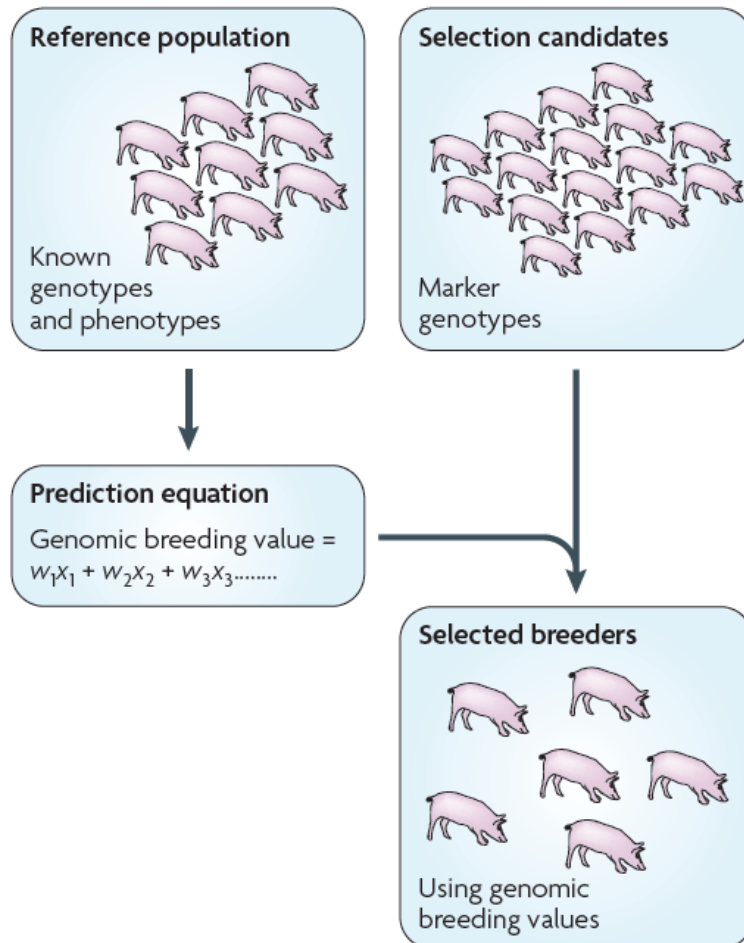# Bayesian Methods For Genomic Prediction

Jian Zeng

Slides credit: Ben Hayes
Queensland Alliance for Agriculture and Food
Innovation, University of Queensland

Bayesian methods for Genomic Prediction

- Alternative assumptions regarding the distribution of SNP effects

- Introduction to Bayesian methods

- Genomic prediction with Bayesian methods

- Comparison of accuracy of methods

Assumptions regarding distribution of SNP effects

Reference population — Known genotypes and phenotypes

Selection candidates — Marker genotypes

Prediction equation — Genomic breeding value = $w_1x_1 + w_2x_2 + w_3x_3\ldots\ldots$

Selected breeders — Using genomic breeding values

- Best Linear Unbiased Prediction
  - GBLUP, SNPBLUP
- GREML
- *Assumes SNP effects are:*
  - *all non-zero*
  - *very small*
  - *normally distributed*
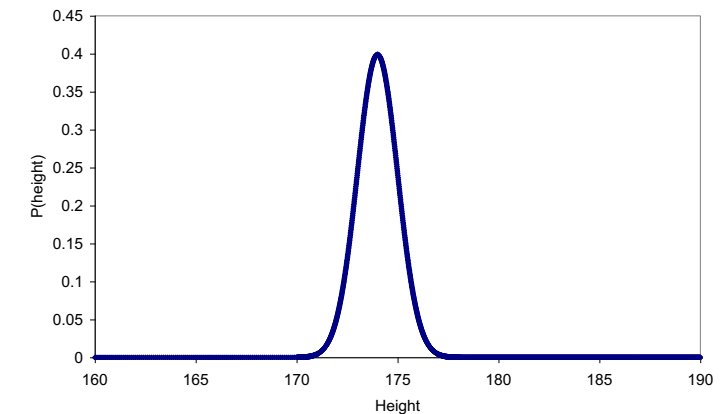
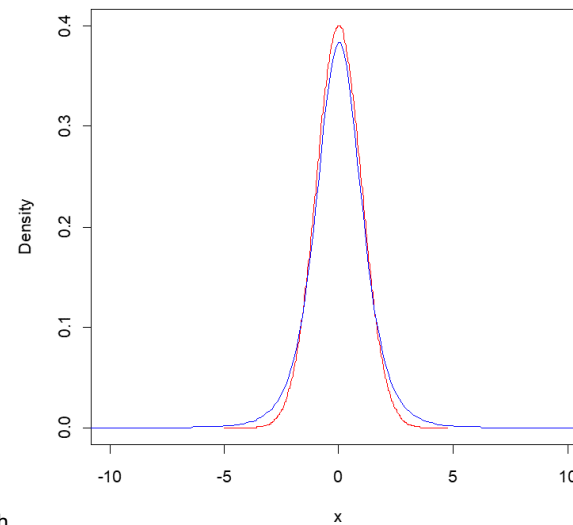# Assumptions regarding distribution of SNP effects
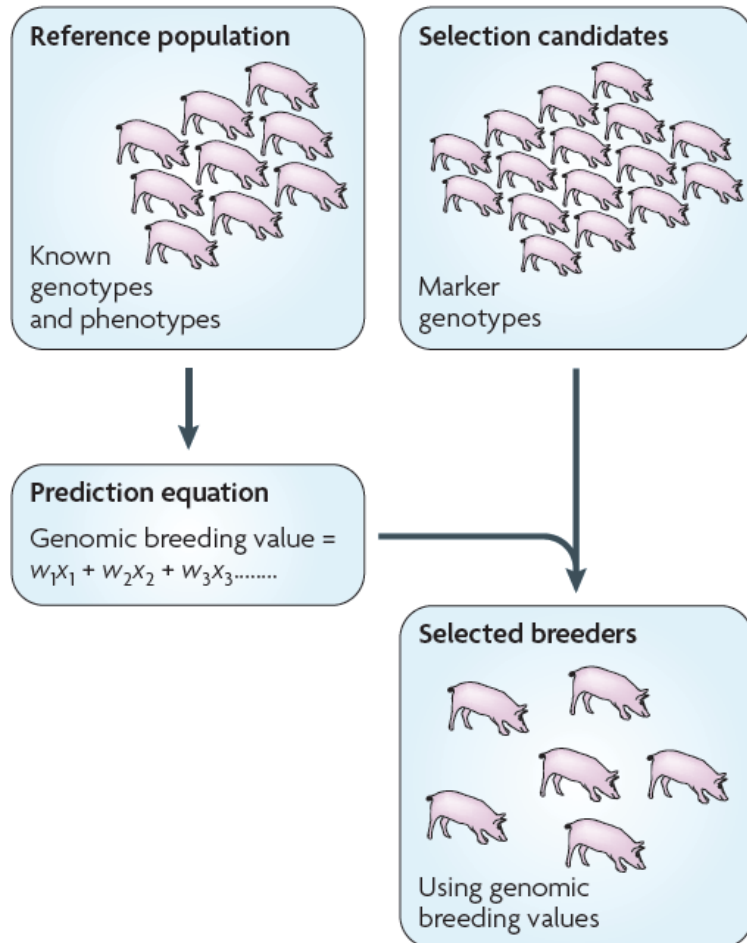
- Alternative distributions?

Assumptions regarding distribution of SNP effects

- Alternative distributions?

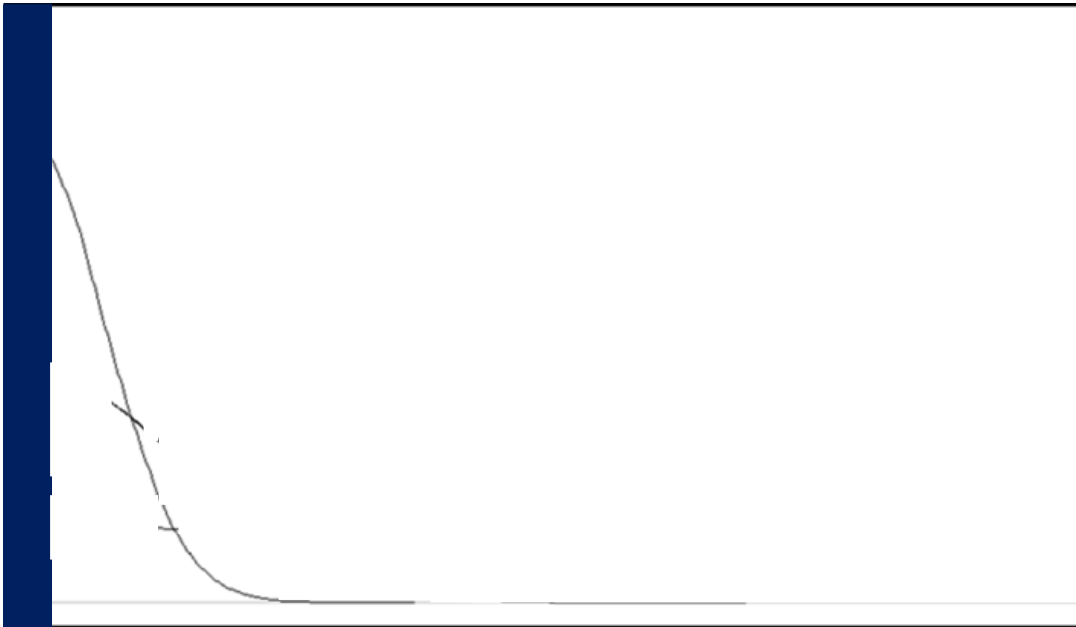| Assumption	Distribution of SNP effects	Method
Small number of moderate to large effects, many small effects	Students t	BayesA

Assumptions regarding distribution of SNP effects

- Alternative distributions?

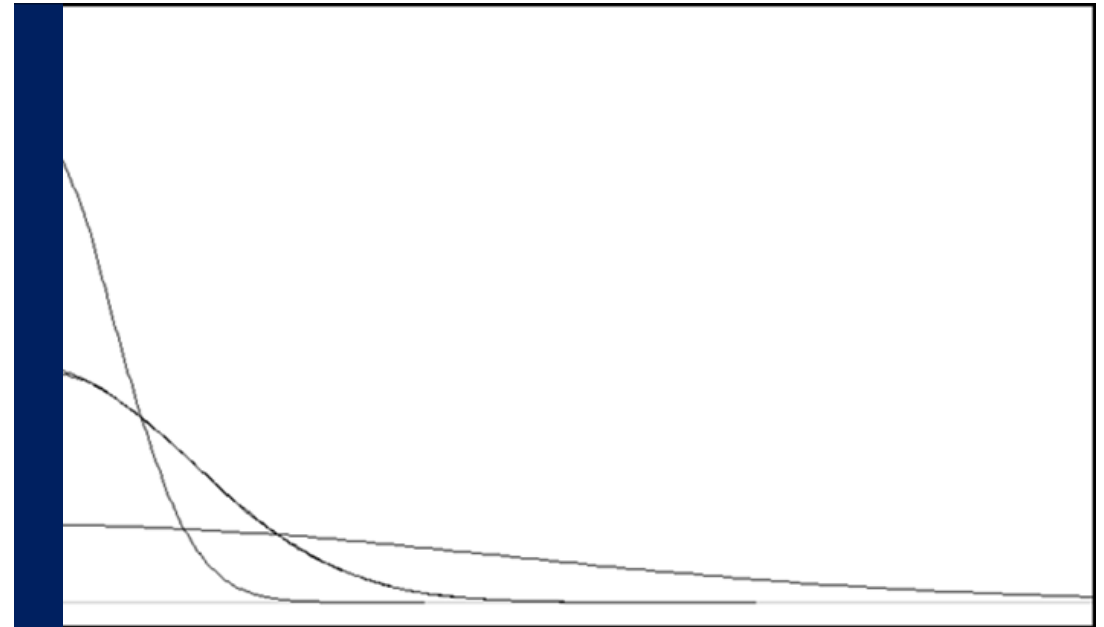| Assumption	Distribution of SNP effects	Method
Small number of moderate to large effects, many small effects	Students t	BayesA
Small number of moderate to large effects, many zero effects	Mixture, spike at zero, Students t	BayesB
Small number of small effects, many zero effects	Mixture, spike at zero, normal distribution	BayesC
Many zero effects, proportion of small effects, some moderate to large effects	Multi-variate normal	BayesR

Assumptions regarding distribution of SNP effects

- Alternative distributions?

BayesC

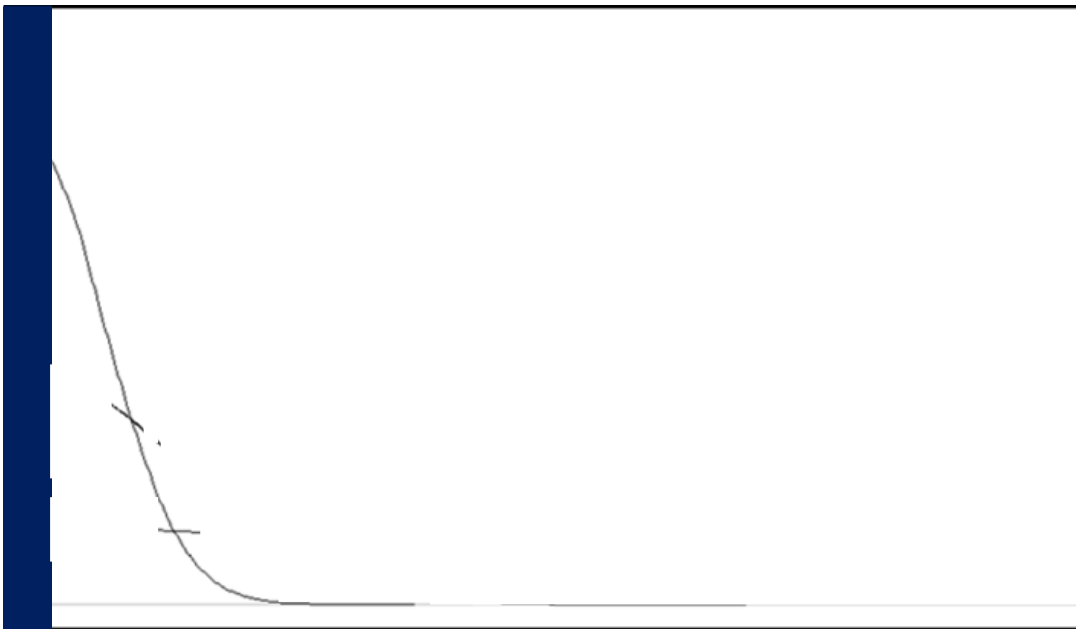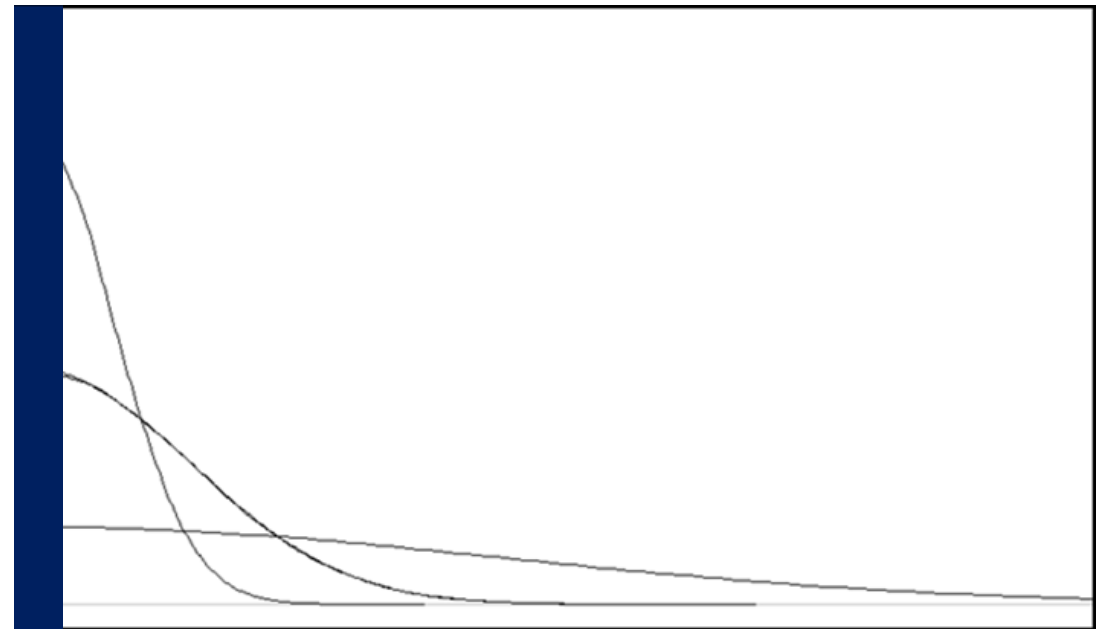
BayesR

Assumptions regarding distribution of SNP effects

- Alternative distributions?

BayesC

BayesR

Bayesian approach allows us to incorporate this prior knowledge in the prediction of SNP effects

Bayesian methods for Genomic Prediction

- Alternative assumptions regarding the distribution of SNP effects

- Introduction to Bayesian methods

- Genomic prediction with Bayesian methods

- Comparison of accuracy of methods

Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x)P(x)$$

Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x)P(x)$$

Probability of
parameters x given
the data y (posterior)

Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x)P(x)$$

Probability of parameters x given the data y (posterior)

Is proportional to

Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x)P(x)$$

Probability of parameters x given the data y (posterior)

Is proportional to

Probability of data y given the x (likelihood of data)

Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x)P(x)$$

Probability of parameters x given the data y (posterior)

Is proportional to

Probability of data y given the x (likelihood of data)

Prior probability of x

Bayesian methods

- Consider an experiment where we measure height of 10 people to estimate average height
- We want to use prior knowledge from many previous studies that average height is 174cm with standard error 5cm
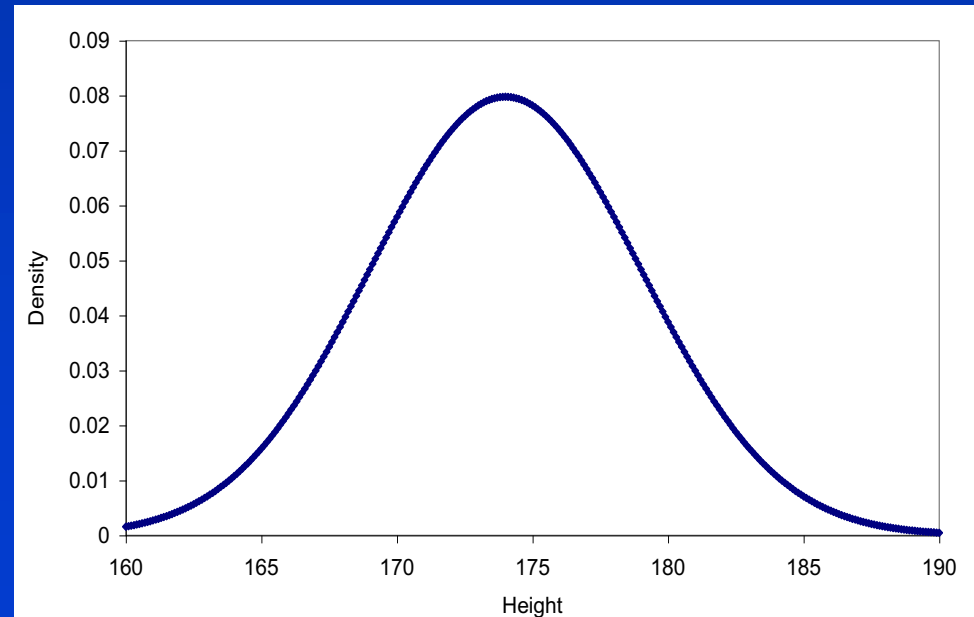
$$y = \text{average height} + e$$

Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x) P(x)$$
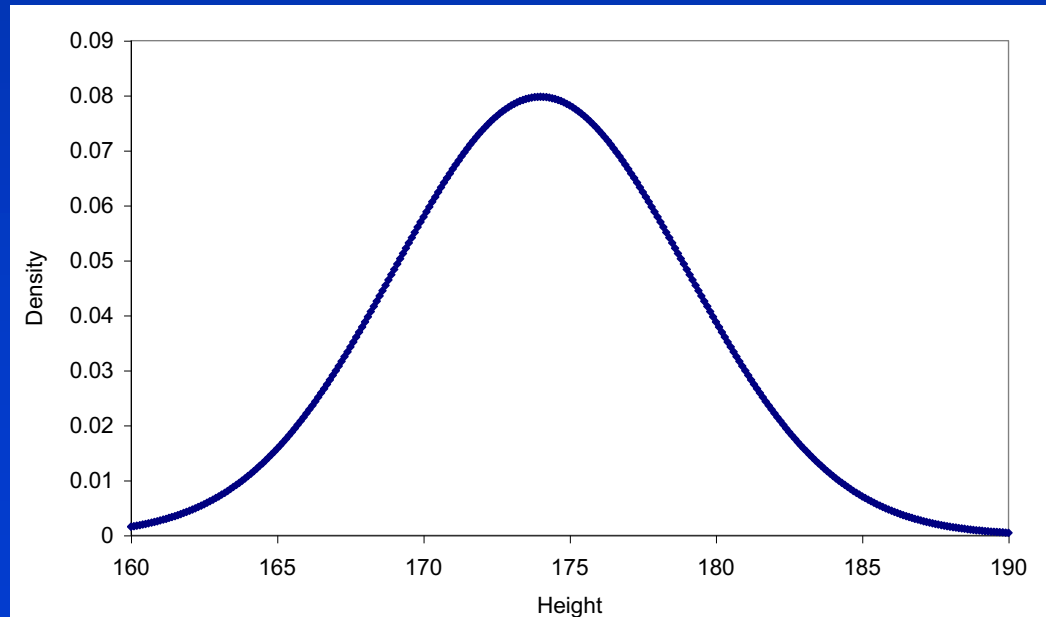
Prior probability of x (average height)

Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x)P(x)$$

Prior probability of x (average height)

From the data……

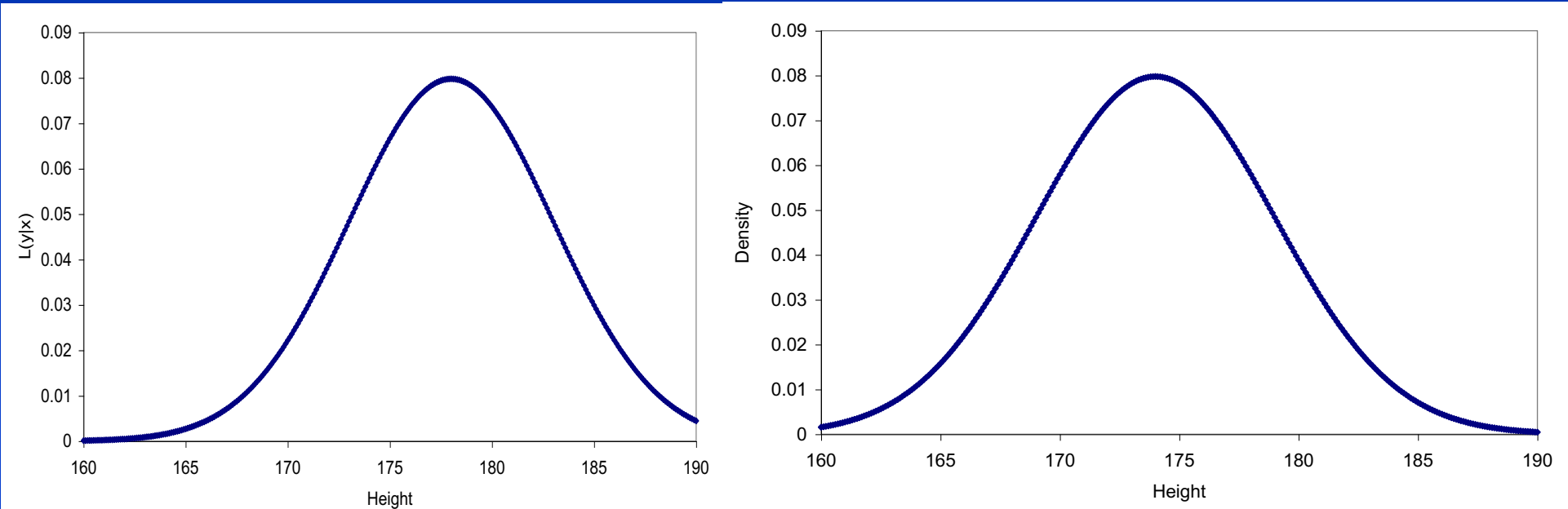$$\overline{x} = 178$$

$$s.e = 5$$

Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x)P(x)$$

Likelihood of data (y) given height x, most likely x = 178cm

Prior probability of x (average height)

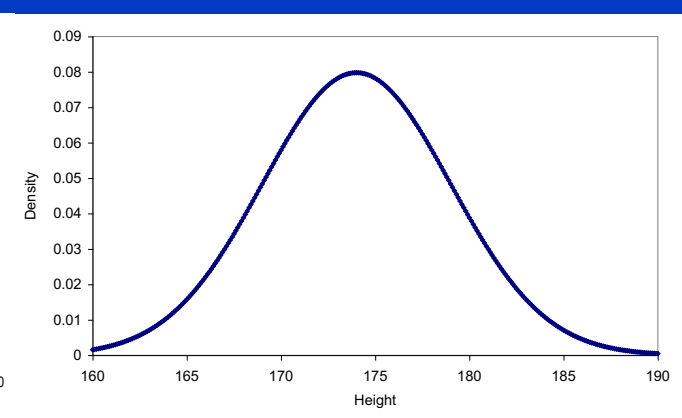# Bayesian methods

- Bayes theorem

$$P(x \mid y) \propto P(y \mid x)P(x)$$

P(x|y) mean = 176cm   L(y|x)   P(x)
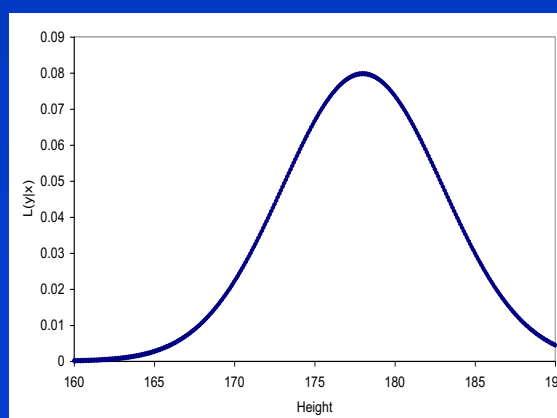
Bayesian methods

- Bayes theorem
- Less certainty about prior information? Use *less* informative (flat) prior

$$P(x \mid y) \propto P(y \mid x)P(x)$$

L(y|x)   P(x)

Bayesian methods

- Bayes theorem
- Less certainty about prior information? Use *less* informative (flat) prior

$$P(x \mid y) \propto P(y \mid x)P(x)$$

P(x|y) mean = 178cm   L(y|x)   P(x)
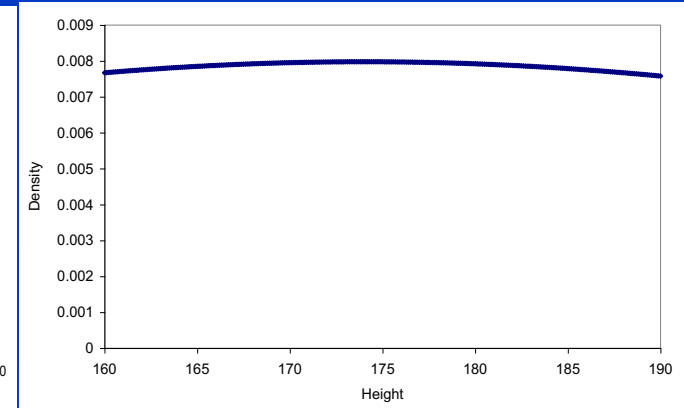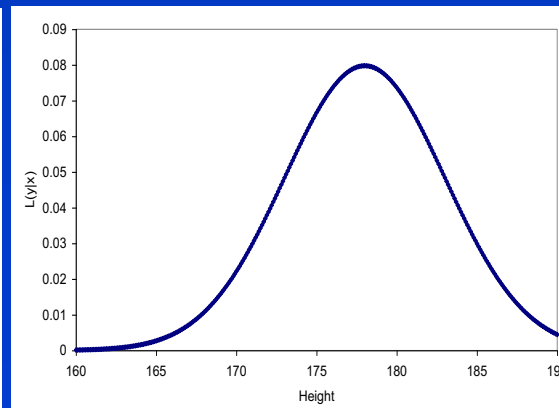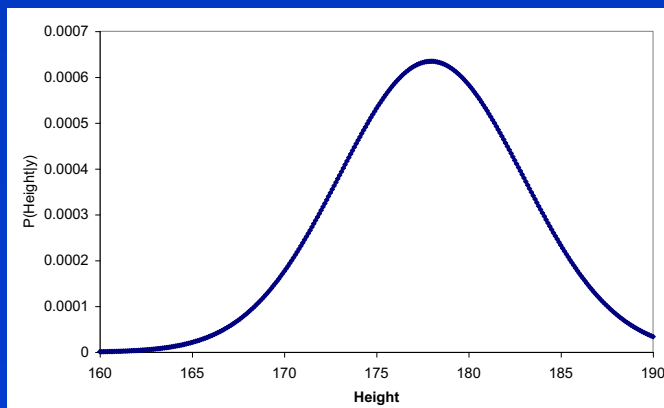
Bayesian methods

- Bayes theorem
- More certainty about prior information? Use *more* informative prior

$$P(x \mid y) \propto P(y \mid x)P(x)$$

L(y|x)   P(x)
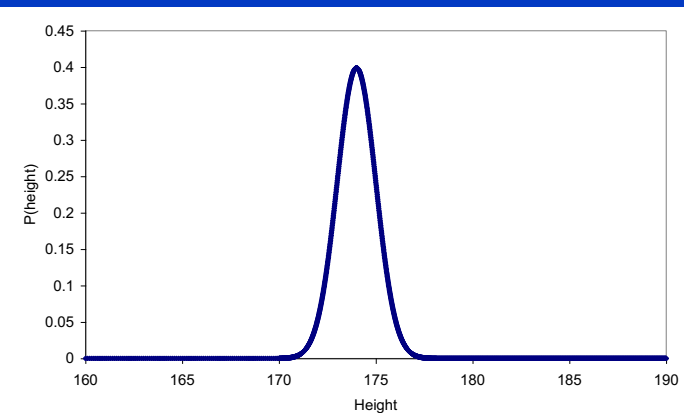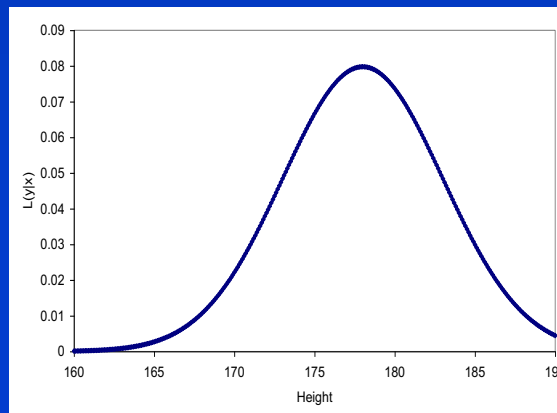
Bayesian methods

- Bayes theorem
- More certainty about prior information? Use *more* informative prior
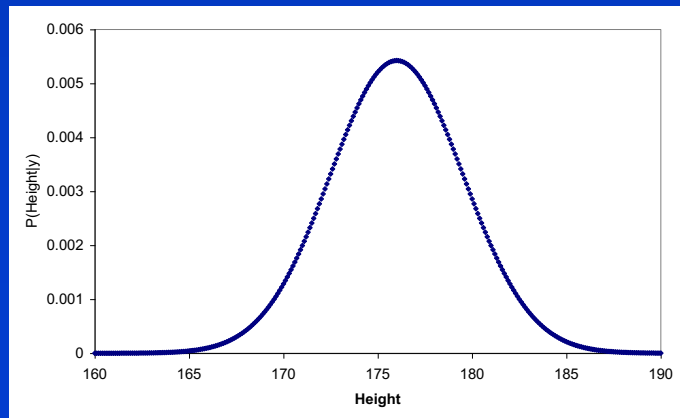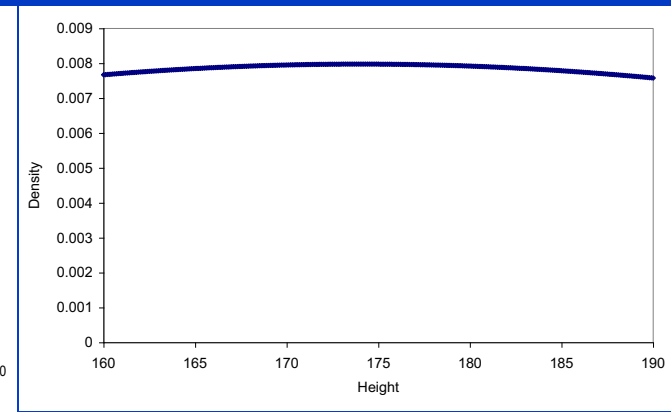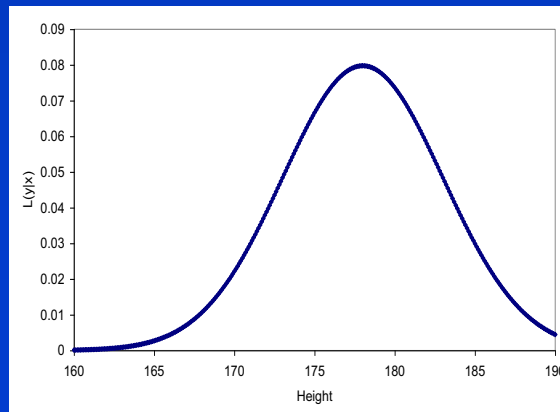
$$P(x\,|\,y) \propto P(y\,|\,x)P(x)$$

P(x|y) mean = 174.5cm   L(y|x)   P(x)
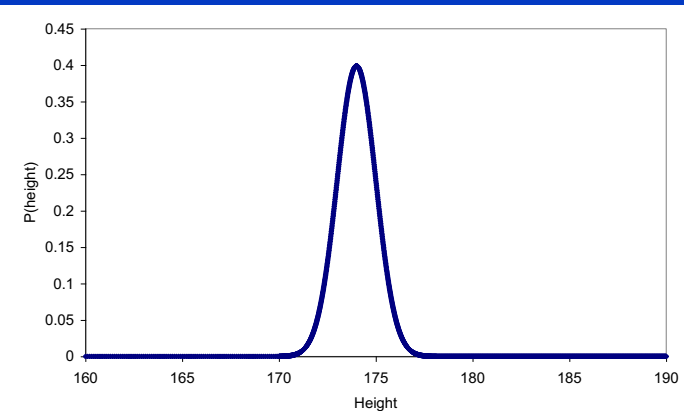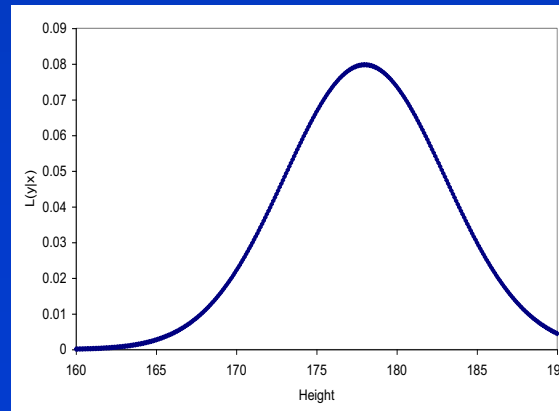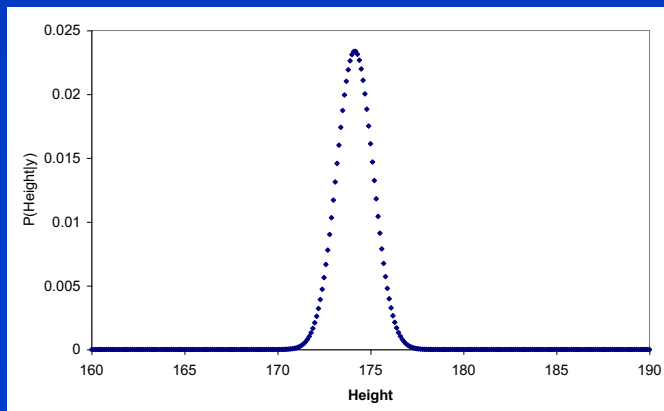
Bayesian methods for Genomic Prediction

- Alternative assumptions regarding the distribution of SNP effects

- Introduction to Bayesian methods

- Genomic prediction with Bayesian methods

- Comparison of accuracy of methods

BayesC

- $y = 1_n\mu + X\beta + e$

$$\beta_j \begin{cases} \sim N(0, \sigma_\beta^2) & \text{with probability } \pi \\ = 0 & \text{with probability } 1 - \pi \end{cases}$$

P(Height|y)

Height
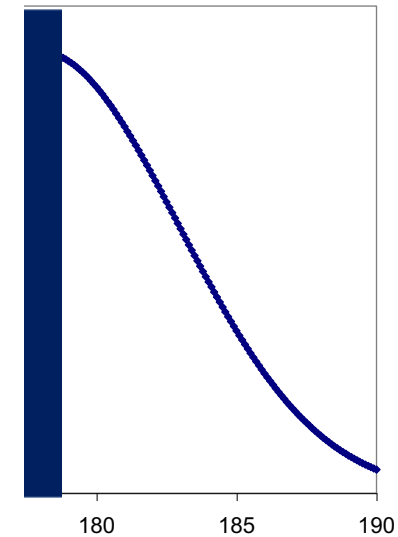
$$P(\boldsymbol{\beta}, \mu | \boldsymbol{y}) \propto P(\boldsymbol{y} | \boldsymbol{\beta}, \mu) P(\boldsymbol{\beta}, \mu)$$

BayesC

- $y = 1_n\mu + X\beta + e$

$$\beta_j \begin{cases} \sim N(0, \sigma_\beta^2) & \text{with probability } \pi \\ = 0 & \text{with probability } 1 - \pi \end{cases}$$

$$P(\boldsymbol{\beta}, \mu | \boldsymbol{y}) \propto P(\boldsymbol{y} | \boldsymbol{\beta}, \mu) P(\boldsymbol{\beta}, \mu)$$

The Queensland Alliance for Agriculture and Food Innovation (QAAFI) is a research institute of The University of Queensland (UQ), supported by the Queensland Government.

CRICOS code 00025B

2
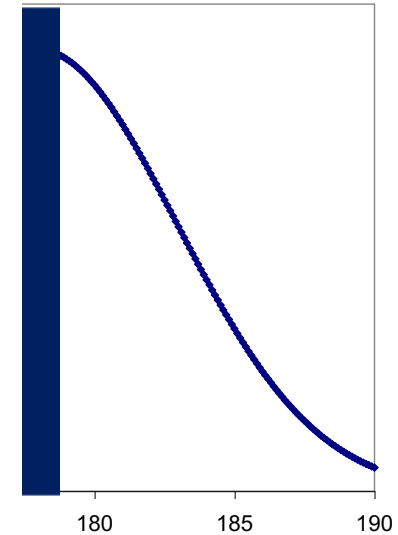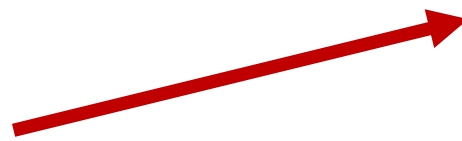
BayesC -> Gibbs Sampling

- Cannot solve directly, as estimates of parameters depend on other parameters -> no closed form solution

- For example, estimate of a SNP effect depends on whether or not the SNP is in the zero variance part of distribution or non-zero variance part of the distribution

- Use Gibbs sampling!

- Sample from posterior distribution of parameter **conditional** on all other parameters

BayesC -> Gibbs Sampling

- Sample from posterior distribution of parameter **conditional** on all other parameters

- For example, for SNP effect β_i
  - First sample if in zero effect or non zero effect part of distribution (δ_i)
  - Then if in non-zero part of the distribution, sample from

$$N\left(\frac{\mathbf{X}_{ij}'\mathbf{y} - \mathbf{X}_{ij}'\mathbf{X}\beta_{(ij=0)} - \mathbf{X}_{ij}'\mathbf{1_n}\mu}{\mathbf{X}_{ij}'\mathbf{X}_{ij} + \sigma_e^2/\sigma_\beta^2}, \sigma_e^2/\left(\mathbf{X}_{ij}'\mathbf{X}_{ij} + \sigma_e^2/\sigma_\beta^2\right)\right)$$

![THE UNIVERSITY OF QUEENSLAND AUSTRALIA]

QAAFI
Queensland Alliance for
Agriculture and Food Innovation

Queensland
Government

BayesC -> Gibbs Sampling

- Sample from posterior distribution of parameter **conditional** on all other parameters

- For example, for SNP effect β_i
  - First sample if in zero effect or non zero effect part of distribution (δ_i)
  - Then if in non-zero part of the distribution, sample from

$$N\left(\frac{\mathbf{X'_{ij}y} - \mathbf{X'_{ij}X}\beta_{(\mathbf{ij=0})} - \mathbf{X'_{ij}1_n}\mu}{\mathbf{X'_{ij}X_{ij}} + \sigma_e^2/\sigma_\beta^2}, \sigma_e^2/\left(\mathbf{X_{ij}'X}_{ij} + \sigma_e^2/\sigma_\beta^2\right)\right)$$

The Queensland Alliance for Agriculture and Food Innovation (QAAFI) is a research institute of The University of Queensland (UQ), supported by the Queensland Government.

CRICOS code 00025B   2

BayesC -> Gibbs Chain

- Set starting values for $(\sigma_e^2, \mu, \delta\ , \beta\ , \sigma_\beta^2, \pi)$

- Then (for many iterations)

  - For each SNP, sample δ_i, β_i conditional on other parameters

  - Sample $\sigma_e^2, \mu, \sigma_\beta^2, \pi$ with updated δ_i, β_i

- Samples reconstruct posterior distributions of parameters

BayesC -> Gibbs Chain

- Set starting values for (σ_e^2, μ, δ, β, σ_β^2, π)

- Then (for many iterations)

  - For each SNP, sample δ_i, β_i conditional on other parameters

  - Sample σ_e^2, μ, σ_β^2, π with updated δ_i, β_i

- Samples reconstruct posterior distributions of parameters

BayesC -> Gibbs Chain
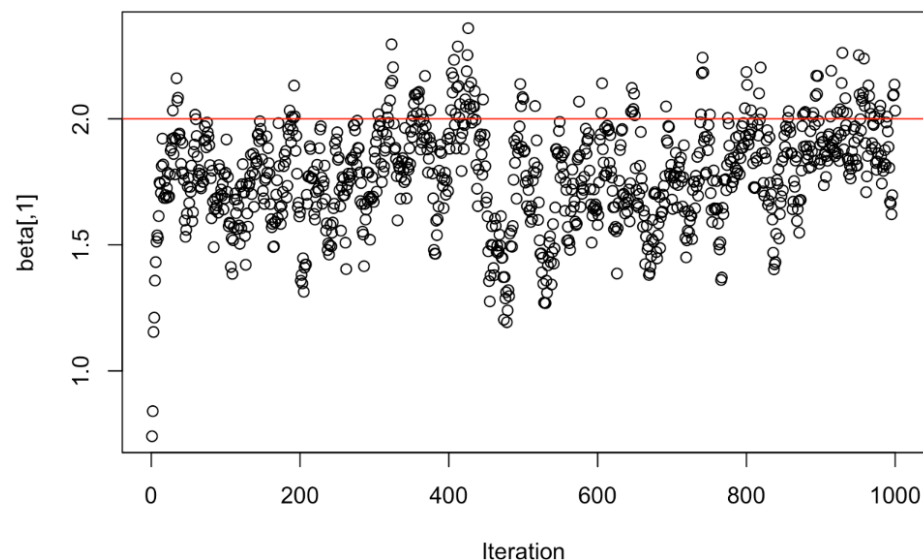
- Set starting values for $(\sigma_e^2, \mu, \delta\ , \beta\ , \sigma_\beta^2, \pi)$

- Then (for many iterations)
  - For each SNP, sample δ_i, β_i conditional on other parameters
  - Sample $\sigma_e^2, \mu, \sigma_\beta^2, \pi$ with updated δ_i, β_i

- Samples reconstruct posterior distributions of parameters

BayesR

- $y = 1_n\mu + X\beta + e$

$$\sigma_\beta^2 = \begin{cases} 0 \cdot \sigma_a^2 \text{ with probability } p_1 \\ 0.0001 \cdot \sigma_a^2 \text{ with probability } p_2 \\ 0.001 \cdot \sigma_a^2 \text{ with probability } p_3 \\ 0.01 \cdot \sigma_a^2 \text{ with probability } p_4 \end{cases}$$

p_{i1} p_{i2} p_{i3} p_{i4}

0
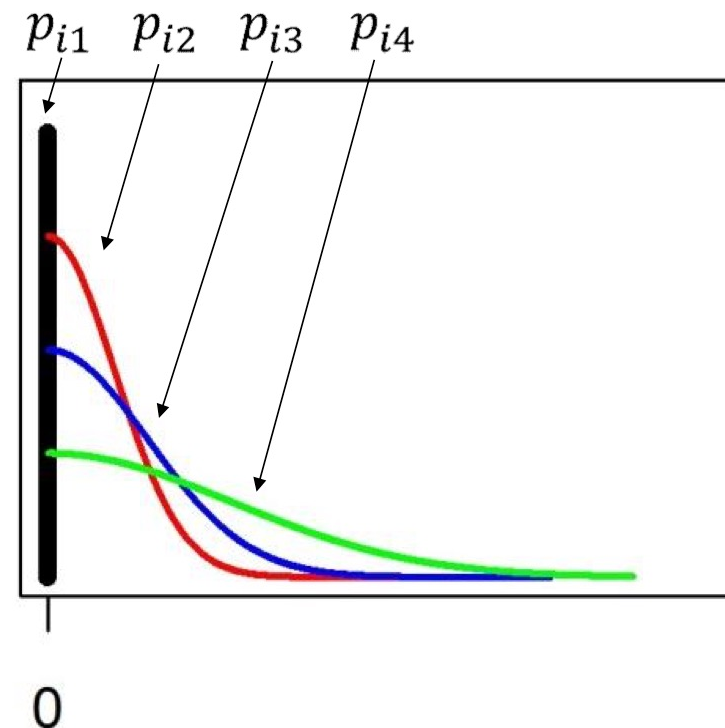
- Each SNP has a probability of being in each of the four distributions
- Use Dirichlet distribution to sample distribution proportions

The Queensland Alliance for Agriculture and Food Innovation (QAAFI) is a research institute of The University of Queensland (UQ), supported by the Queensland Government.

CRICOS code 00025B

Bayesian methods for Genomic Prediction

- Alternative assumptions regarding the distribution of SNP effects

- Introduction to Bayesian methods

- Genomic prediction with Bayesian methods

- Comparison of accuracy of methods

Real Data, 800K

- Reference
  - Holstein = 3049 bulls, 8478 cows
  - Jersey = 770 bulls, 3917 cows

- Validation
  - Holstein = 262 bulls
  - Jersey = 105 bulls
  - *Australian Reds = 114 bulls*

- GEBV with GBLUP, BayesR
- (Kemper et al GSE, 2014)
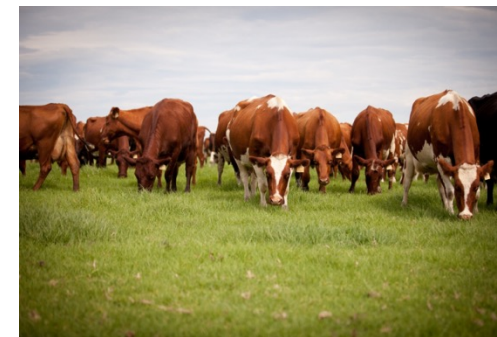
Real Data, 800K

- r(GEBV,DTD)

	Fat	Milk	Protein	Fat%	Protein%	**Average**
Holstein						
GBLUP	0.60	0.59	0.58	0.72	0.83	**0.66**
BAYESR	0.64	0.62	0.57	0.81	0.84	**0.69**
Jersey						
GBLUP	0.56	0.62	0.67	0.64	0.76	**0.65**
BAYESR	0.56	0.69	0.71	0.76	0.79	**0.70**
Australian Reds						
GBLUP	0.20	0.16	0.11	0.32	0.34	**0.22**
BAYES	0.26	0.21	0.13	0.44	0.36	**0.28**
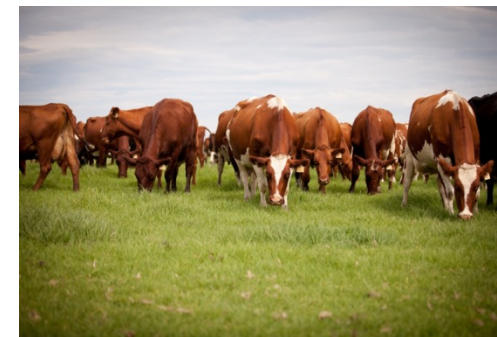
BayesR
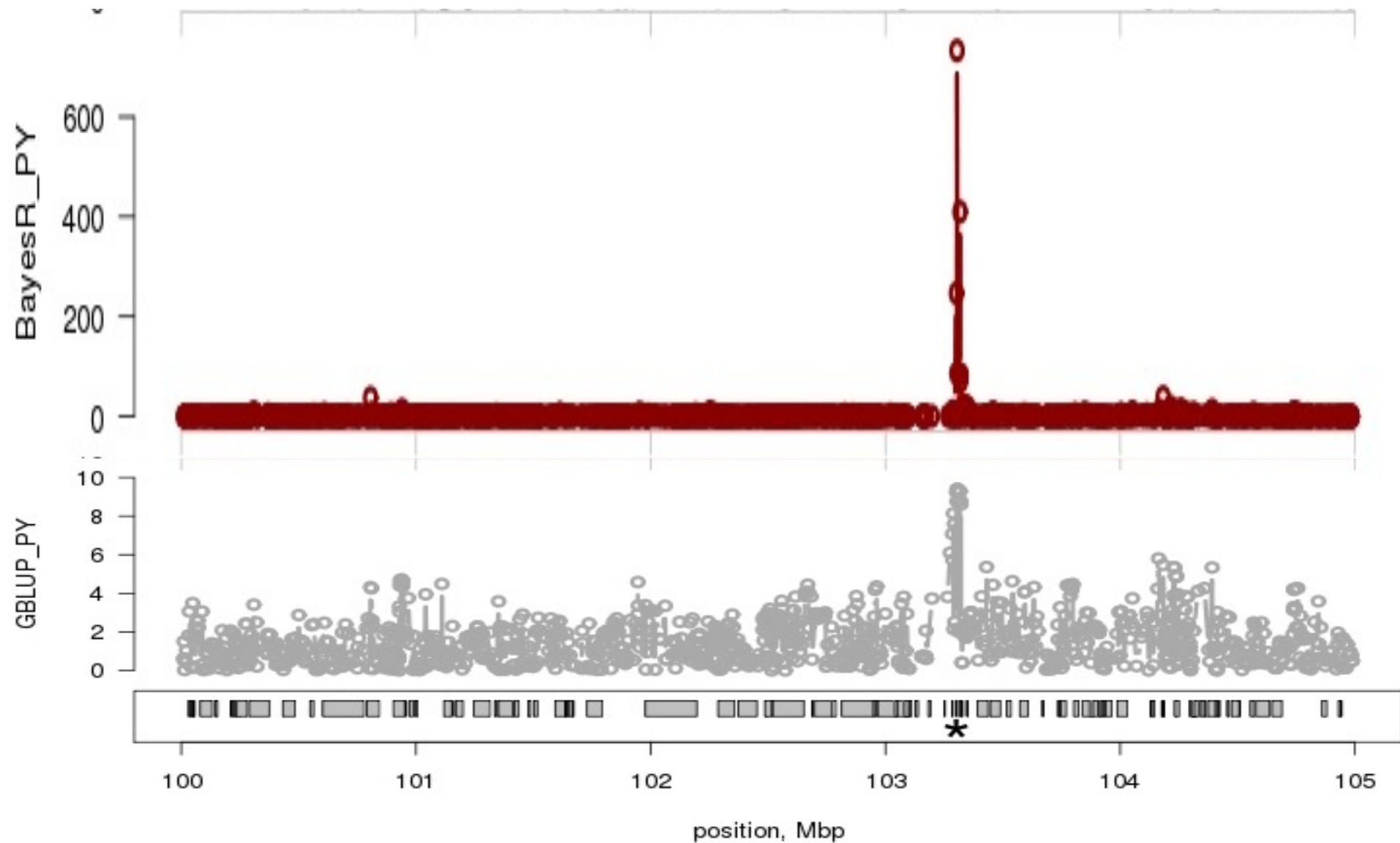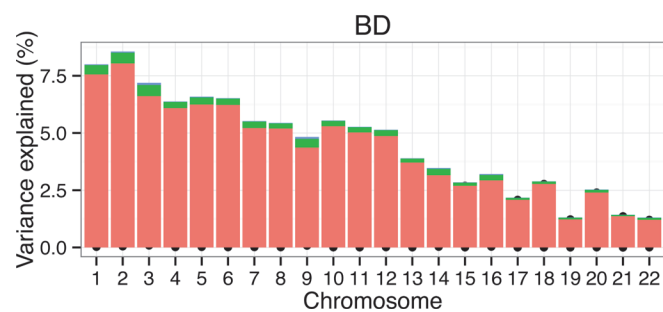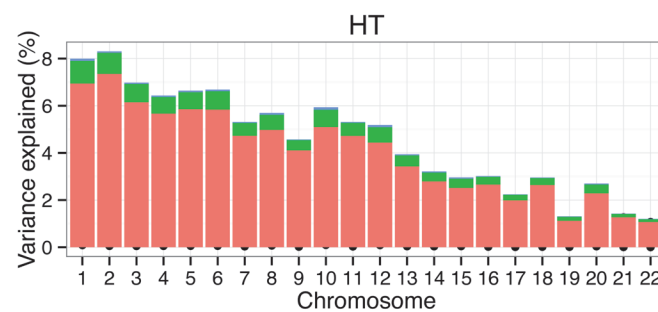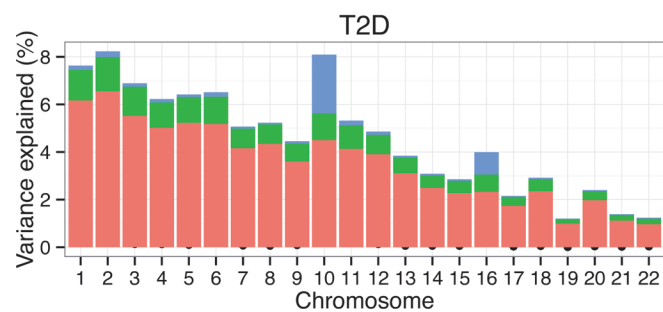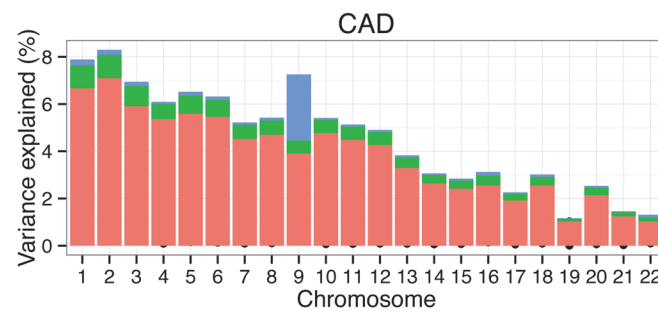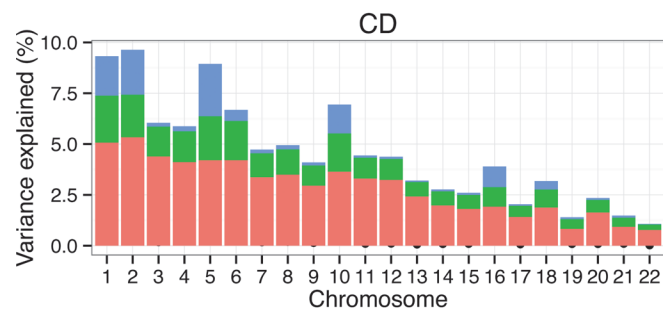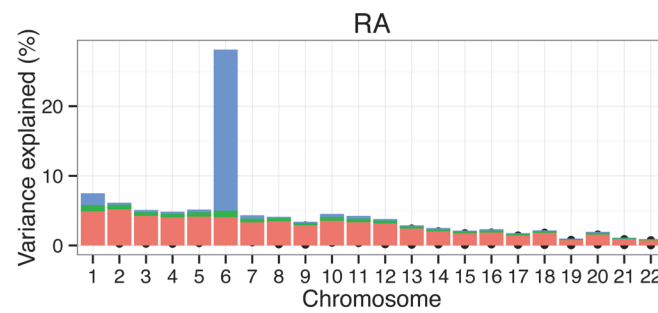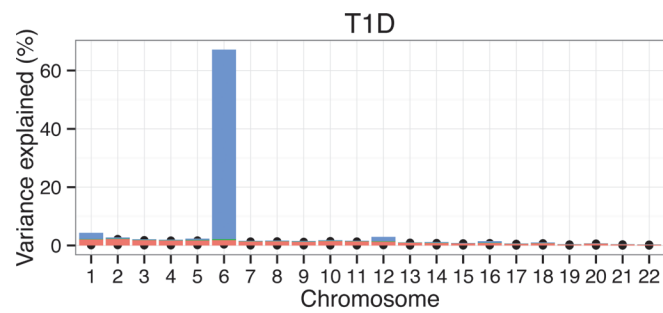
Prediction of disease risk in humans
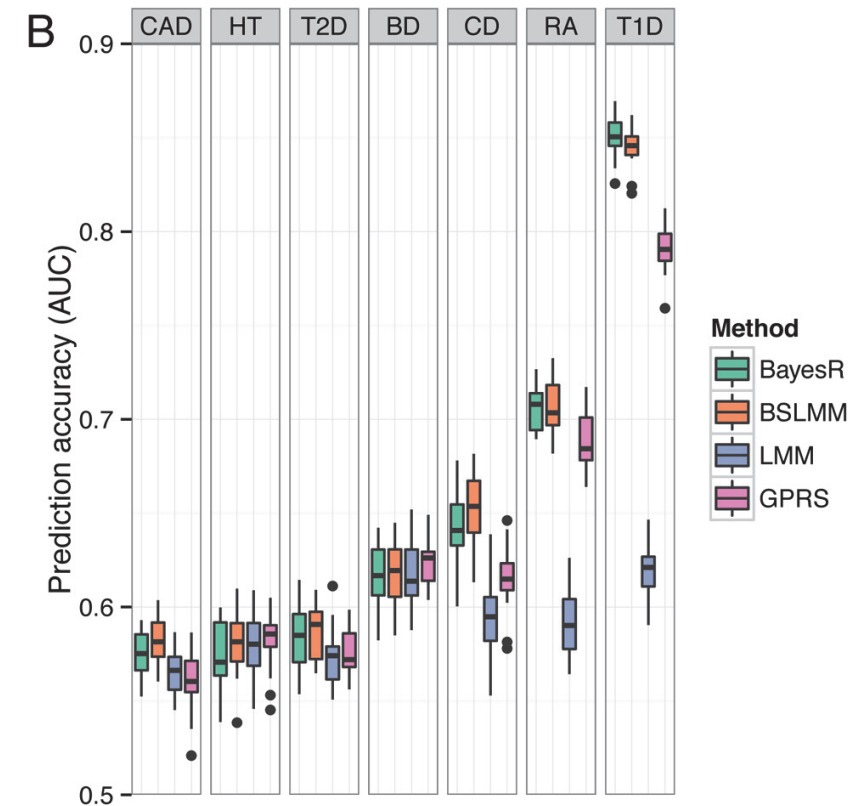
Bayesian methods for Genomic Prediction

Bayesian approach allows us to incorporate prior knowledge in prediction of SNP effects

Bayesian methods can have an advantage when:

QTL of moderate to large effect on the trait (eg Fat%, DGAT1)

Very large numbers of SNP (800K, sequence) -> set some SNP effects to zero

Integrates genomic prediction and QTL fine-mapping

Reference

BayesA, BayesB:

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes[†] and M. E. Goddard[†,‡]

*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, [†]Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and [‡]Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

Meuwissen et al is the paper coined genomic selection.

BayesC:

BMC Bioinformatics

RESEARCH ARTICLE   **Open Access**

Extension of the bayesian alphabet for genomic selection

David Habier[1*], Rohan L Fernando[1], Kadir Kizilkaya[1,2] and Dorian J Garrick[2,3]

BayesR:

Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels

M. Erbe,*[1] B. J. Hayes,†‡§[1,2] L. K. Matukumalli,# S. Goswami,|| P. J. Bowman,†‡ C. M. Reich,†‡ B. A. Mason,†‡ and M. E. Goddard†¶

PLOS GENETICS

RESEARCH ARTICLE

Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model

Gerhard Moser[1*], Sang Hong Lee[1], Ben J. Hayes[2,3], Michael E. Goddard[2,4], Naomi R. Wray[1], Peter M. Visscher[1,5]

Practical 4: Polygenic prediction using Bayesian methods

https://cnsgenomics.com/data/teaching/GNGWS23/model5/Practical4_Bayes.html

Log into the cluster

cd to your working directory in scratch: cd /scratch/[your folder]

You will learn how to run MCMC using the toy example data set in R.

You will use GCTB to run BayesR in the simulated data set based on real genotypes.

❖ Compared to C+PT, does BayesR improve prediction accuracy? Why or why not?

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix..

Full conditionals for single-site Gibbs

▶ $(\mu | \boldsymbol{y}, \boldsymbol{\alpha}, \sigma_e^2) \sim N(\frac{\boldsymbol{1}'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha})}{n}, \frac{\sigma_e^2}{n})$

▶ $(\alpha_j | \boldsymbol{y}, \mu, \boldsymbol{\alpha}_{j\_}, \sigma_e^2) \sim N(\hat{\alpha}_j, \frac{\sigma_e^2}{c_j})$

   ▶
$$\hat{\alpha}_j = \frac{\boldsymbol{x}_j' \boldsymbol{w}}{c_j}$$

   ▶
$$\boldsymbol{w} = \boldsymbol{y} - \boldsymbol{1}\mu - \sum_{j' \neq j} \boldsymbol{x}_{j'} \alpha_{j'}$$

   ▶
$$c_j = (\boldsymbol{x}_j' \boldsymbol{x}_j + \frac{\sigma_e^2}{\sigma_\alpha^2})$$

▶ $(\sigma_e^2 | \boldsymbol{y}, \mu, \boldsymbol{\alpha}) \sim [(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\theta}) + \nu_e S_e^2] \chi^{-2}_{(\nu_e + n)}$

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix.

Derive: full conditional for α_j

From Bayes' Theorem,

$$f(\alpha_j | \boldsymbol{y}, \mu, \boldsymbol{\alpha}_{j_-}, \sigma_e^2) = \frac{f(\alpha_j, \boldsymbol{y}, \mu, \boldsymbol{\alpha}_{j_-}, \sigma_e^2)}{f(\boldsymbol{y}, \mu, \boldsymbol{\alpha}_{j_-}, \sigma_e^2)}$$

$$\propto f(\boldsymbol{y} | \alpha_j, \mu, \boldsymbol{\alpha}_{j_-}, \sigma_e^2) f(\alpha_j) f(\mu, \boldsymbol{\alpha}_{j_-}, \sigma_e^2)$$

$$\propto (\sigma_e^2)^{-n/2} \exp\{-\frac{(\boldsymbol{w} - \boldsymbol{x}_j \alpha_j)'(\boldsymbol{w} - \boldsymbol{x}_j \alpha_j)}{2\sigma_e^2}\} (\sigma_\alpha^2)^{-1/2} \exp\{-\frac{\alpha_j^2}{2\sigma_\alpha^2}\}$$

where

$$\boldsymbol{w} = \boldsymbol{y} - \boldsymbol{1}\mu - \sum_{j \neq j'} \boldsymbol{x}_{j'} \alpha_{j'}$$

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix.

Derive: full conditional for α_j

The exponential terms in the joint density can be written as:

$$-\frac{1}{2\sigma_e^2}\{\boldsymbol{w}'\boldsymbol{w} - 2\boldsymbol{x}_j'\boldsymbol{w}\alpha_j + [\boldsymbol{x}_j'\boldsymbol{x}_j + \frac{\sigma_e^2}{\sigma_\alpha^2}]\alpha_j^2\}$$

Completing the square in this expression with respect to α_j
gives

$$-\frac{1}{2\sigma_e^2}\{c_j(\alpha_j - \hat{\alpha}_j)^2 + \boldsymbol{w}'\boldsymbol{w} - c_j\hat{\alpha}_j^2\}$$

where

$$\hat{\alpha}_j = \frac{\boldsymbol{x}_j\boldsymbol{w}}{c_j}$$

So,

$$f(\alpha_j|\boldsymbol{y}, \mu, \boldsymbol{\alpha}_{j_-}, \sigma_e^2) \propto \exp\{-\frac{(\alpha_j - \hat{\alpha}_j)^2}{2\frac{\sigma_e^2}{c_j}}\}$$

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix. In this slide, treat r_j as α_j.

Full-conditional for δ_j:

$$\Pr(\delta_j | \boldsymbol{y}, \mu, \boldsymbol{\alpha}_{-j}, \boldsymbol{\delta}_{-j}, \sigma_\alpha^2, \sigma_e^2, \pi) =$$
$$\Pr(\delta_j | r_j, \boldsymbol{\theta}_{j\_})$$

$$\Pr(\delta_j | r_j, \boldsymbol{\theta}_{j\_}) = \frac{f(\delta_j, r_j | \boldsymbol{\theta}_{j\_})}{f(r_j | \boldsymbol{\theta}_{j\_})}$$

$$= \frac{f(r_j | \delta_j, \boldsymbol{\theta}_{j\_}) \Pr(\delta_j | \pi)}{f(r_j | \delta_j = 0, \boldsymbol{\theta}_{j\_})\pi + f(r_j | \delta_j = 1, \boldsymbol{\theta}_{j\_})(1 - \pi)}$$

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix.

Full conditional for σ_e^2

From Bayes' theorem,

$$f(\sigma_e^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}) = \frac{f(\sigma_e^2, \boldsymbol{y}, \mu, \boldsymbol{\alpha})}{f(\boldsymbol{y}, \mu, \boldsymbol{\alpha})}$$

$$\propto f(\boldsymbol{y}|\sigma_e^2, \mu, \boldsymbol{\alpha})f(\sigma_e^2)f(\mu, \boldsymbol{\alpha})$$

where

$$f(\boldsymbol{y}|\sigma_e^2, \mu, \boldsymbol{\alpha}) \propto (\sigma_e^2)^{-n/2} \exp\{-\frac{(\boldsymbol{w} - \boldsymbol{x}_j\alpha_j)'(\boldsymbol{w} - \boldsymbol{x}_j\alpha_j)}{2\sigma_e^2}\}$$

and

$$f(\sigma_e^2) = \frac{(S_e^2\nu_e/2)^{\nu_e/2}}{\Gamma(\nu/2)}(\sigma_e^2)^{-(2+\nu_e)/2} \exp(-\frac{\nu_e S_e^2}{2\sigma_e^2})$$

The Queensland Alliance for Agriculture and Food Innovation (QAAFI) is a research institute of The University of Queensland (UQ), supported by the Queensland Government.

CRICOS code 00025B

4

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

QAAFI
Queensland Alliance for
Agriculture and Food Innovation

Queensland
Government

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix.

Full conditional for σ_e^2

So,

$$f(\sigma_e^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}) \propto (\sigma_e^2)^{-(2+n+\nu_e)/2} \exp(-\frac{SSE + \nu_e S_e^2}{2\sigma_e^2})$$

where

$$SSE = (\boldsymbol{w} - \boldsymbol{x}_j \alpha_j)'(\boldsymbol{w} - \boldsymbol{x}_j \alpha_j)$$

So,

$$f(\sigma_e^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}) \sim \tilde{\nu}_e \tilde{S}_e^2 \chi_{\tilde{\nu}_e}^{-2}$$

where

$$\tilde{\nu}_e = n + \nu_e; \quad \tilde{S}_e^2 = \frac{SSE + \nu_e S_e^2}{\tilde{\nu}_e}$$

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

QAAFI
Queensland Alliance for
Agriculture and Food Innovation

Queensland
Government

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix.

Full conditional for σ_α^2

This can be written as

$$f(\sigma_\alpha^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2) \propto f(\boldsymbol{y}|\sigma_\alpha^2, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2)f(\sigma_\alpha^2, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2)$$

But, can see that

$$f(\boldsymbol{y}|\sigma_\alpha^2, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2) \propto f(\boldsymbol{y}|\mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2)$$

So,

$$f(\sigma_\alpha^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2) \propto f(\sigma_\alpha^2, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2)$$

Note that σ_α^2 appears only in $f(\boldsymbol{\alpha}|\sigma_\alpha^2)$ and $f(\sigma_\alpha^2)$:

$$f(\boldsymbol{\alpha}|\sigma_\alpha^2) \propto (\sigma_\alpha^2)^{-k/2} \exp\{-\frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}}{2\sigma_\alpha^2}\}$$

and

$$f(\sigma_\alpha^2) \propto (\sigma_\alpha^2)^{-(\nu_\alpha+2)/2} \exp\{\frac{\nu_\alpha S_\alpha^2}{2\sigma_\alpha^2}\}$$

The Queensland Alliance for Agriculture and Food Innovation (QAAFI) is a research institute of The University of Queensland (UQ), supported by the Queensland Government.

CRICOS code 00025B

4

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

QAAFI
Queensland Alliance for
Agriculture and Food Innovation

Queensland
Government

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix.

Full conditional for σ_α^2

Combining these two densities gives:

$$f(\sigma_\alpha^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2) \propto (\sigma_\alpha^2)^{-(k+\nu_\alpha+2)/2} \exp\{\frac{\boldsymbol{\alpha}'\boldsymbol{\alpha} + \nu_\alpha S_\alpha^2}{2\sigma_\alpha^2}\}$$

So,

$$(\sigma_\alpha^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma_e^2) \sim \tilde{\nu}_\alpha \tilde{S}_\alpha^2 \chi_{\tilde{\nu}_\alpha}^{-2}$$

where

$$\tilde{\nu}_\alpha = k + \nu_\alpha$$

and

$$\tilde{S}_\alpha^2 = \frac{\boldsymbol{\alpha}'\boldsymbol{\alpha} + \nu_\alpha S_\alpha^2}{\tilde{\nu}_\alpha}$$

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

QAAFI
Queensland Alliance for
Agriculture and Food Innovation

Queensland
Government

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix.

Hyper parameter: S_α^2

If σ^2 is distributed as a scaled, inverse chi-square random variable with scale parameter S^2 and degrees of freedom ν

$$\mathsf{E}(\sigma^2) = \frac{\nu S^2}{\nu - 2}$$

Recall that under some assumptions

$$\sigma_\alpha^2 = \frac{V_a}{\sum_j 2p_j q_j}$$

So, we take

$$S_\alpha^2 = \frac{(\nu_\alpha - 2) V_a}{\nu_\alpha k (1 - \pi) 2\overline{pq}}$$

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

QAAFI
Queensland Alliance for
Agriculture and Food Innovation

Queensland
Government

Appendix - Full conditional distributions for BayesC model parameters

The SNP effect β_j is denoted as α_j in this Appendix.

Full conditional for π

Using Bayes' theorem,

$$f(\pi|\boldsymbol{\delta}, \mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_e^2, \boldsymbol{y}) \propto f(\boldsymbol{y}|\pi, \boldsymbol{\delta}, \mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_e^2) f(\pi, \boldsymbol{\delta}, \mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_e^2)$$

But,

- Conditional on $\boldsymbol{\delta}$ the likelihood is free of π
- Further, π only appears in probability of the vector of bernoulli variables: $\boldsymbol{\delta}$

Thus,

$$f(\pi|\boldsymbol{\delta}, \mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_e^2, \boldsymbol{y}) = \pi^{(k-m)}(1-\pi)^m$$

where $m = \boldsymbol{\delta}'\boldsymbol{\delta}$, and k is the number of markers. Thus, π is sampled from a beta distribution with $a = k - m + 1$ and $b = m + 1$.