{y})$ is not a frequentist probability.

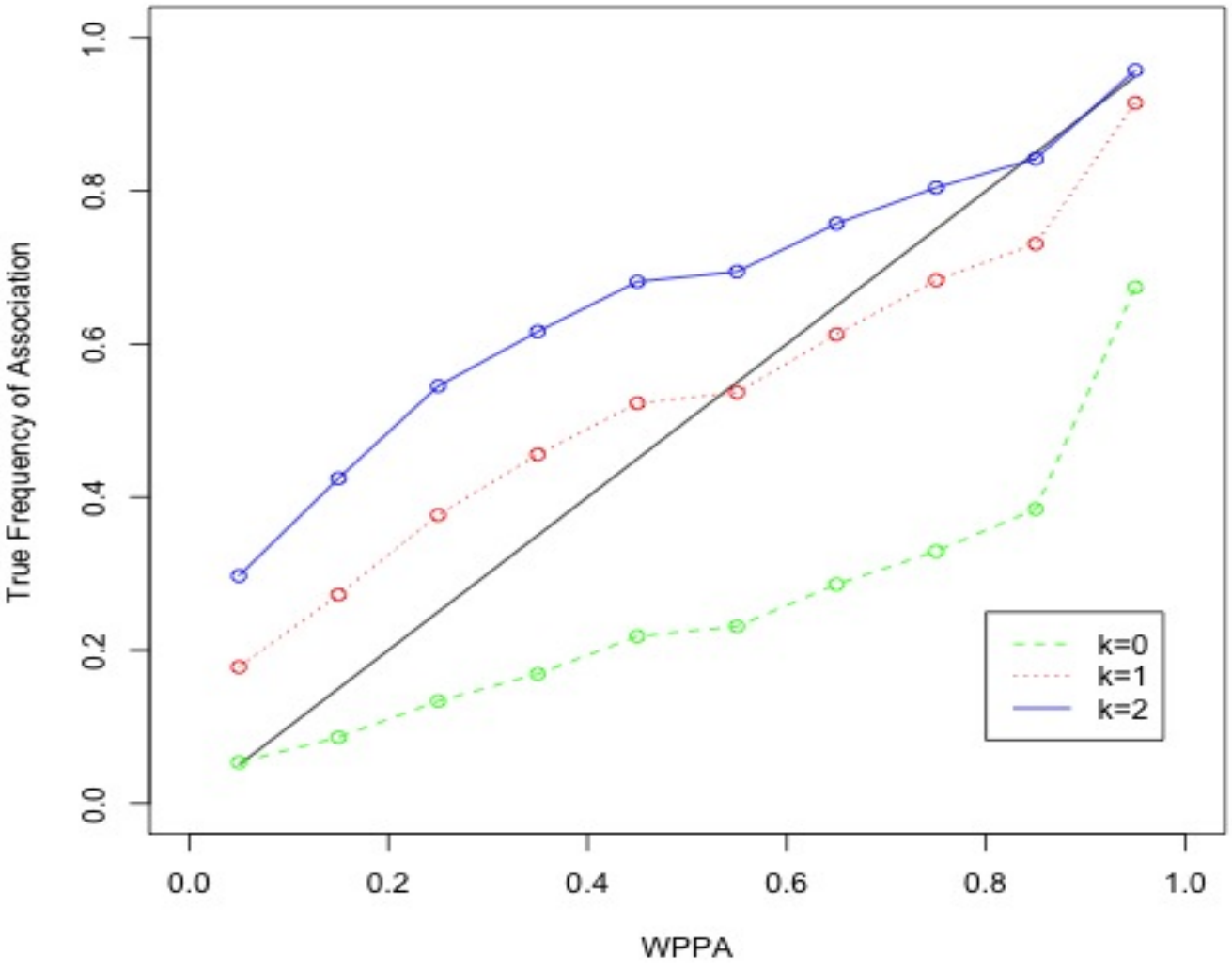
Simulation

- 52k SNP genotypes from 3,570 Angus bulls
- 100 data sets of size 1000 or 3,570 were randomly sampled
- marker effects randomly sampled according to BayesC with $\pi = 0.995$
- markers with non-zero effects (QTL) were not included in marker panel
- $h^2 = 0.9$

Results for N=1000



Results for N=3,570



Summary

- Genomic window based inference multiple regression models
- When PFP is used to manage false positives, no multiple-test penalty
- Bayesian posterior probabilities can be used to control PFP
  - $\Pr(H_0)$, and power of test can be treated as unknown
  - Do not need to know the distribution of test statistic
  - Simple to determine significance threshold

Summary

- Genomic window based inference multiple regression models
- When PFP is used to manage false positives, no multiple-test penalty
- Bayesian posterior probabilities can be used to control PFP
  - $\Pr(H_0)$, and power of test can be treated as unknown
  - Do not need to know the distribution of test statistic
  - Simple to determine significance threshold

Summary

- Genomic window based inference multiple regression models
- When PFP is used to manage false positives, no multiple-test penalty
- Bayesian posterior probabilities can be used to control PFP
  - $\Pr(H_0)$, and power of test can be treated as unknown
  - Do not need to know the distribution of test statistic
  - Simple to determine significance threshold

Acknowledgements

- Funding:
  - NIH Grant R01GM099992
  - USDA/AFRI project EBIGS