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

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

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

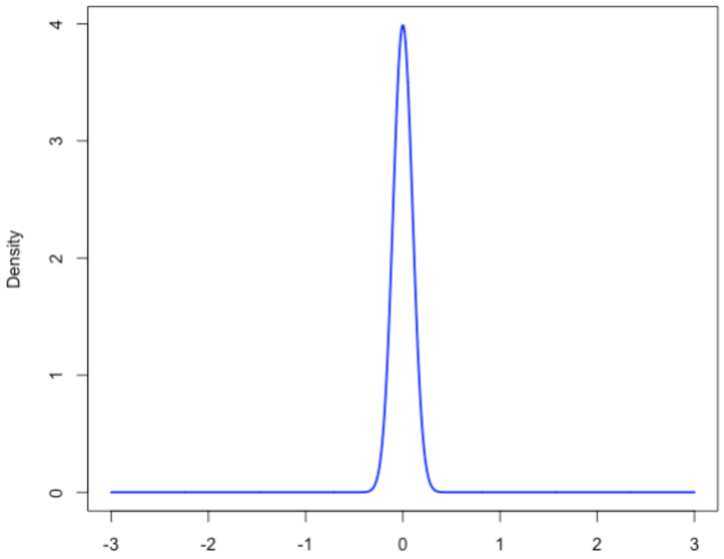
$\lambda = \sigma_e^2 \,/\, \sigma_\beta^2$

BLUP

- 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

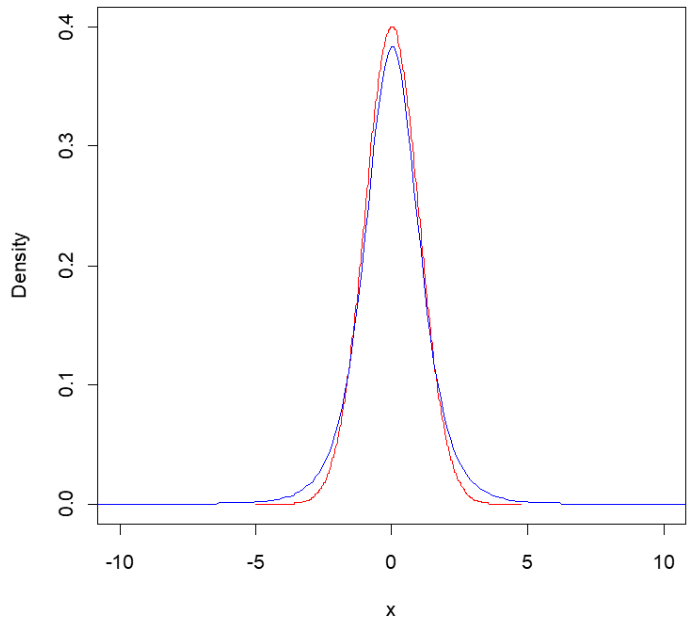
- Allow alternative assumptions regarding the distribution of SNP effects

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

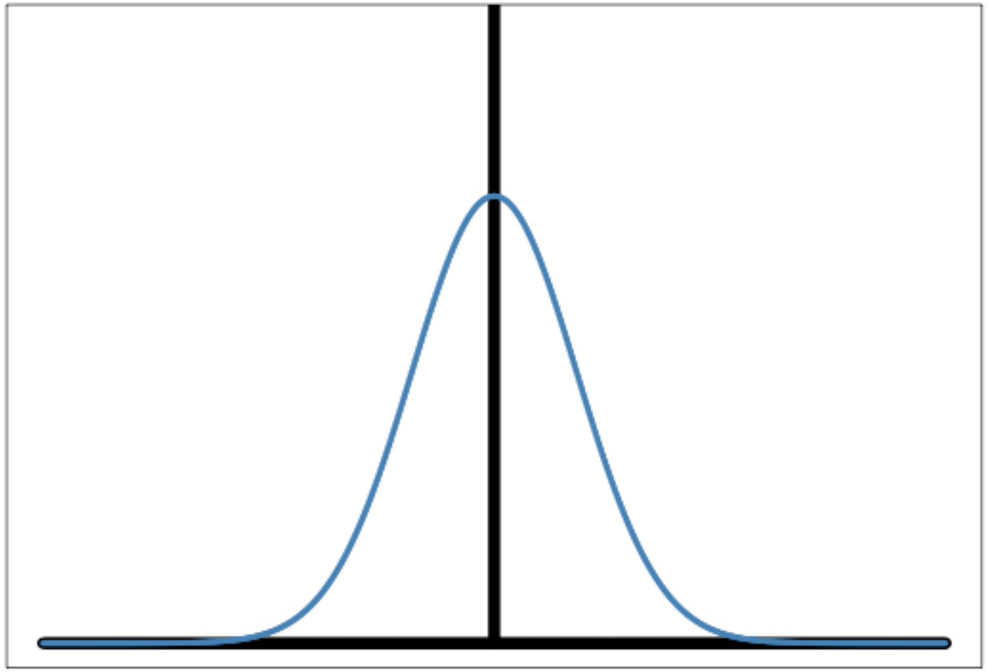
The University of Queensland
Australia

Alternative distributions

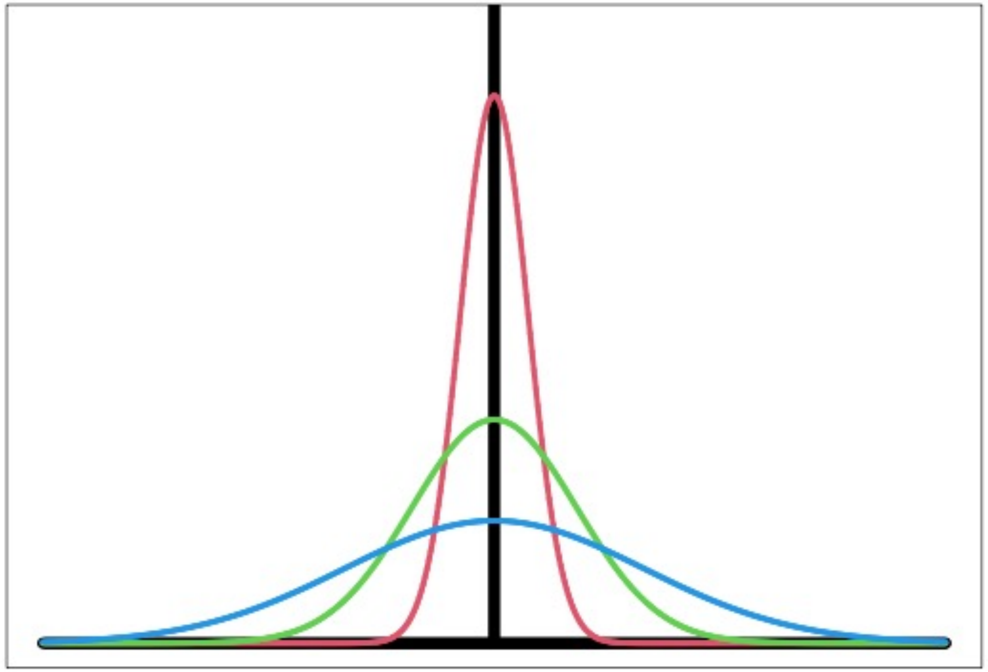
Assumption	Distribution of SNP effects	Method
Small number of moderate to large effects, many small effects	Students t	BayesA
Small number of moderate to large effects, many zero effects	Mixture, spike at zero, Students t	BayesB
Small number of small effects, many zero effects	Mixture, spike at zero, 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

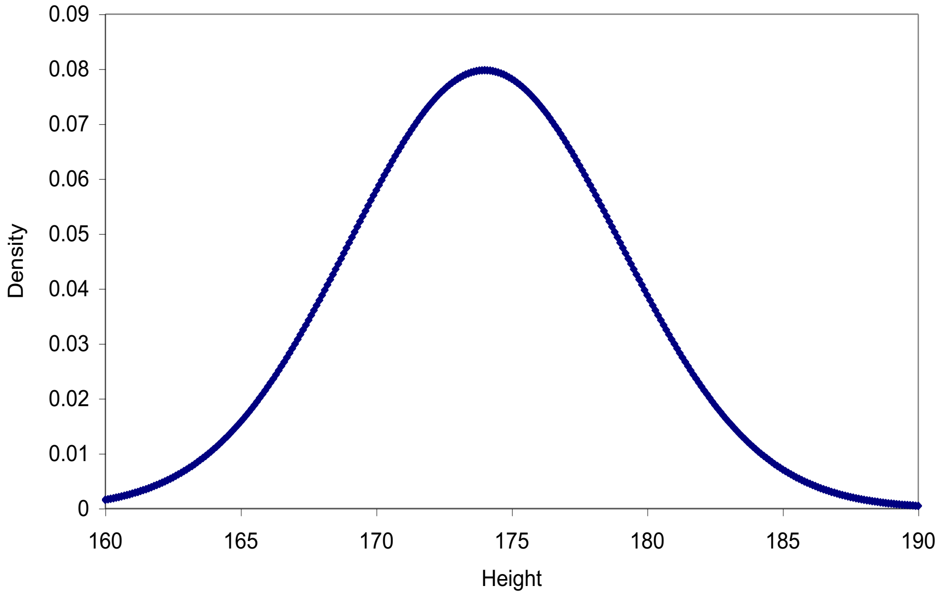
$$y = \text{average height} + e$$

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Bayes theorem

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

Prior probability of x (average height)



Bayes theorem

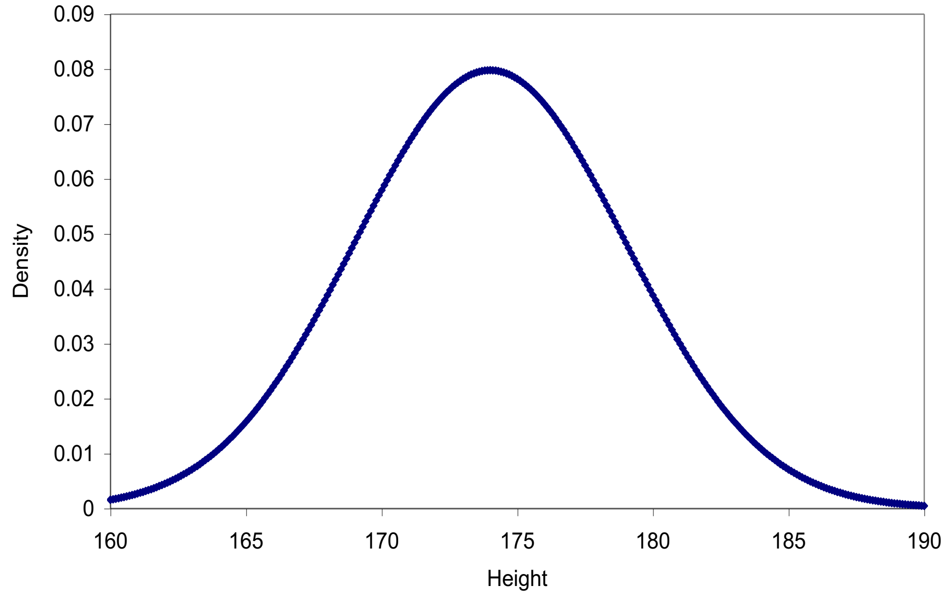
$$P(x \mid y) \propto P(y \mid x) P(x)$$

From the data……

$$\bar{x} = 178$$

$$s.e = 5$$

Prior probability of x (average height)

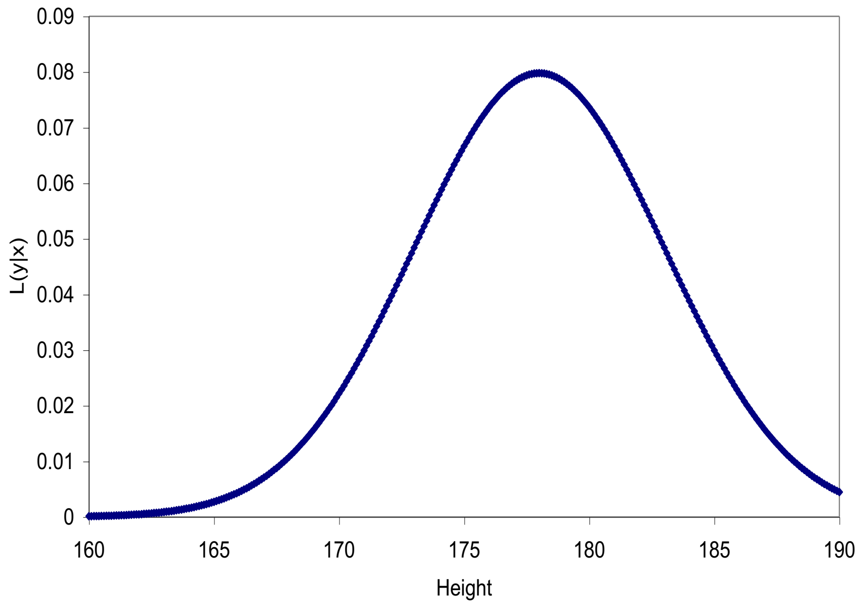


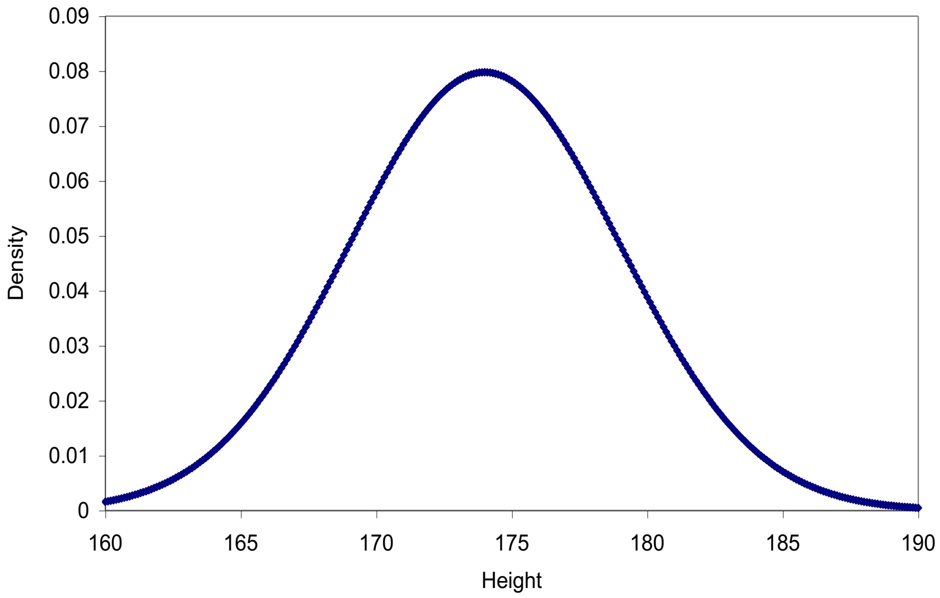
Bayes theorem

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

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

Prior probability of x (average height)

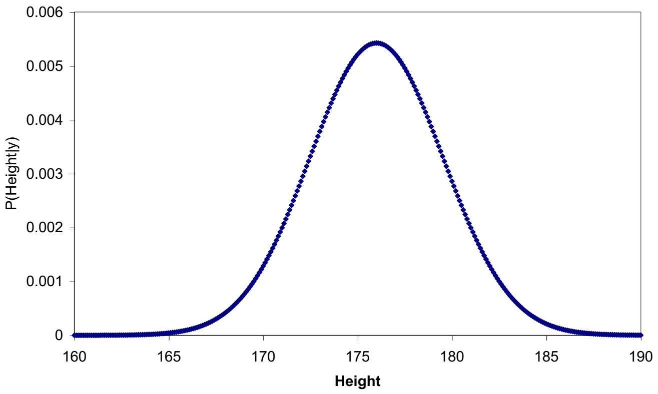




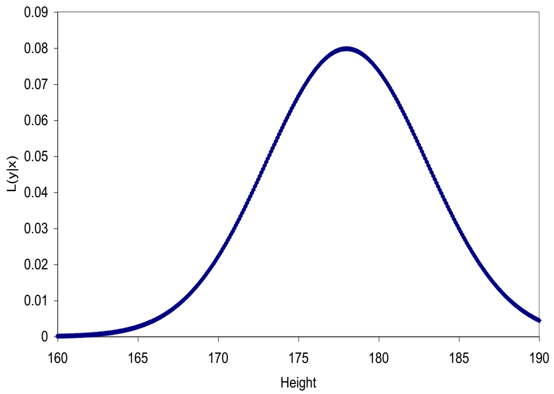
Bayes theorem

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

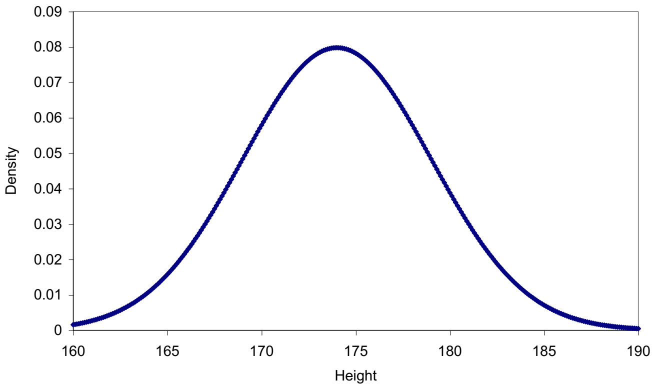
P(x|y) mean = 176cm



L(y|x)



P(x)



Introduction to Bayesian methods

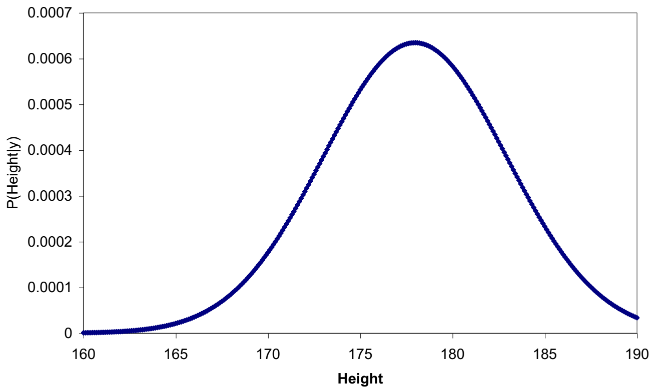
Less certainty about prior information? Use *less* informative (flat) prior

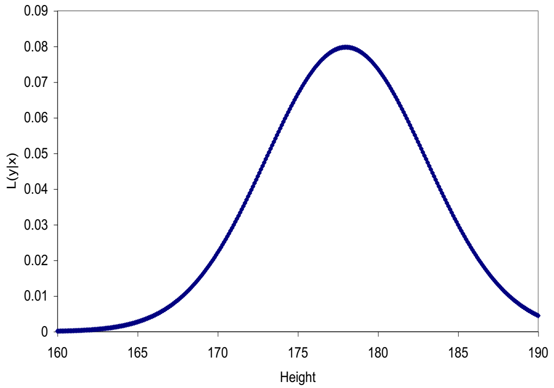
$$P(x \mid y) \propto P(y \mid x) P(x)$$

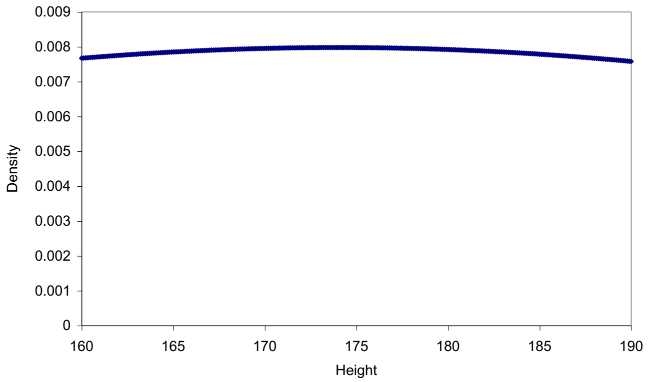
P(x|y) mean = 178cm

L(y|x)

P(x)



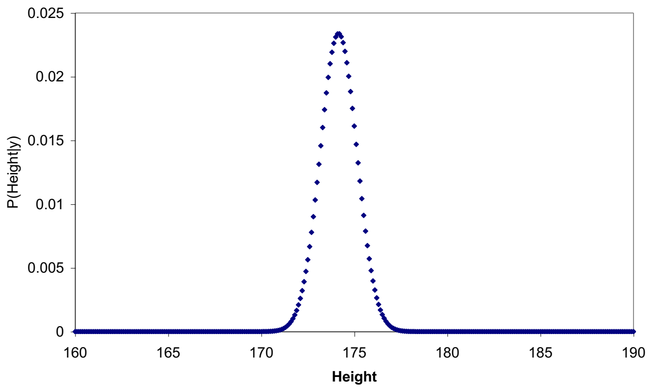




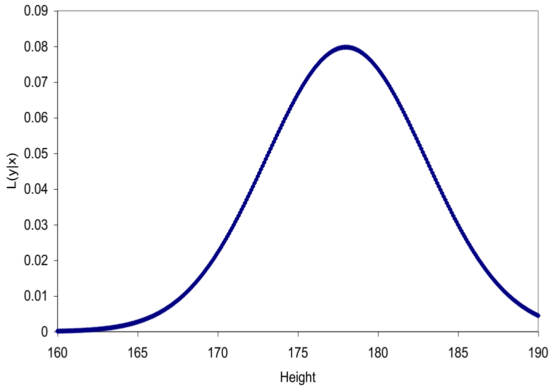
More certainty about prior information? Use *more* informative prior

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

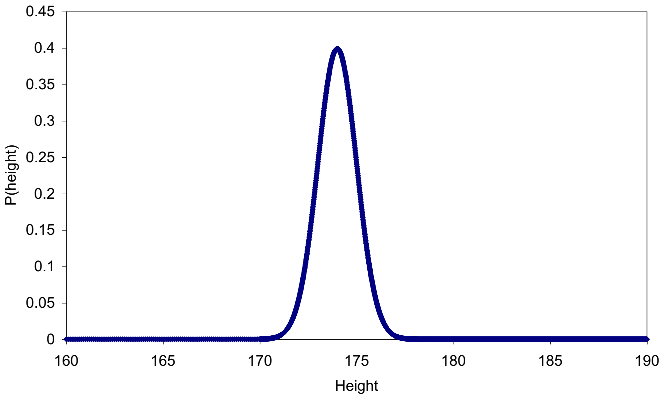
P(x|y) mean = 174.5cm



L(y|x)



P(x)

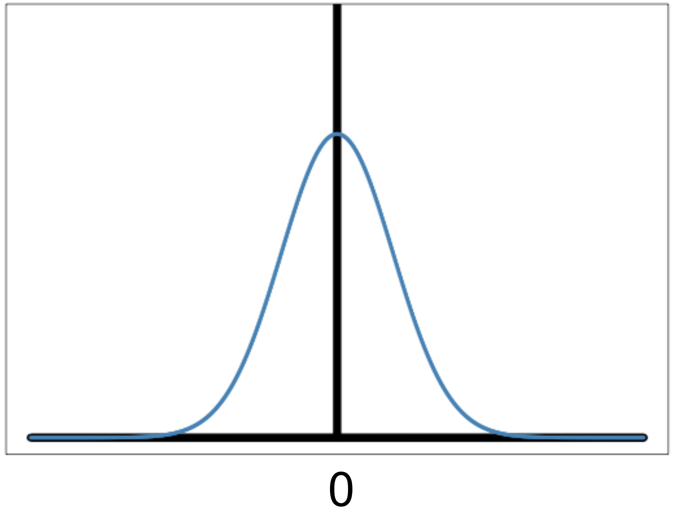


PGS prediction with Bayesian methods

Model

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

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



BLUP is a special case of BayesC when $\pi = 1$

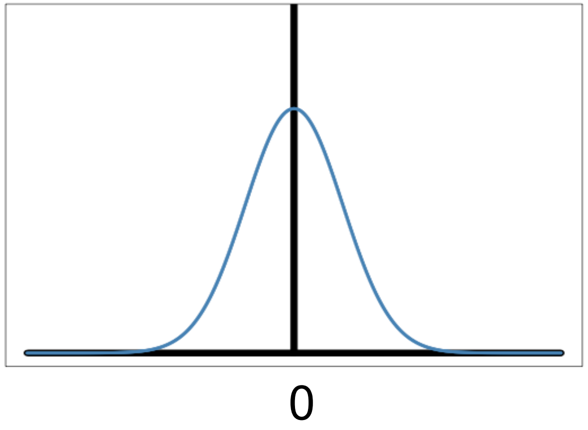
Posterior inference on SNP effects

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

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

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma_e^2}\right\} \prod_{j=1}^{m}\left[(\sigma_\beta^2)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi + \varphi_0(1-\pi)\right]$$



SNP effect estimates:

$$\widehat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\boldsymbol{y}) = \int \boldsymbol{\beta} P(\boldsymbol{\beta}|\boldsymbol{y})\mathrm{d}\boldsymbol{\beta}$$

$$= \int_{\beta_1} \dots \int_{\beta_m} (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma_e^2}\right\} \prod_{j=1}^{m}\left[(\sigma_\beta^2)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi + \varphi_0(1-\pi)\right] \mathrm{d}\beta_1 \dots \mathrm{d}\beta_m$$

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Posterior inference on SNP effects

$$\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\boldsymbol{y}) = \int_{\beta_1} \dots \int_{\beta_m} (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma_e^2}\right\} \prod_{j=1}^{m} \left[(\sigma_\beta^2)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi + \varphi_0(1-\pi)\right] \mathrm{d}\beta_1 \dots \mathrm{d}\beta_m$$

- Cannot solve directly → no closed form solution
- Estimates of parameters depend on other parameters
- Use Markov chain Monte Carlo (MCMC) algorithm!

MCMC algorithm

Markov chain

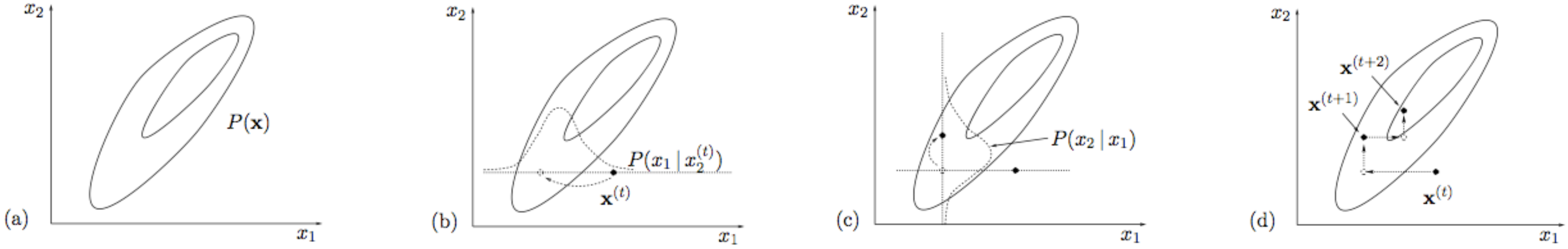
A sequence of samples where each sample depends only on the previous one (memoryless). This property allows the algorithm to gradually explore the distribution.

Monte Carlo

Using random sampling to perform numerical estimation, e.g., integrating over a probability distribution by averaging over samples.

Gibbs Sampling

A special case of MCMC to sample from posterior distribution of each parameter **conditional** on all other parameters.



The key is to derive $P(x_1|x_2)$ and $P(x_2|x_1)$

Figure source

BayesC

To run Gibbs sampling, we need to derive the full conditional distribution for each parameter

- $P(\mu | \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2)$
- $P(\beta_j | \boldsymbol{y}, \boldsymbol{\beta}_{-j}, \sigma_\beta^2, \pi, \sigma_e^2)$
- $P(\sigma_\beta^2 | \boldsymbol{y}, \boldsymbol{\beta}, \pi, \sigma_e^2)$
- $P(\pi | \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\beta^2, \sigma_e^2)$
- $P(\sigma_e^2 | \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\beta^2, \pi)$

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

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

Posterior joint distribution

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

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

Scaled inverse chi-square distribution

Likelihood

Flat prior

Point-normal mixture

Beta distribution

Posterior joint distribution

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

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$ ← Likelihood

$$\times \prod_{j=1}^{m}\left[(\sigma_\beta^2)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi + \varphi_0(1-\pi)\right]$$ ← Prior for $\boldsymbol{\beta}$: point-normal mixture

$$\times (\sigma_\beta^2)^{-\frac{v_\beta+2}{2}} exp\left\{-\frac{v_\beta\tau_\beta^2}{2\sigma_\beta^2}\right\}$$ ← Prior for σ_β^2 : scaled inverse chi-square distribution

$$\times (\sigma_e^2)^{-\frac{v_e+2}{2}} exp\left\{-\frac{v_e\tau_e^2}{2\sigma_e^2}\right\}$$ ← Prior for σ_e^2 : scaled inverse chi-square distribution

$$\times \pi^{a-1}(1-\pi)^{b-1}$$ ← Prior for π : beta distribution

Full conditional distribution for μ

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

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}}\exp\left\{-\frac{\left(\mathbf{y}-\mathbf{1}\mu-\sum_j\mathbf{X}_j\beta_j\right)'\left(\mathbf{y}-\mathbf{1}\mu-\sum_j\mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times\prod_{j=1}^{m}\left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi+\varphi_0(1-\pi)\right]$$

$$\times\left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta+2}{2}}exp\left\{-\frac{\upsilon_\beta\tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times\left(\sigma_e^2\right)^{-\frac{\upsilon_e+2}{2}}exp\left\{-\frac{\upsilon_e\tau_e^2}{2\sigma_e^2}\right\}$$

$$\times\pi^{a-1}(1-\pi)^{b-1}$$

Full conditional distribution for μ

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\sim N\left(\frac{\mathbf{1}'\left(\mathbf{y} - \sum_j \mathbf{X}_j\beta_j\right)}{n}, \frac{\sigma_e^2}{n}\right)$$

Full conditional distribution for β_j

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

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times \prod_{j=1}^{m} \left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\} \pi + \varphi_0 (1 - \pi)\right]$$

$$\times \left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta + 2}{2}} exp\left\{-\frac{\upsilon_\beta \tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times \left(\sigma_e^2\right)^{-\frac{\upsilon_e + 2}{2}} exp\left\{-\frac{\upsilon_e \tau_e^2}{2\sigma_e^2}\right\}$$

$$\times \pi^{a-1} (1 - \pi)^{b-1}$$

Full conditional distribution for β_j

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y}-\mathbf{1}\mu-\sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y}-\mathbf{1}\mu-\sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi + \varphi_0(1-\pi)$$

Let’s introduce an indicator variable δ_j

If $\delta_j = 1,$ then β_j is in non-zero component

If $\delta_j = 0,$ then $\beta_j = 0$

Full conditional distribution for β_j

If $\delta_j = 1$

$$P\left(\beta_j \middle| \mathbf{y}, \delta_j = 1, \boldsymbol{\beta}_{-j}, \sigma_\beta^2, \pi, \sigma_e^2\right)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)}{2\sigma_e^2}\right\} \times \left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}$$

$$\sim N\left(\frac{\mathbf{X}_j'(\mathbf{y} - \mathbf{1}\mu - \sum_{k \neq j} \mathbf{X}_k' \beta_k)}{\mathbf{X}_j'\mathbf{X}_j + \sigma_e^2/\sigma_\beta^2}, \frac{\sigma_e^2}{\mathbf{X}_j'\mathbf{X}_j + \sigma_e^2/\sigma_\beta^2}\right)$$

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Full conditional distribution for σ_β^2

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

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}}\exp\left\{-\frac{\left(\mathbf{y}-\mathbf{1}\mu-\sum_j\mathbf{X}_j\beta_j\right)'\left(\mathbf{y}-\mathbf{1}\mu-\sum_j\mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times\prod_{j=1}^{m}\left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi+\varphi_0(1-\pi)\right]$$

$$\times\left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta+2}{2}}exp\left\{-\frac{\upsilon_\beta\tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times(\sigma_e^2)^{-\frac{\upsilon_e+2}{2}}exp\left\{-\frac{\upsilon_e\tau_e^2}{2\sigma_e^2}\right\}$$

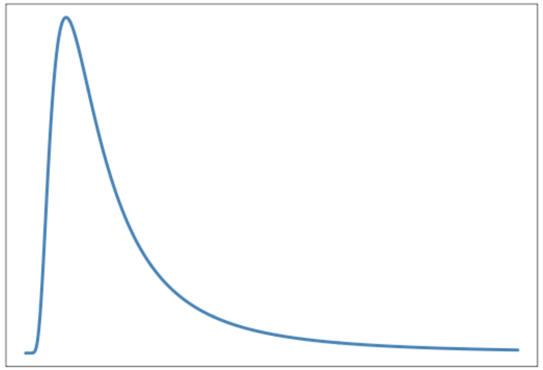
$$\times\pi^{a-1}(1-\pi)^{b-1}$$

Full conditional distribution for σ_β^2

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

$$\propto \prod_{j=1}^{m} \left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\right]^{\delta_j} \times \left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta+2}{2}} exp\left\{-\frac{\upsilon_\beta \tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\sim \chi^{-2}\left(\tilde{\upsilon}_\beta = \upsilon_\beta + \sum_j \delta_j \,, \tilde{\tau}_\beta^2 = \frac{\sum_j \beta_j^2 + \upsilon_\beta \tau_\beta^2}{\tilde{\upsilon}_\beta}\right)$$



Full conditional distribution for π

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

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times \prod_{j=1}^{m} \left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\} \pi + \varphi_0(1-\pi)\right]$$

$$\times \left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta+2}{2}} exp\left\{-\frac{\upsilon_\beta \tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times \left(\sigma_e^2\right)^{-\frac{\upsilon_e+2}{2}} exp\left\{-\frac{\upsilon_e \tau_e^2}{2\sigma_e^2}\right\}$$

$$\times \pi^{a-1}(1-\pi)^{b-1}$$

Full conditional distribution for π

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

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

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times \prod_{j=1}^{m}\left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi + \varphi_0(1-\pi)\right]$$

$$\times\left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta+2}{2}} exp\left\{-\frac{\upsilon_\beta\tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times(\sigma_e^2)^{-\frac{\upsilon_e+2}{2}} exp\left\{-\frac{\upsilon_e\tau_e^2}{2\sigma_e^2}\right\}$$

$$\times \pi^{a-1}(1-\pi)^{b-1}$$

→ Only depends on the indicator variable δ_j

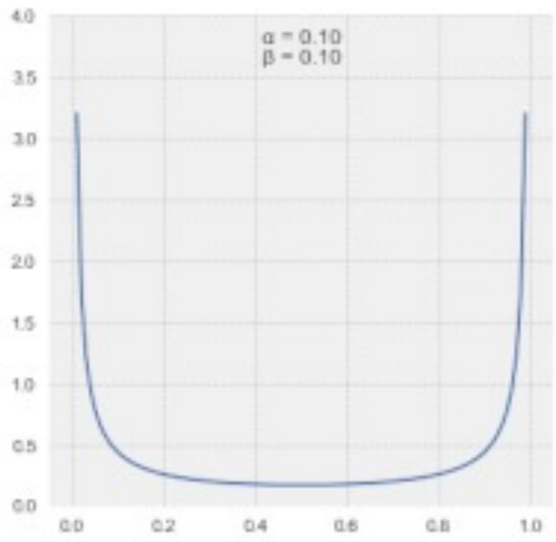
$$\prod_{j=1}^{m}\left[\pi^{\delta_j} + (1-\pi)^{(1-\delta_j)}\right]$$

Full conditional distribution for π

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

$$\propto \prod_{j=1}^{m} \left[\pi^{\delta_j} + (1-\pi)^{(1-\delta_j)}\right] \times \pi^{a-1}(1-\pi)^{b-1}$$

$$\sim Beta\left(a + \sum\nolimits_j \delta_j \, , b + m - \sum\nolimits_j \delta_j\right)$$



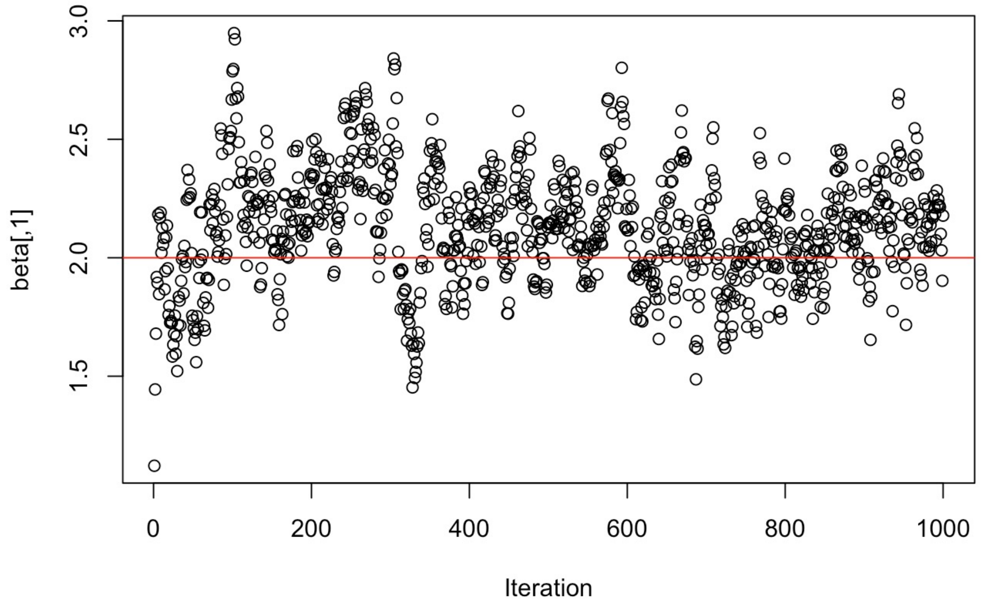
Gibbs sampling

- 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_j, β_j conditional on other parameters
  - Sample $\mu, \sigma_\beta^2, \pi, \sigma_e^2$ with updated $\boldsymbol{\delta}, \boldsymbol{\beta}$

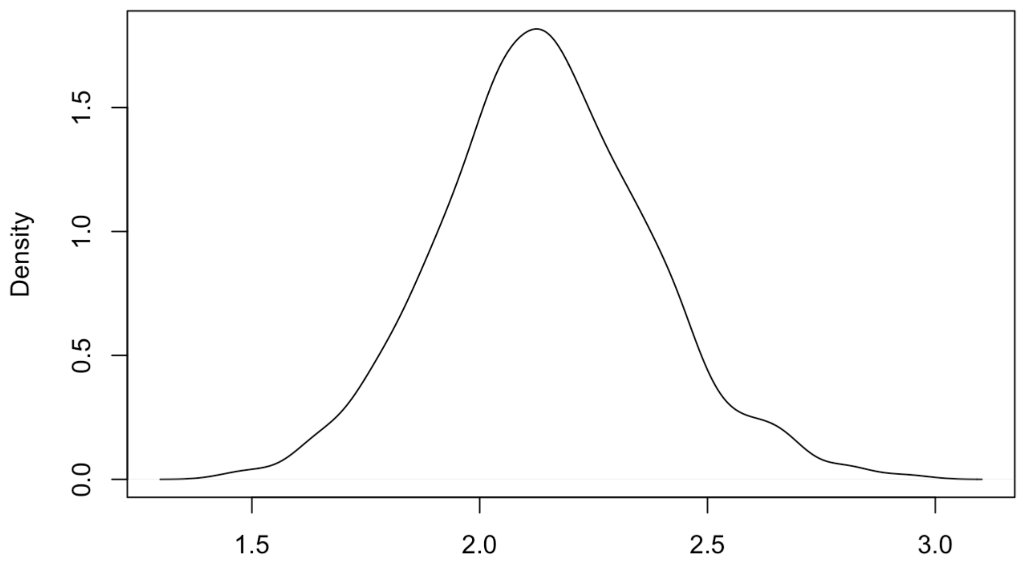
Samples reconstruct posterior distributions of parameters

Gibbs sampling

Trace plot



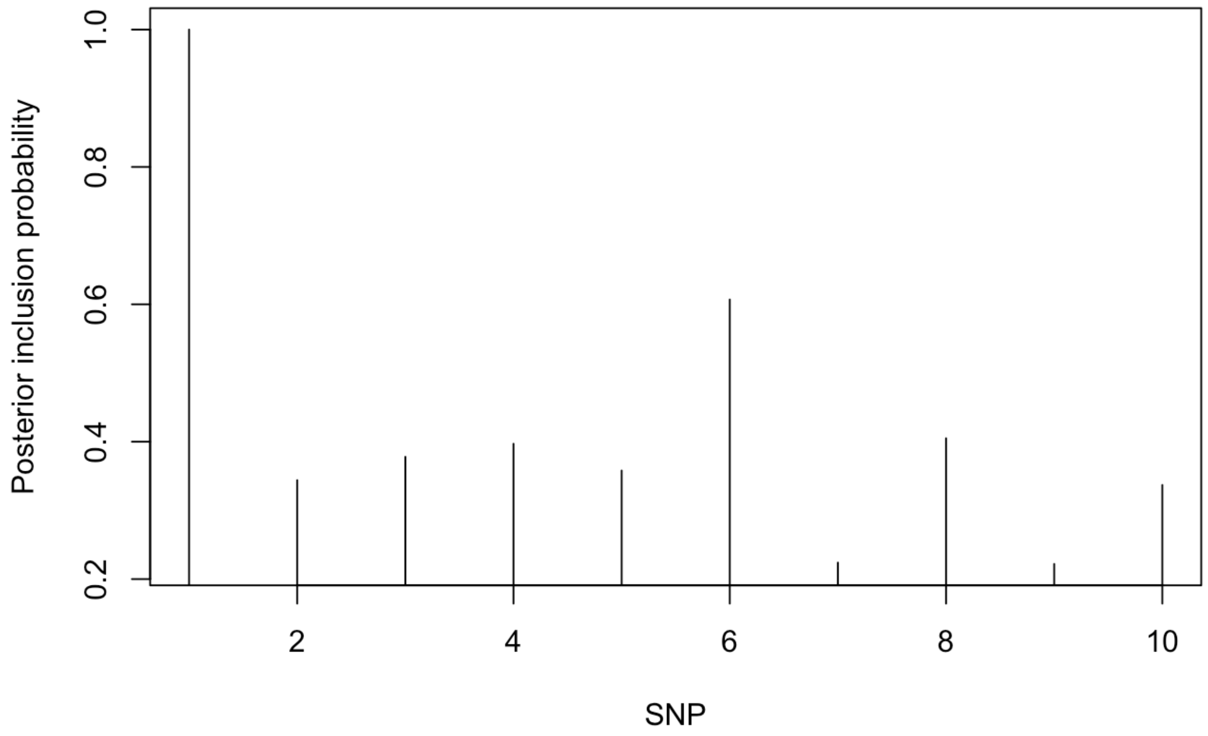
Posterior distribution



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

As a method of fine-mapping

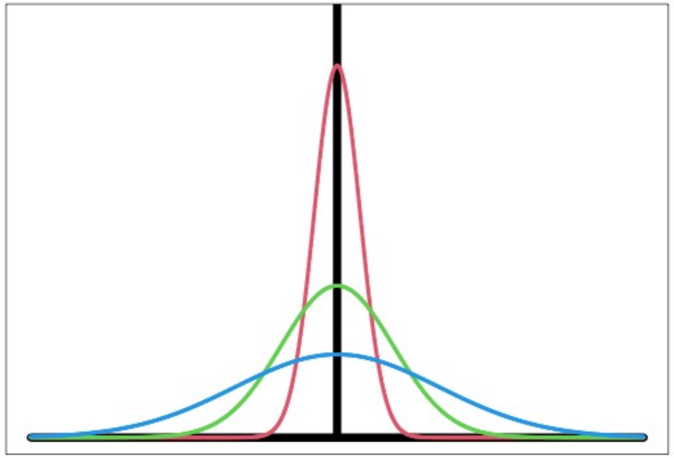
Posterior inclusion probability (PIP):
probability that the SNP is included in the model with a nonzero effect.



Model

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

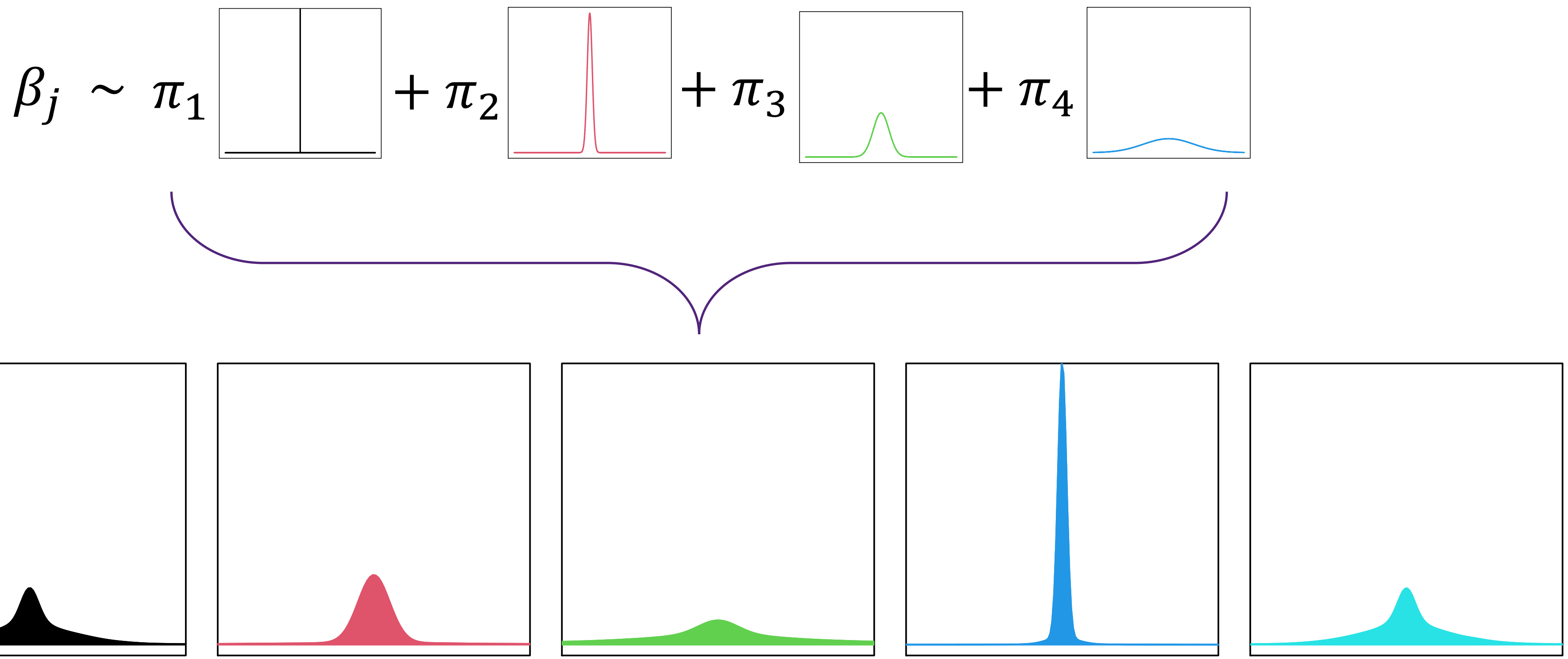
$$\beta_j|\pi, \sigma_\beta^2 = \begin{cases} 0 & \text{with probability } \pi_1, \\ \sim N(0, \gamma_2\sigma_\beta^2) & \text{with probability } \pi_2, \\ \vdots & \\ \sim N(0, \gamma_C\sigma_\beta^2) & \text{with probability } 1 - \sum_{c=1}^{C-1} \pi_c, \end{cases}$$



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

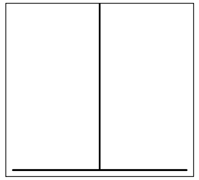
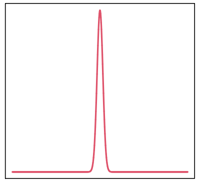
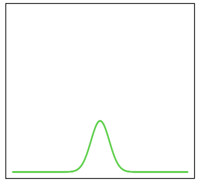
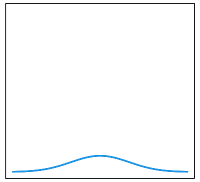
BayesC is a special case of BayesR with two components

Why use multi-normal mixture?



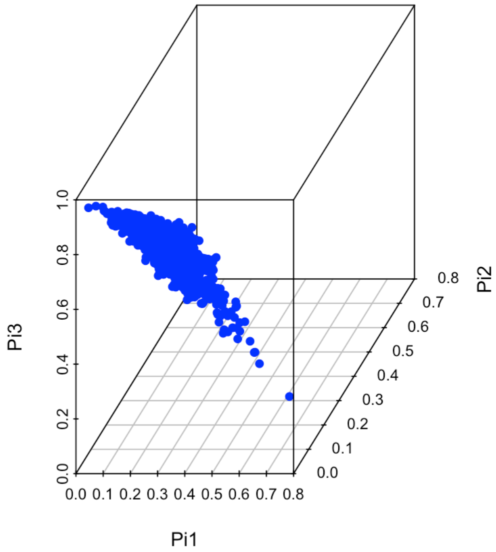
Account for almost any distribution!

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)

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



More details in Lloyd-Jones et al Nat Comm 2019

Applications of BayesR

Cattle, 800K SNPs

- Training
  - 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)

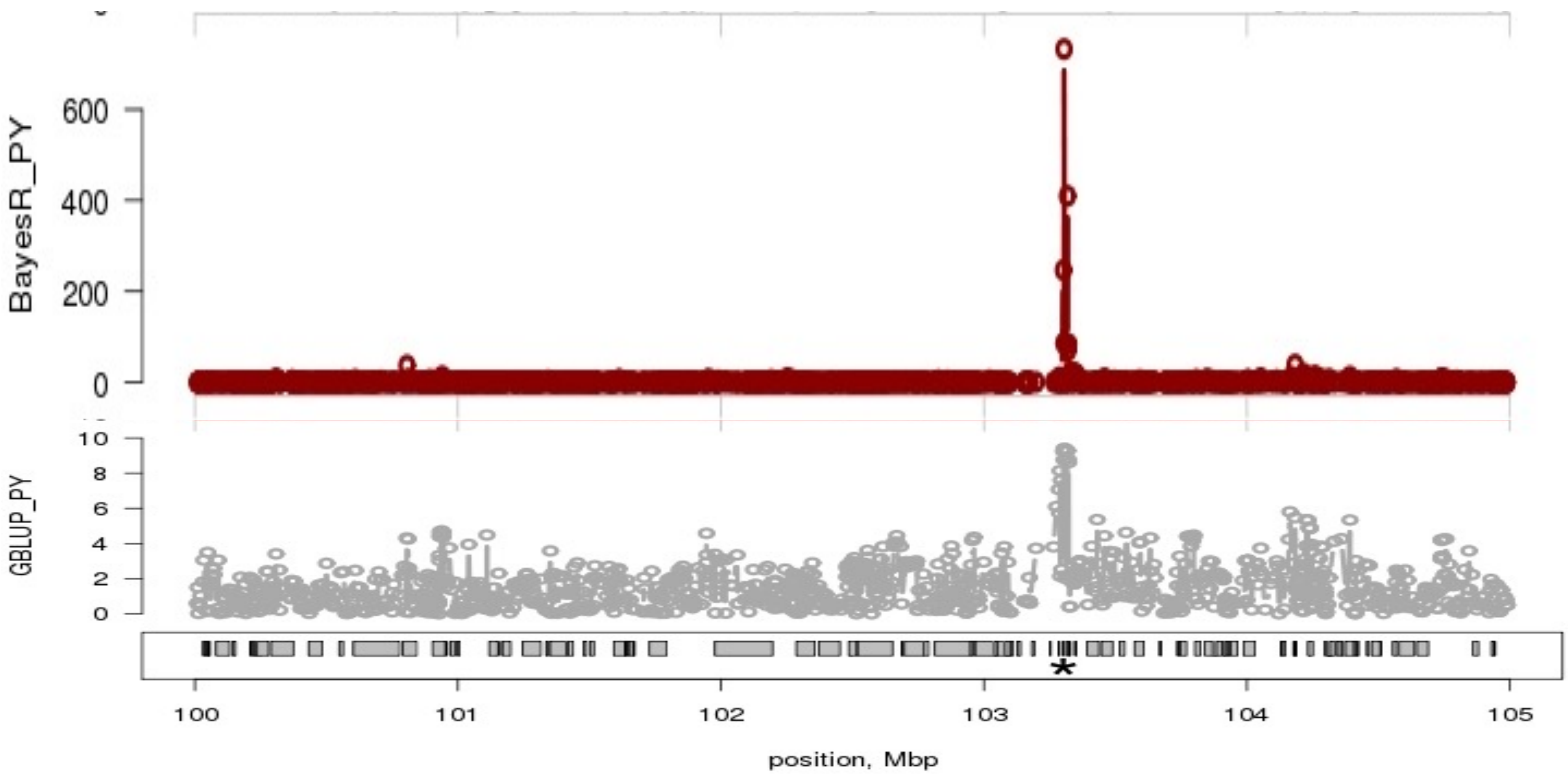
The University of Queensland Australia

Cattle, 800K SNPs

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

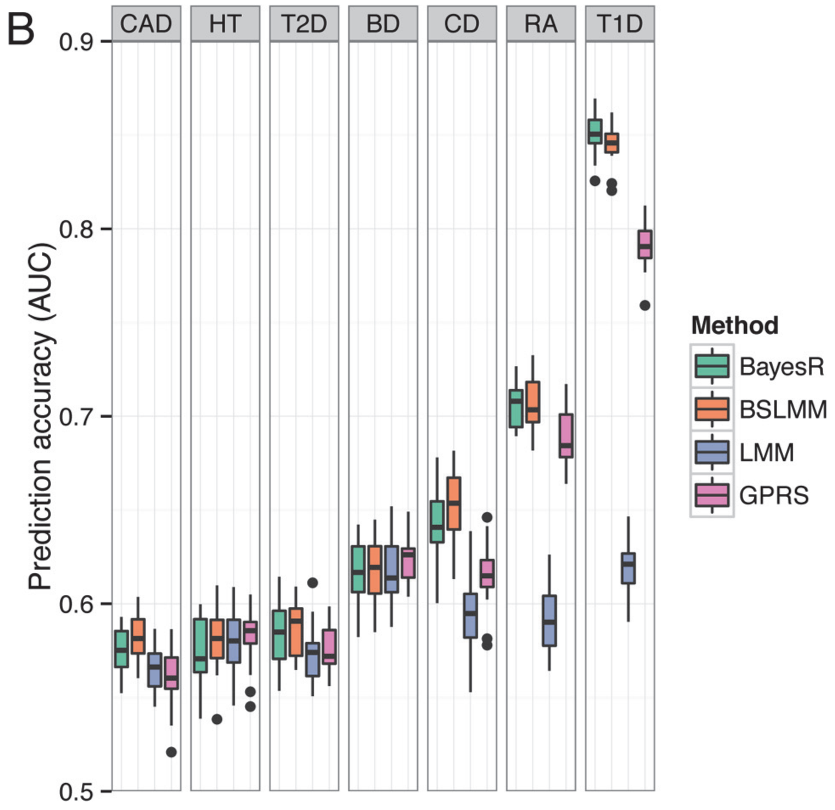
	Fat	Milk	Protein	Fat%	Protein%	**Average**
Holstein						
GBLUP	0.60	0.59	0.58	0.72	0.83	**0.66**
BAYESR	0.64	0.62	0.57	0.81	0.84	**0.69**
Jersey						
GBLUP	0.56	0.62	0.67	0.64	0.76	**0.65**
BAYESR	0.56	0.69	0.71	0.76	0.79	**0.70**
Australian Reds						
GBLUP	0.20	0.16	0.11	0.32	0.34	**0.22**
BAYES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

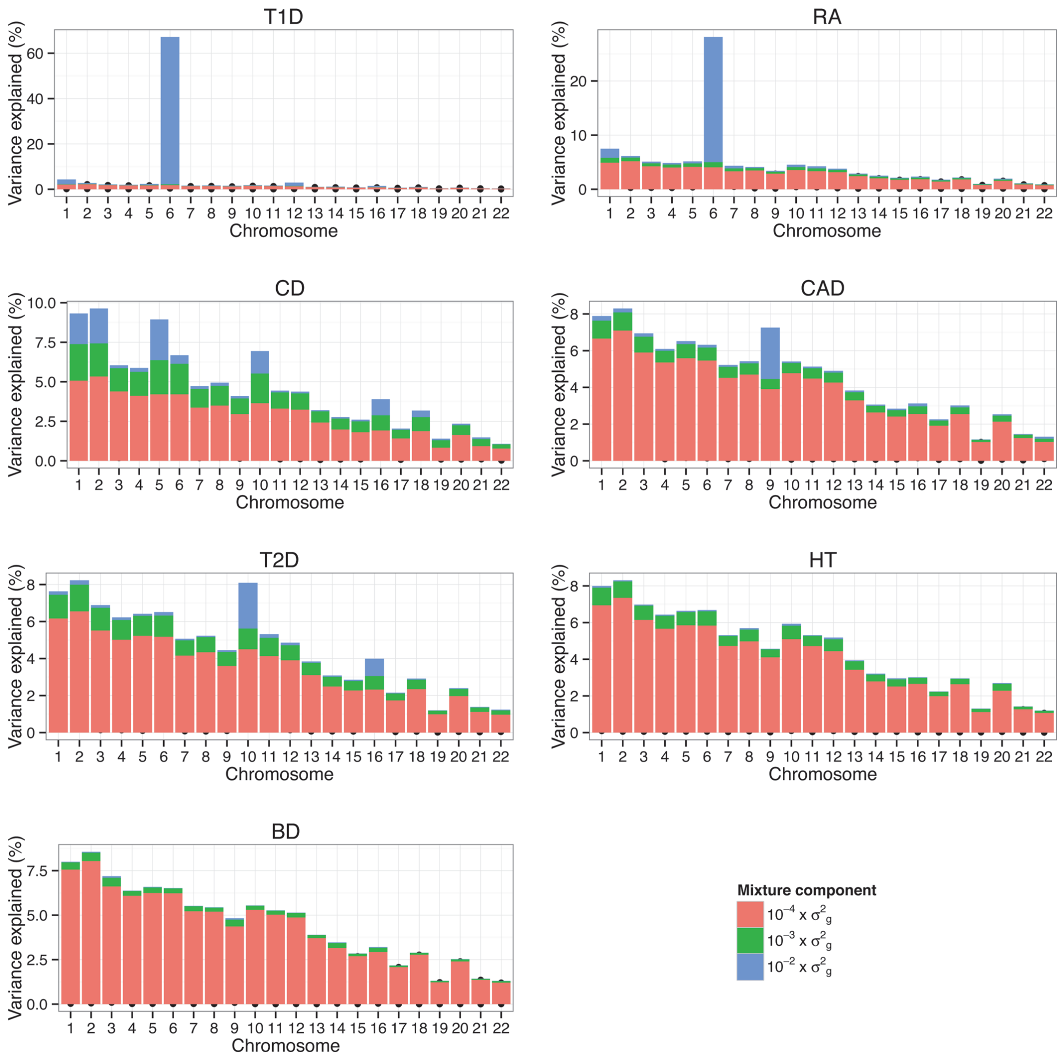
BayesR


BayesR_PY
600
400
200
0
GBLUP_PY
10
8
6
4
2
0
100
101
102
103
104
105
position, Mbp

BayesR application

Prediction of disease risk in humans





Moser et al PLoS Genetics 2015

Summary

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Bayesian methods for Genomic Prediction

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

Markov chain Monte Carlo (MCMC) is a technique to draw samples from a posterior distribution for Bayesian inference of model parameters.

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:

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

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]

Questions?

Practical 4: Bayesian methods

https://cnsgenomics.com/data/teaching/GNGWS25/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.