THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Bayesian Methods for PGS prediction

Jian Zeng
j.zeng@uq.edu.au

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
Institute for Molecular Bioscience

Program in Complex Trait Genomics

Slides credit: Ben Hayes

Best linear unbiased prediction (BLUP)

Linear mixed model

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

BLUP solutions

$$\begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n}'\mathbf{1_n} & \mathbf{1_n'X} \\ \mathbf{X'1_n} & \mathbf{X'X} + \mathbf{I}\lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{X'y} \end{bmatrix}$$

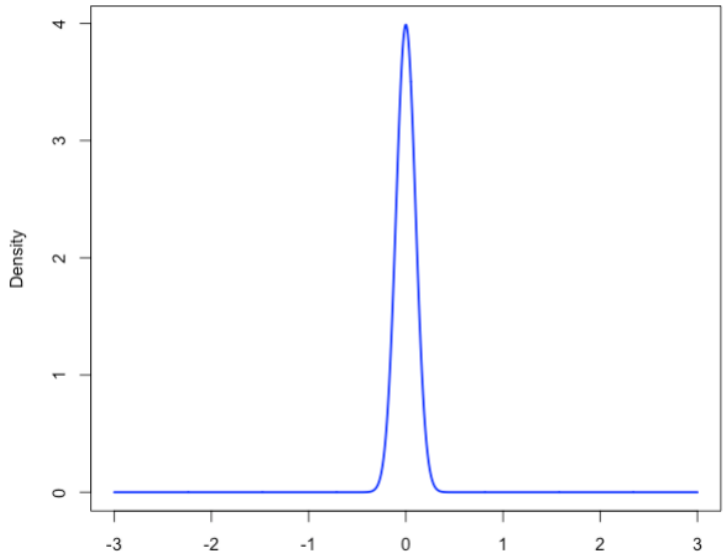
I = identity matrix (dimensions *m* x *m*)

$\lambda = \sigma_e^2 \,/\, \sigma_\beta^2$

- How to determine the shrinkage parameter λ?
  - ➢ Estimate the variance components using GREML
  - ➢ Cross-validation with various input values for λ

- Assumes SNPs effects are:
  - all non-zero
  - very small
  - normally distributed

How realistic is it?



Bayesian methods

- Bayesian methods can estimate all parameters including SNP effects simultaneously

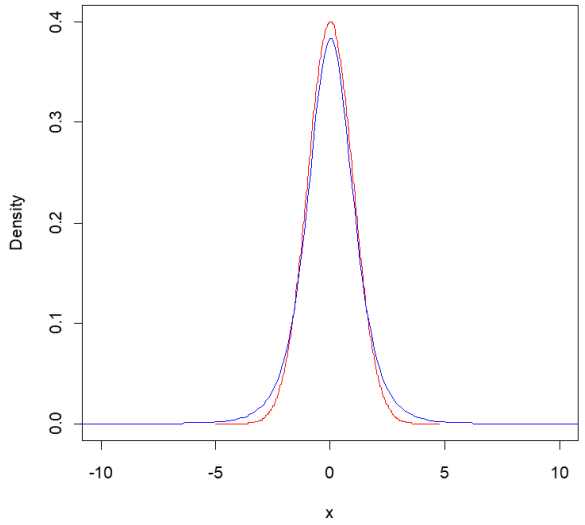
- Allow alternative assumptions regarding the distribution of SNP effects

What are alternative distributions that make sense?

Assumptions for SNP effect distribution

Alternative distributions

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



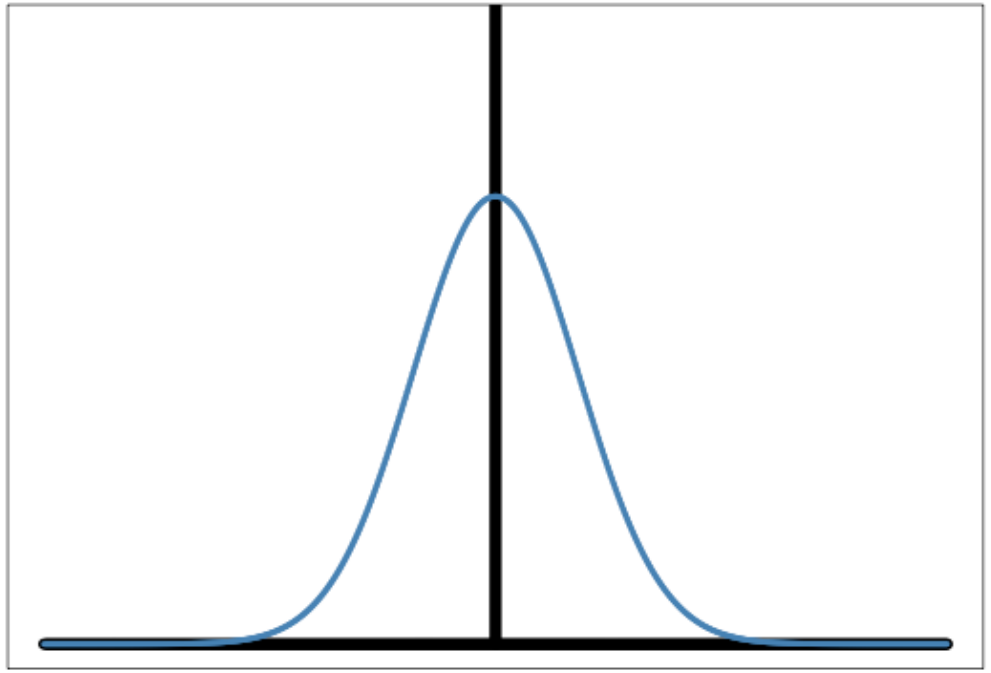
Assumptions for SNP effect distribution

Alternative distributions

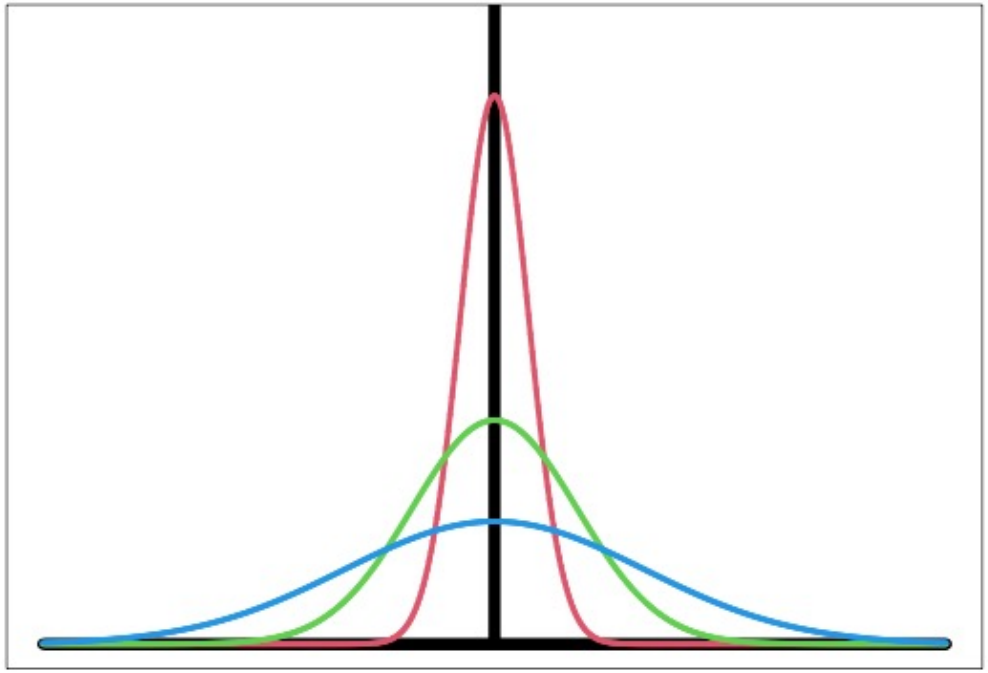
Assumption	Distribution of SNP effects	Method
Small number of moderate to large effects, many small effects	Students t	BayesA
Small number of moderate to large effects, many zero effects	Mixture, spike at zero, Students t	BayesB
Small number of small effects, many zero effects	Mixture, spike at zero,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

Assumptions for SNP effect distribution

BayesC



BayesR

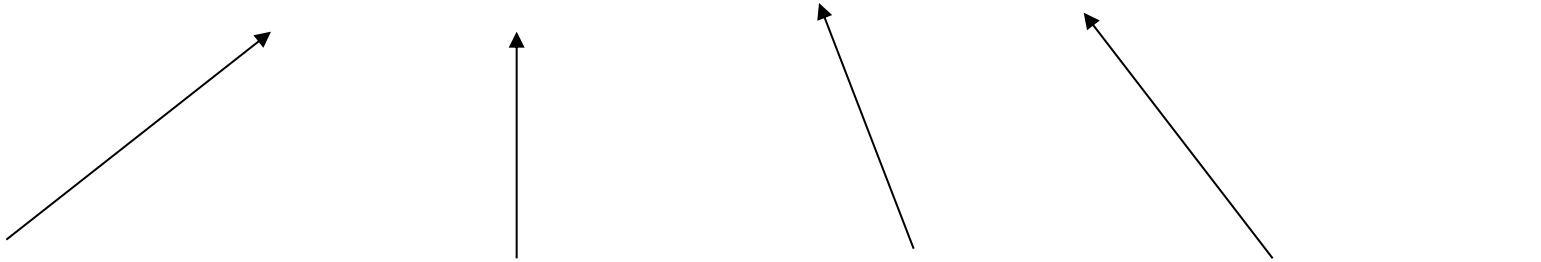


How to incorporate this prior knowledge in the estimation of SNP effects?

Introduction 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)

Prior probability of x

Introduction to Bayesian methods

Consider an experiment where we measure height of 10 people to estimate average height

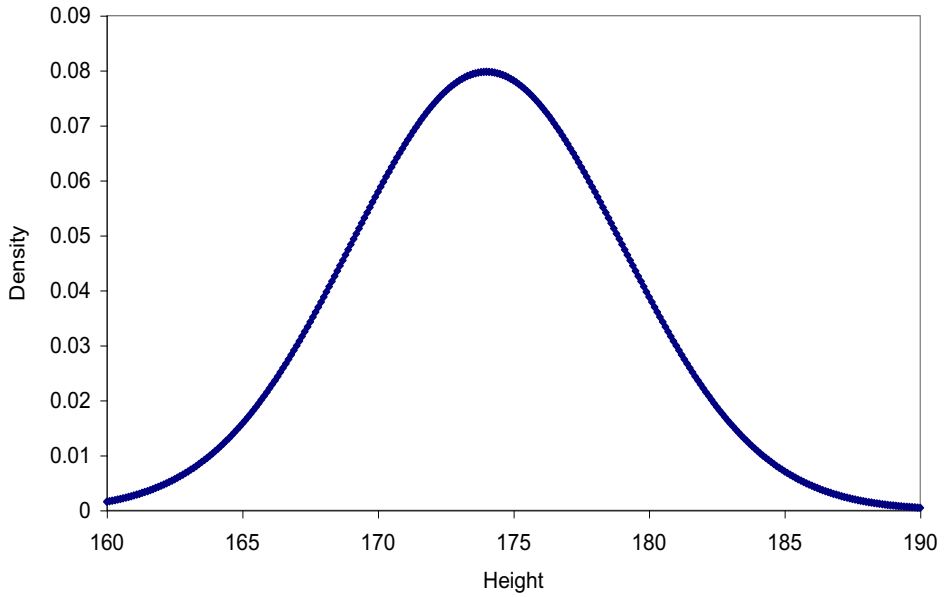
We want to use prior knowledge from many previous studies that average height is 174cm with standard error 5cm

$$y = \text{average height} + e$$

Bayes theorem

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

Prior probability of x (average height)



Bayes theorem

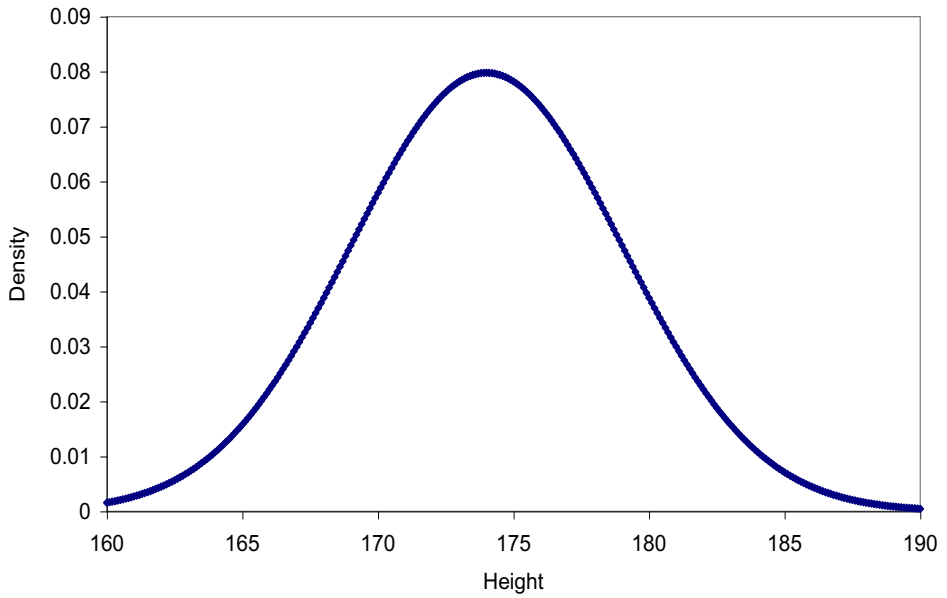
$$P(x \mid y) \propto P(y \mid x)P(x)$$

From the data……

$$\bar{x} = 178$$

$$s.e = 5$$

Prior probability of x (average height)

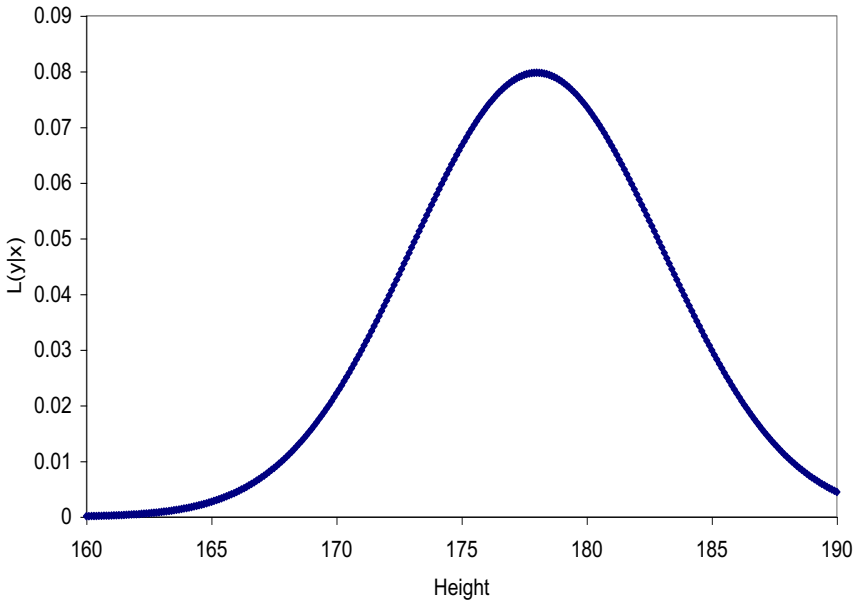


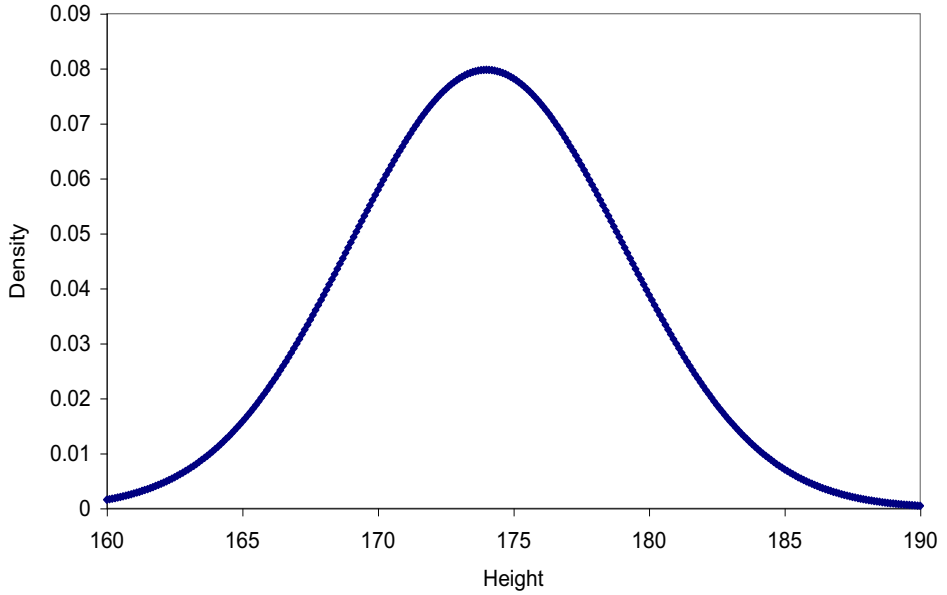
Bayes theorem

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

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

Prior probability of x (average height)





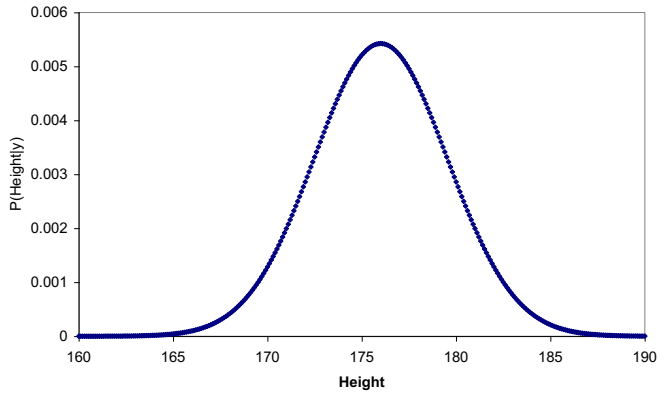
Bayes theorem

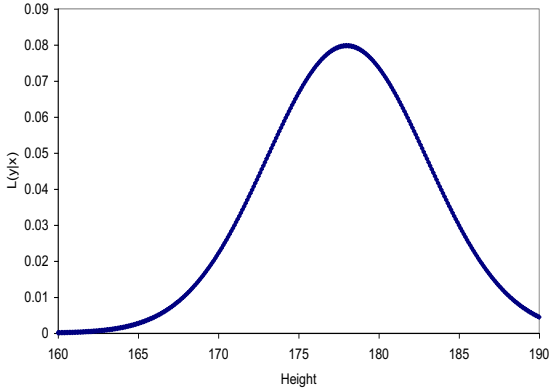
$$P(x \mid y) \propto P(y \mid x)P(x)$$

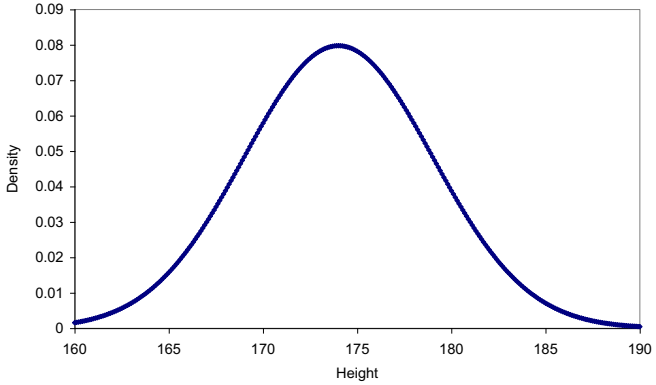
P(x|y) mean = 176cm

L(y|x)

P(x)







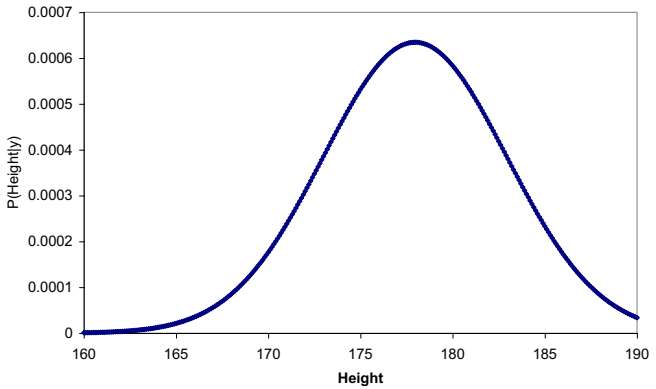
Less certainty about prior information? Use *less* informative (flat) prior

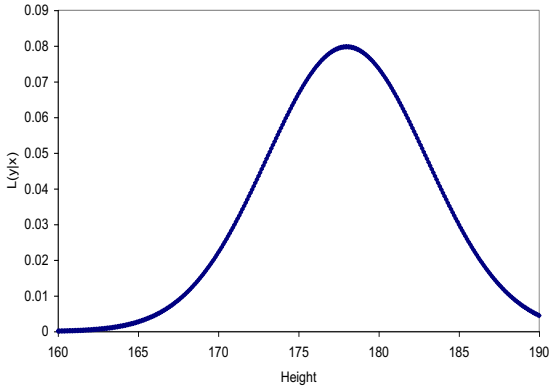
$$P(x \mid y) \propto P(y \mid x)P(x)$$

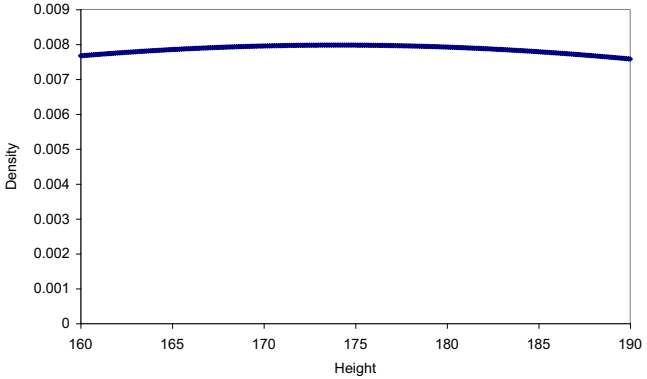
P(x|y) mean = 178cm

L(y|x)

P(x)







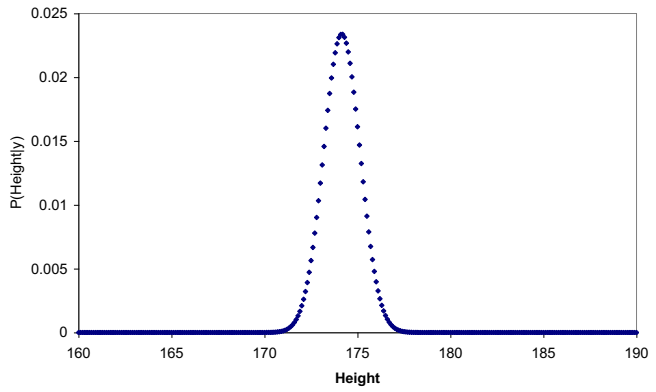
More certainty about prior information? Use *more* informative prior

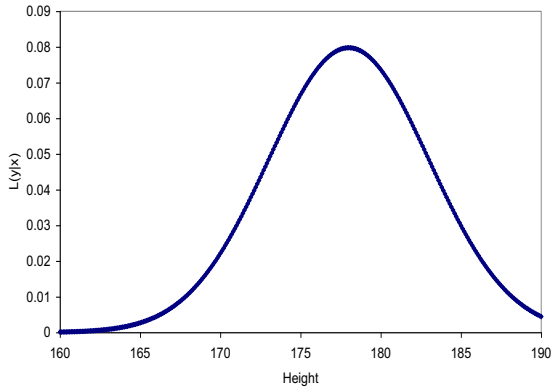
$$P(x \mid y) \propto P(y \mid x)P(x)$$

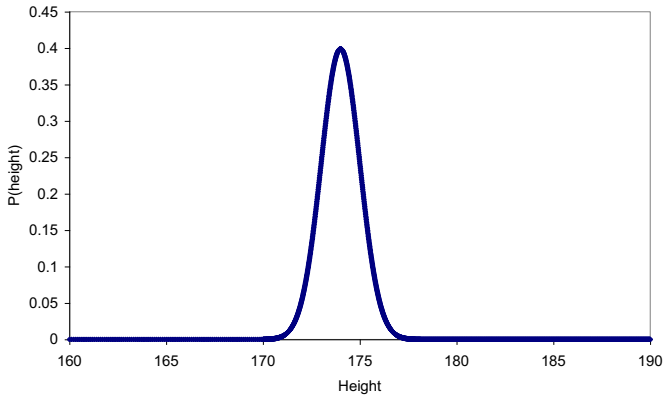
P(x|y) mean = 174.5cm

L(y|x)

P(x)





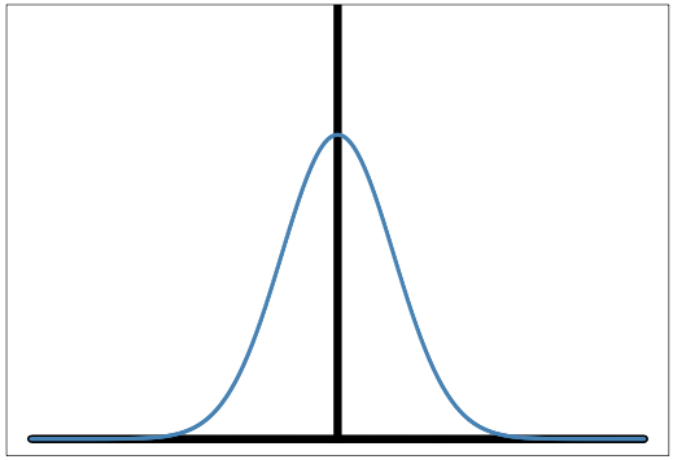


PGS prediction with Bayesian methods

Model

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

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



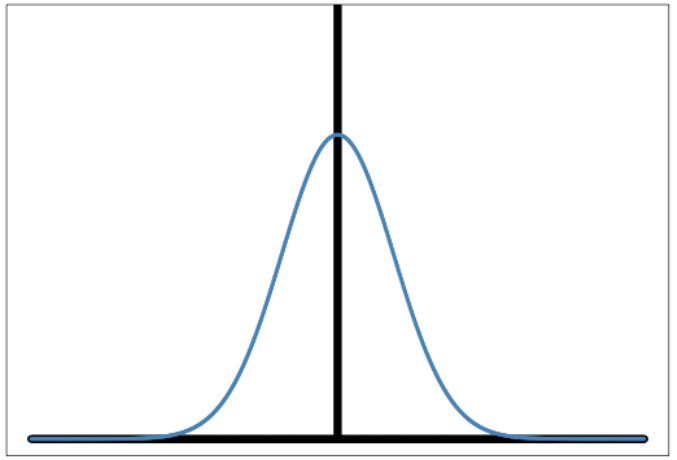
BLUP is a special case of BayesC when π = 1

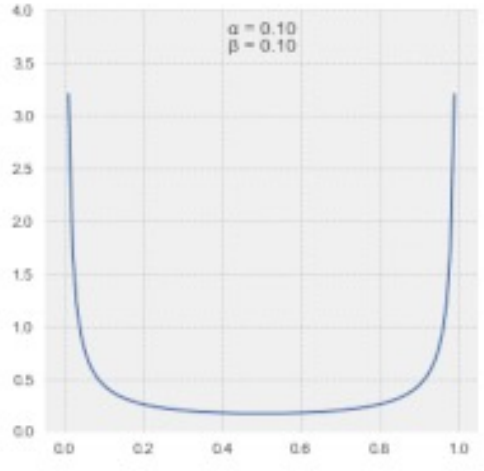
When π < 1, it introduces extra shrinkage on the SNP effect

Model

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

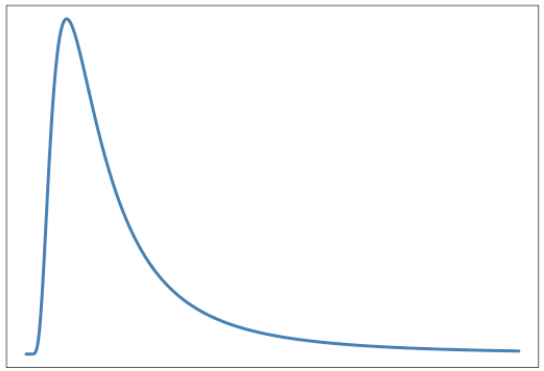
$$\beta_j \begin{cases} \sim N(0, \sigma^2_\beta) & \text{with probability } \pi \\ = 0 & \text{with probability } 1 - \pi \end{cases}$$





$$\sigma^2_\beta \sim \chi^{-2}(\upsilon, S)$$

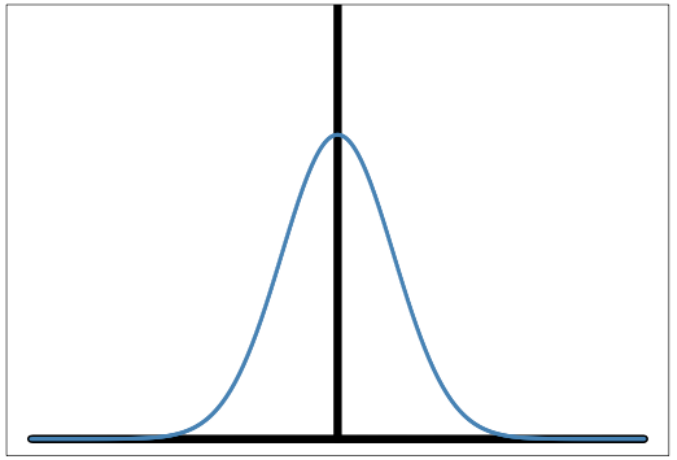
$$\pi \sim Beta(1,1)$$



Model

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

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



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

"conjugate prior"

The posterior distribution is also a point-normal mixture

Posterior joint distribution

$$P\left(\mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2 \middle| \boldsymbol{y}\right)$$

$$\propto P\left(\boldsymbol{y} \middle| \mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2\right) P(\mu) P\left(\boldsymbol{\beta} \middle| \sigma_\beta^2, \pi\right) P\left(\sigma_\beta^2\right) P(\pi) P\left(\sigma_\beta^2\right)$$

- 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 it is zero or nonzero component of the distribution
- Use Gibbs sampling!

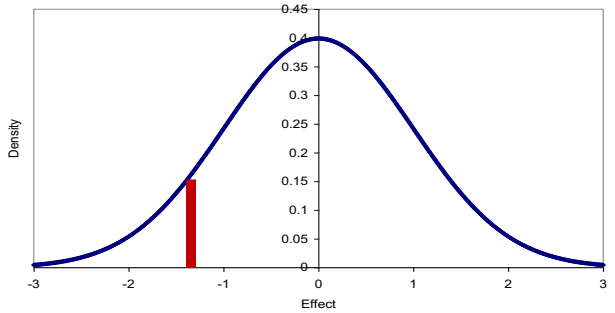
Gibbs Sampling

A technique to sample from posterior distribution of the parameter **conditional** on all other parameters.

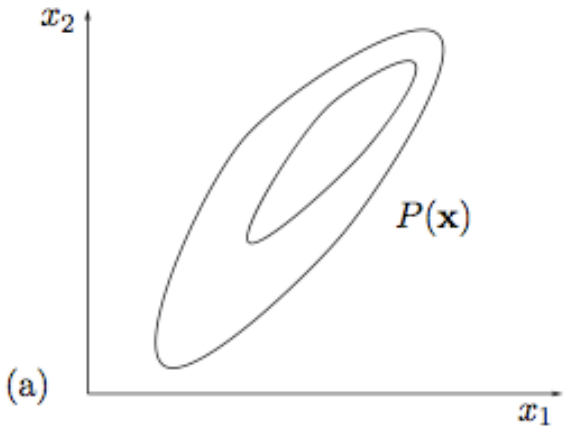
For example, for SNP effect β_i

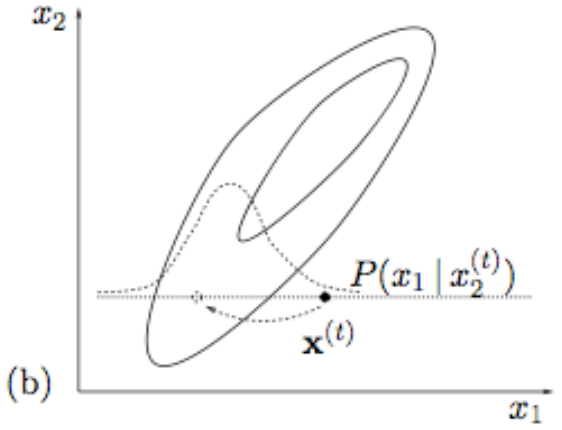
- First sample if in zero or nonzero effect component of distribution (δ_i)
- Then if in nonzero component of the distribution, sample from

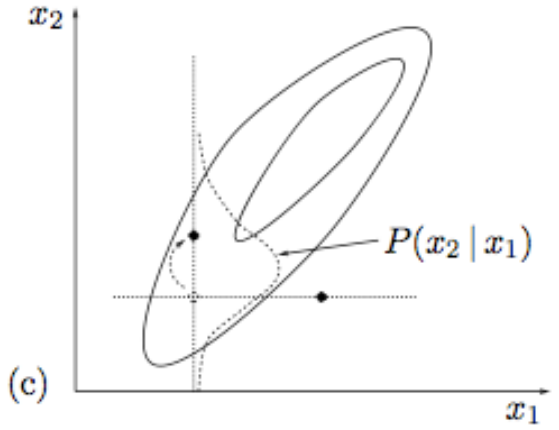
$$N\left(\frac{\mathbf{X'_{ij}y} - \mathbf{X'_{ij}X}\beta_{(\mathbf{ij=0})} - \mathbf{X'_{ij}1_n}\mu}{\mathbf{X'_{ij}X_{ij}} + \sigma_e^2/\sigma_\beta^2}, \sigma_e^2/\left(\mathbf{X_{ij}}'\mathbf{X}_{ij} + \sigma_e^2/\sigma_\beta^2\right)\right)$$



Gibbs sampling







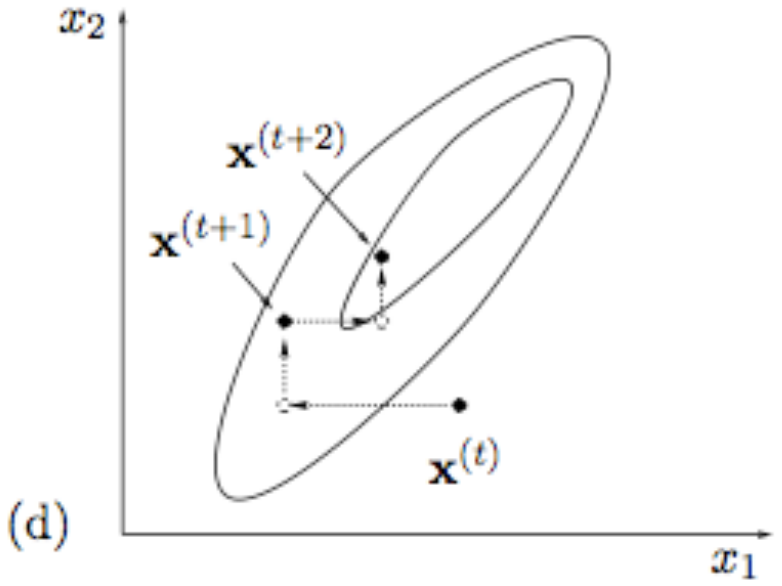


Figure source

Gibbs chain

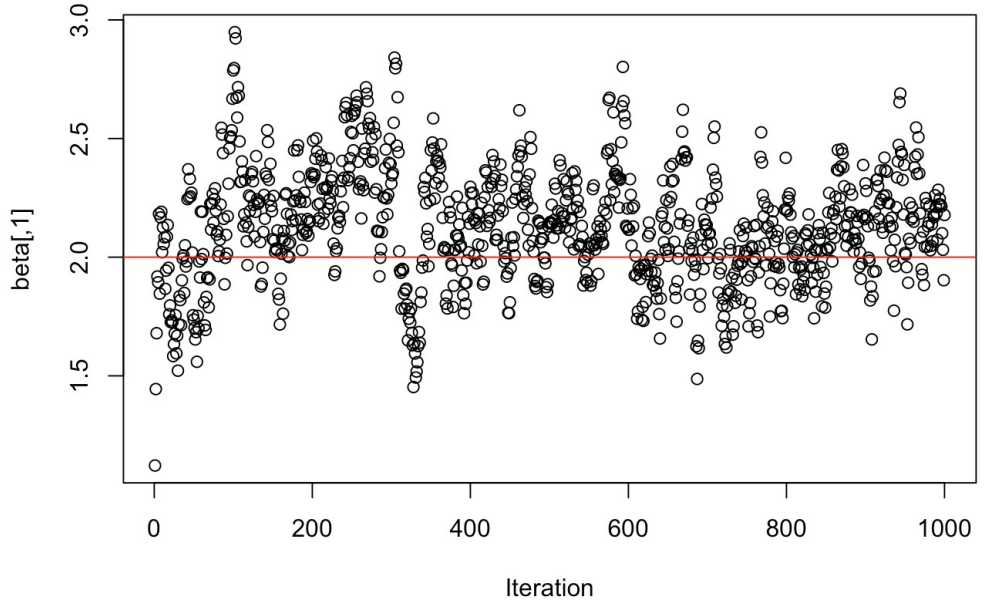
- Set starting values for $(\mu, \boldsymbol{\delta}, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2)$
- Then (for many iterations)
  - For each SNP, sample δ_i, β_i conditional on other parameters
  - Sample $\mu, \sigma_\beta^2, \pi, \sigma_e^2$ with updated $\boldsymbol{\delta}, \boldsymbol{\beta}$

Samples reconstruct posterior distributions of parameters

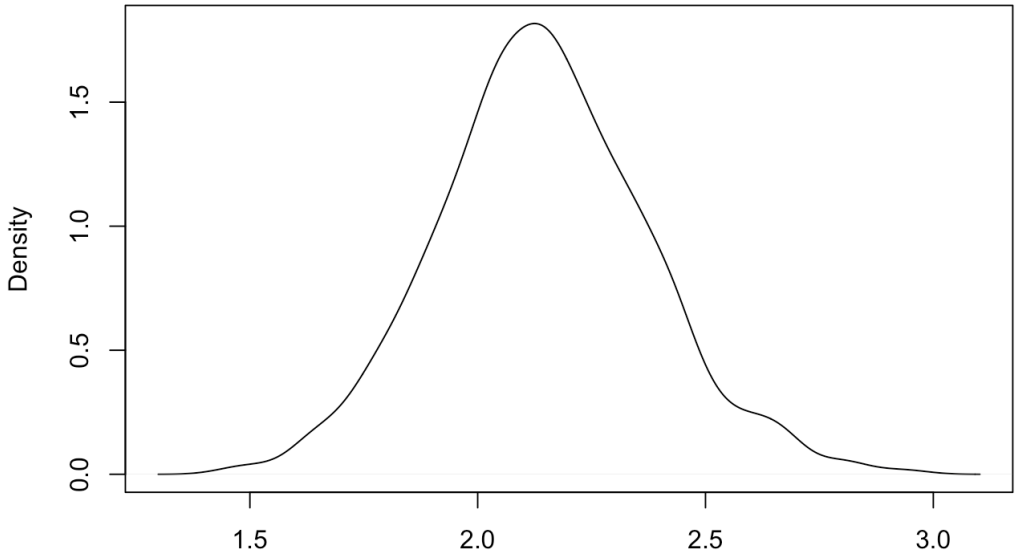
Gibbs sampling is Markov chain Monte Carlo (MCMC) algorithm

Gibbs chain

Trace plot



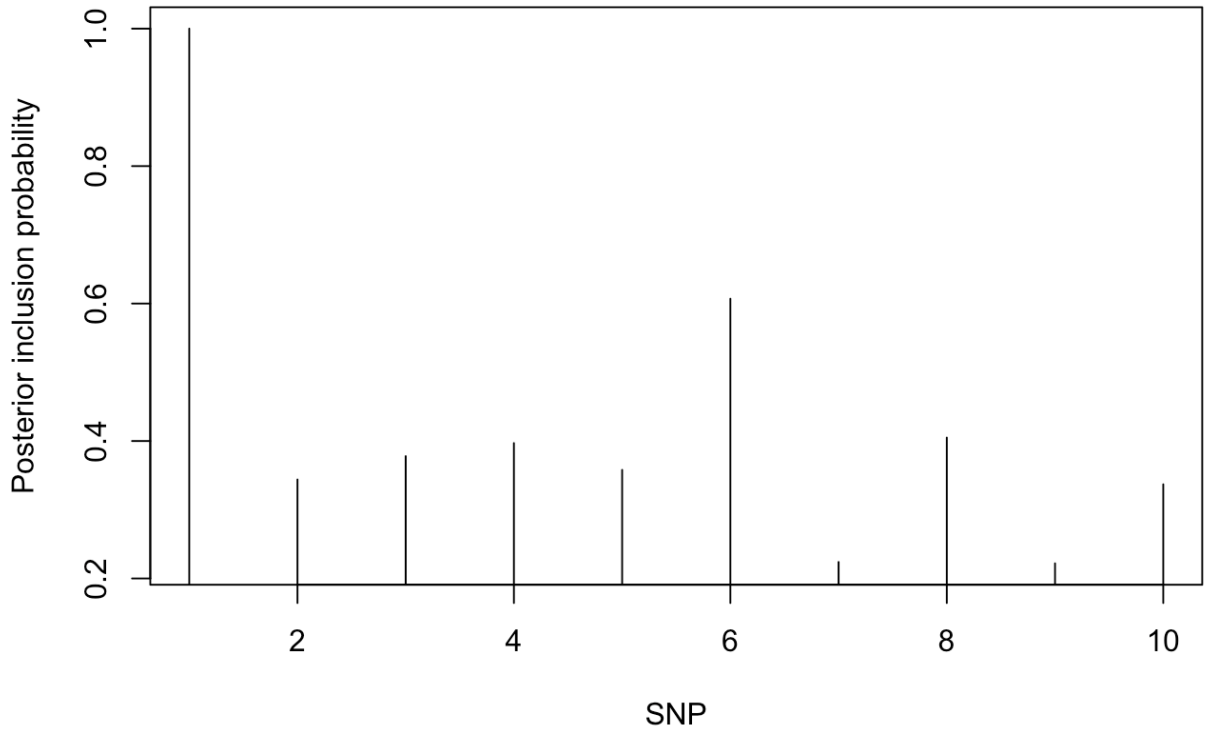
Posterior distribution



Posterior mean is often used as the point estimate of the SNP effect

As a method of fine-mapping

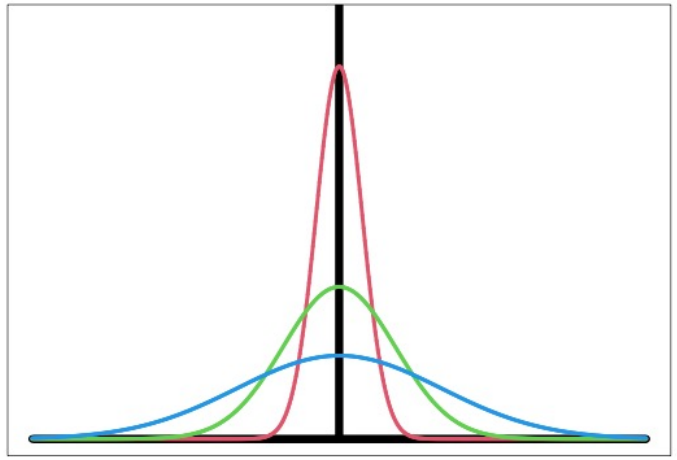
Posterior inclusion probability (PIP):
probability that the SNP is included in the model with a nonzero effect.



Model

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

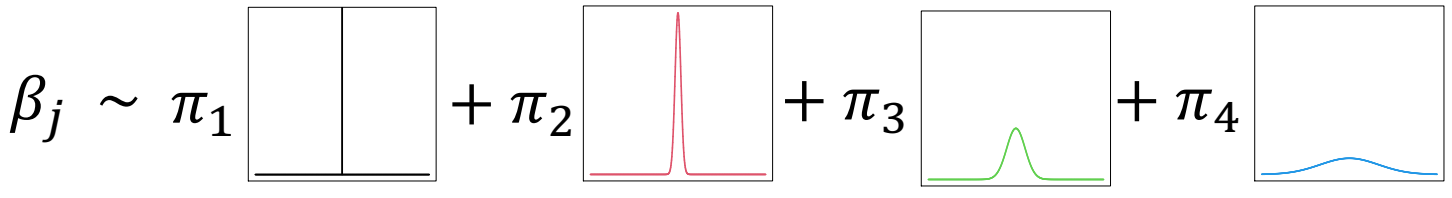
$$\beta_j|\pi, \sigma^2_\beta = \begin{cases} 0 & \text{with probability } \pi_1, \\ \sim N(0, \gamma_2\sigma^2_\beta) & \text{with probability } \pi_2, \\ \vdots & \\ \sim N(0, \gamma_C\sigma^2_\beta) & \text{with probability } 1 - \sum_{c=1}^{C-1} \pi_c, \end{cases}$$

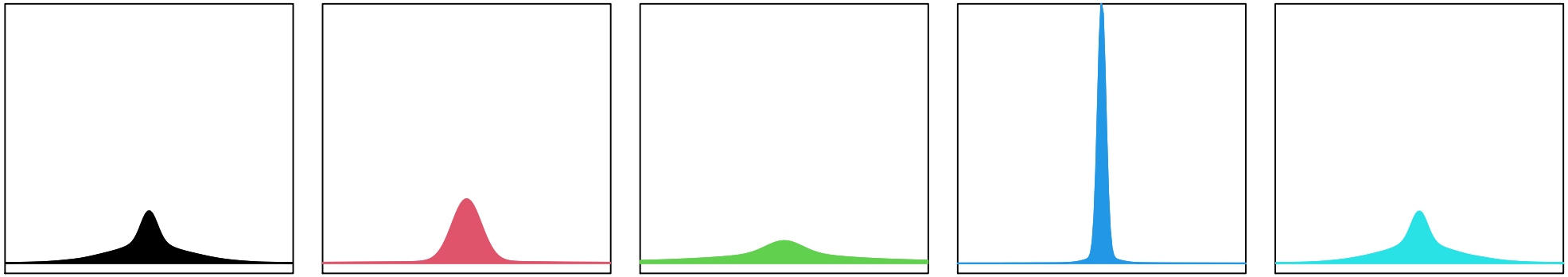


$$\boldsymbol{\gamma} = (0, 0.01, 0.1, 1.0)'$$

BayesC is a special case of BayesR with two components

Why use multi-normal mixture?

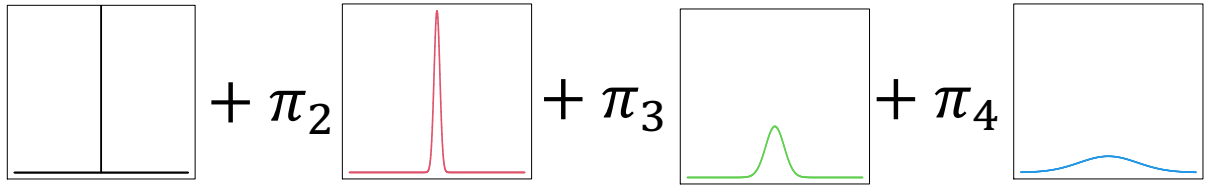




Account for almost any distribution!

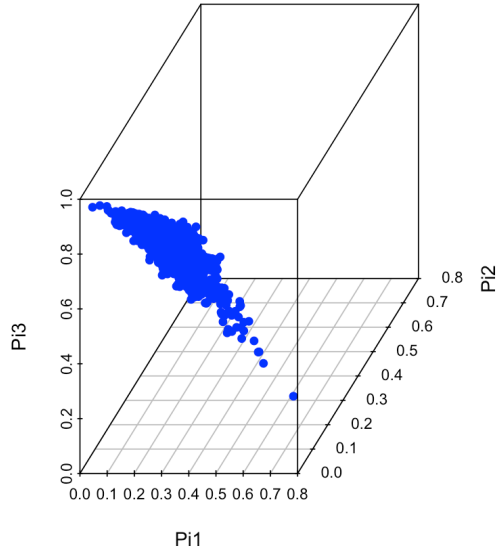
Estimate $\boldsymbol{\pi}$ from the data

$$\beta_j \sim \pi_1 + \pi_2 + \pi_3 + \pi_4$$



Sample $\boldsymbol{\pi}$ from a Dirichlet distribution (multivariate Beta distribution)

$$[\pi_1, \pi_2, \pi_3, \pi_4]' \sim Dirichlet(a_1, a_2, a_3, a_4)$$



Applications of BayesR

Cattle, 800K SNPs

- 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)

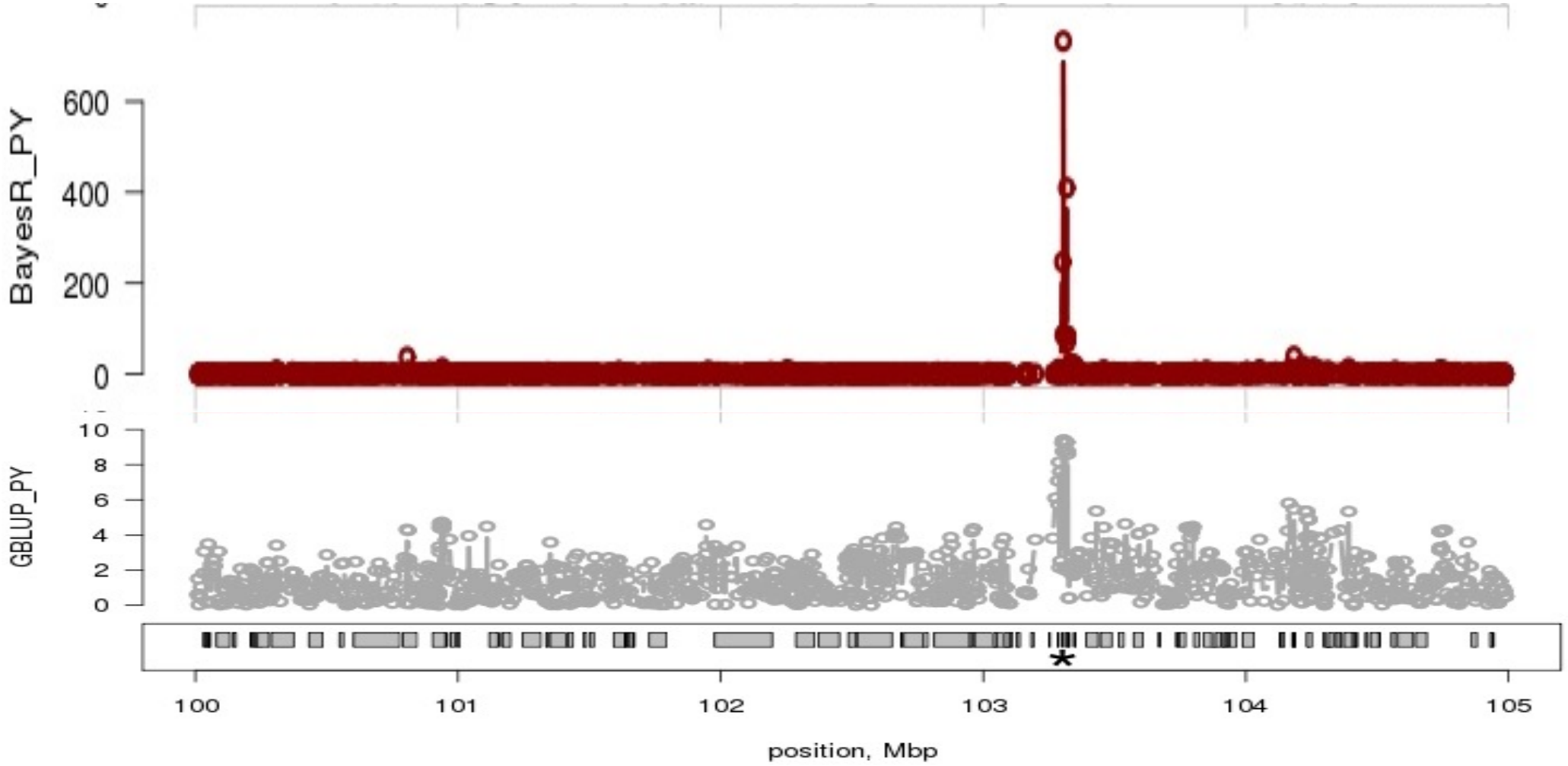
Cattle, 800K SNPs

- Prediction accuracy r($\hat{g}$,y)

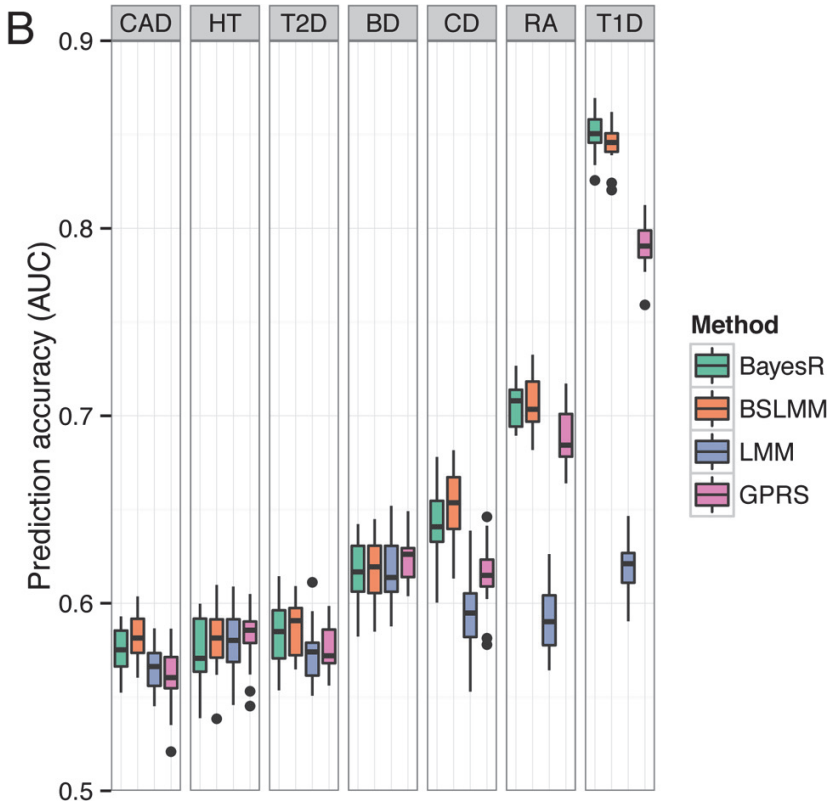
	Fat	Milk	Protein	Fat%	Protein%	**Average**
Holstein						
GBLUP	0.60	0.59	0.58	0.72	0.83	**0.66**
BAYESR	0.64	0.62	0.57	0.81	0.84	**0.69**
Jersey						
GBLUP	0.56	0.62	0.67	0.64	0.76	**0.65**
BAYESR	0.56	0.69	0.71	0.76	0.79	**0.70**
Australian Reds						
GBLUP	0.20	0.16	0.11	0.32	0.34	**0.22**
BAYES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

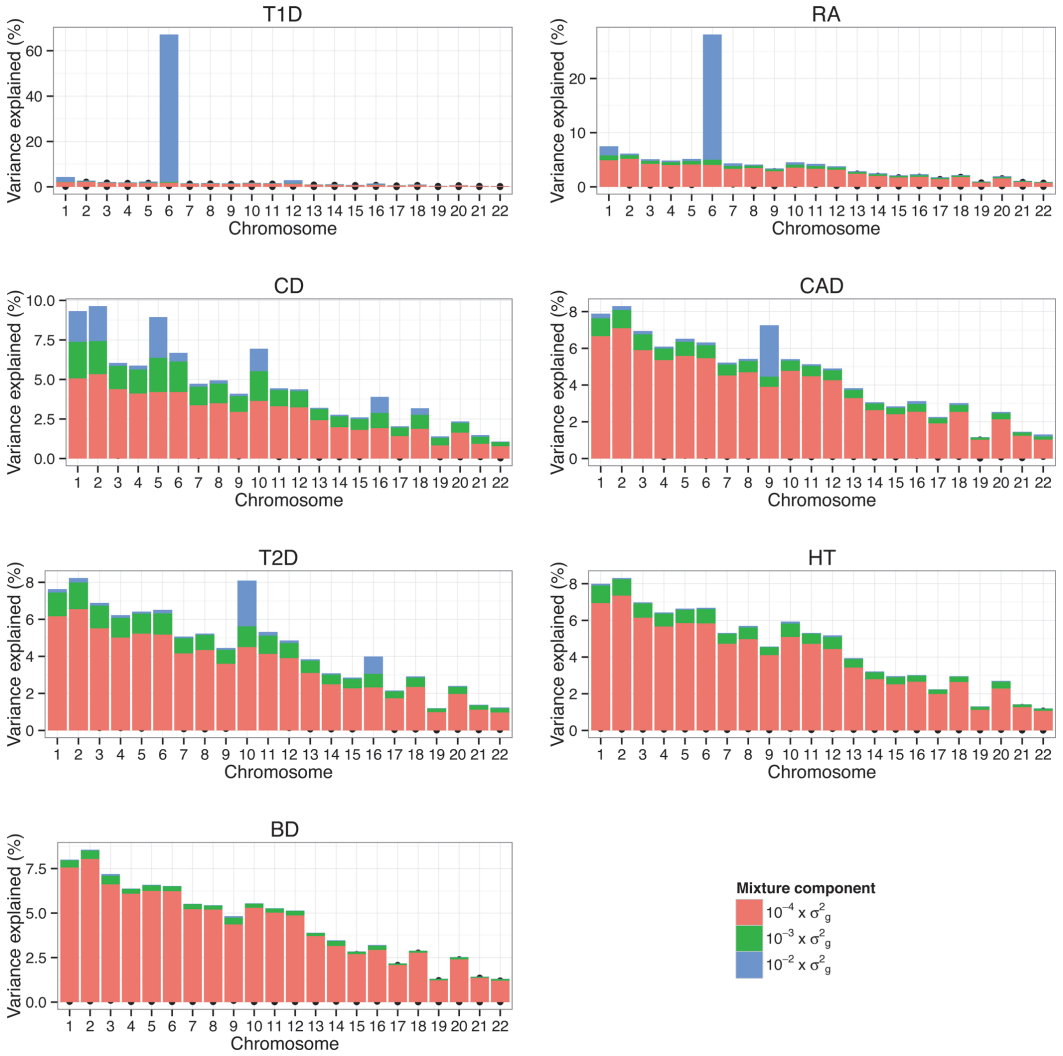
BayesR application

BayesR



Prediction of disease risk in humans





Moser et al PLoS Genetics 2015

Bayesian methods for Genomic Prediction

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

Bayesian methods can have an advantage when:

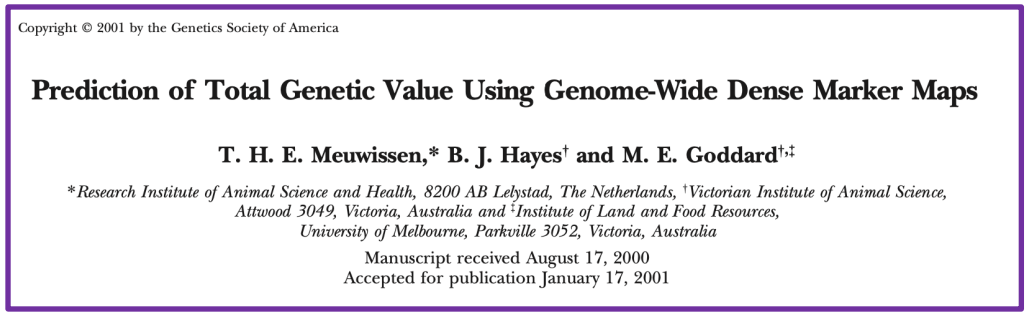
QTL of moderate to large effect on the trait (eg Fat%, DGAT1)

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

Integrates polygenic prediction and genetic fine-mapping

Reference

BayesA, BayesB:



Copyright © 2001 by the Genetics Society of America

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*

Manuscript received August 17, 2000
Accepted for publication January 17, 2001

Meuwissen et al is the paper coined genomic selection.

BayesC:

Habier *et al. BMC Bioinformatics* 2011, **12**:186
http://www.biomedcentral.com/1471-2105/12/186

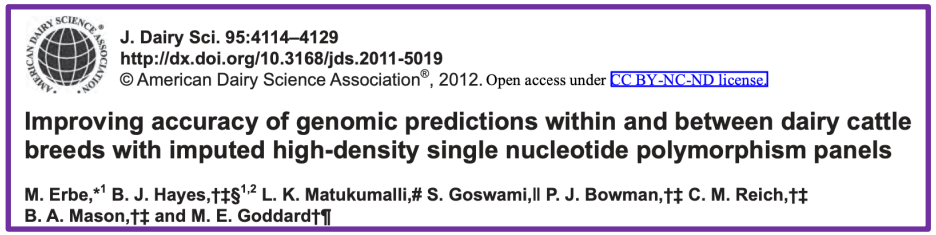
BMC Bioinformatics

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]

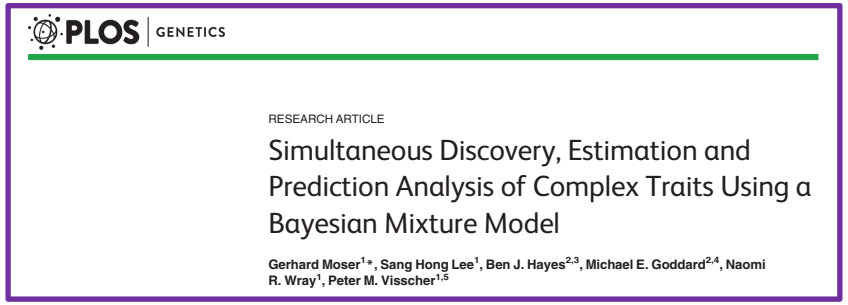
BayesR:



J. Dairy Sci. 95:4114–4129
http://dx.doi.org/10.3168/jds.2011-5019
© American Dairy Science Association®, 2012. Open access under CC BY-NC-ND license.

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: Bayesian methods

https://cnsgenomics.com/data/teaching/GNGWS24/module5/Practical4_Bayes.html

To log into your server, type command below in **Terminal** for Mac/Linux users or in **Command Prompt** or **PowerShell** for Windows users.

```
ssh username@hostname
```

And then key in the provided password.