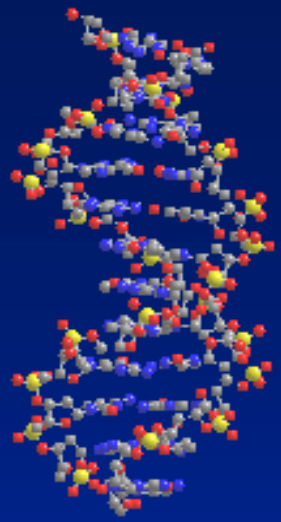
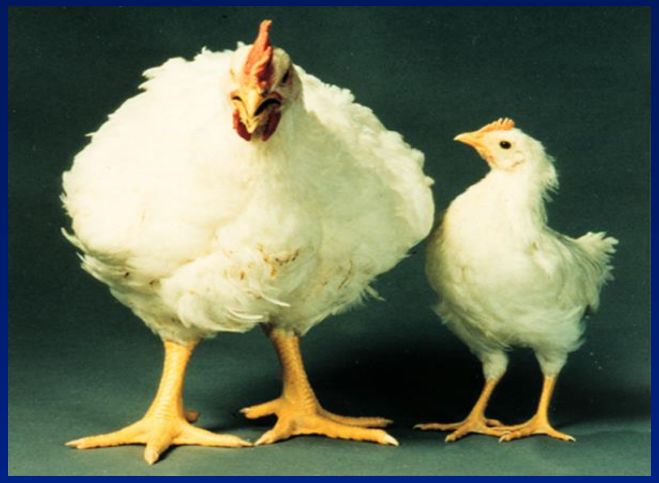
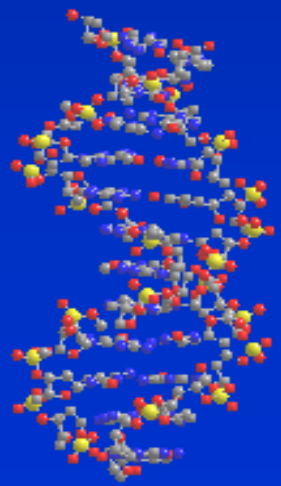




From sequence data to genomic prediction



Course overview

- Day 1
  - Introduction
  - Generation, quality control, alignment of sequence data
  - Detection of variants, quality control and filtering
- Day 2
  - Imputation from SNP array genotypes to sequence data
- Day 3
  - Genome wide association studies with SNP array and sequence variant genotypes
- Day 4 & 5
  - Genomic prediction with SNP array and sequence variant genotypes (BLUP and Bayesian methods)
  - Use of genomic selection in breeding programs

Genomic prediction

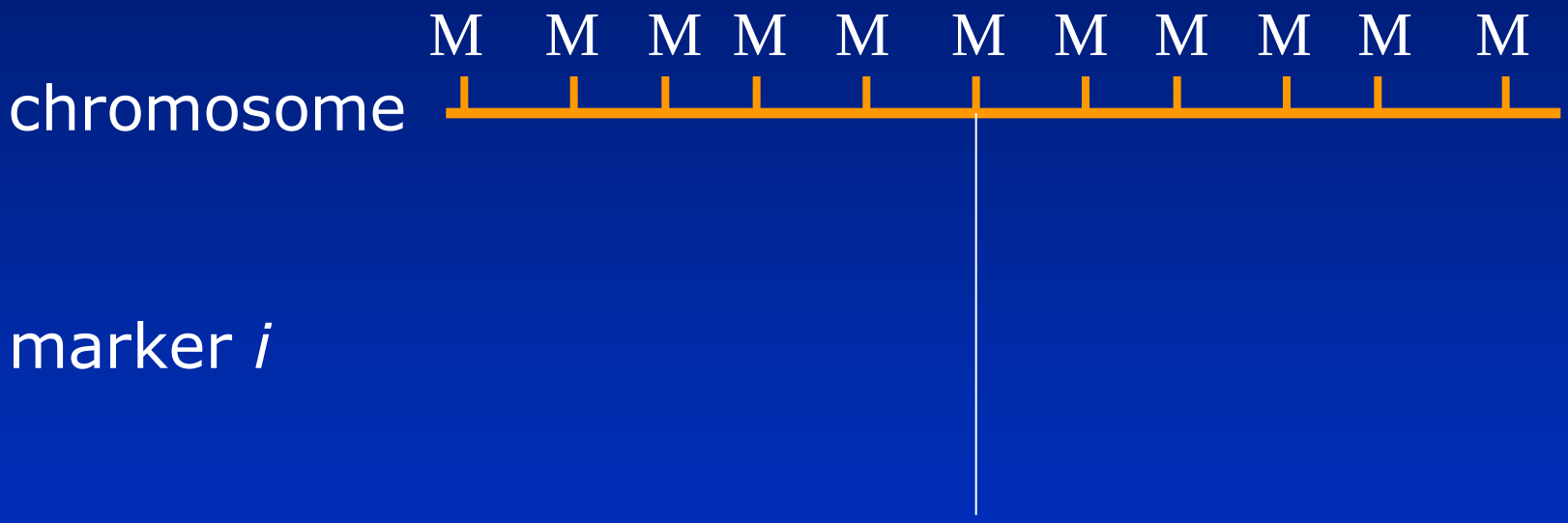
- Introduction to genomic selection
- Genomic prediction with BLUP
- Genomic prediction with Bayesian methods
- Examples in real data

Genomic prediction

- Problem marker assisted selection is only a proportion of genetic variance is tracked with markers
  - Eg. 10 QTL << 5% of the genetic variance
- Alternative is to trace all segments of the genome with markers
  - Divide genome into chromosome segments based on marker intervals?
  - Capture all QTL = all genetic variance

Genomic selection

Genomic prediction



Genomic prediction

Genomic prediction

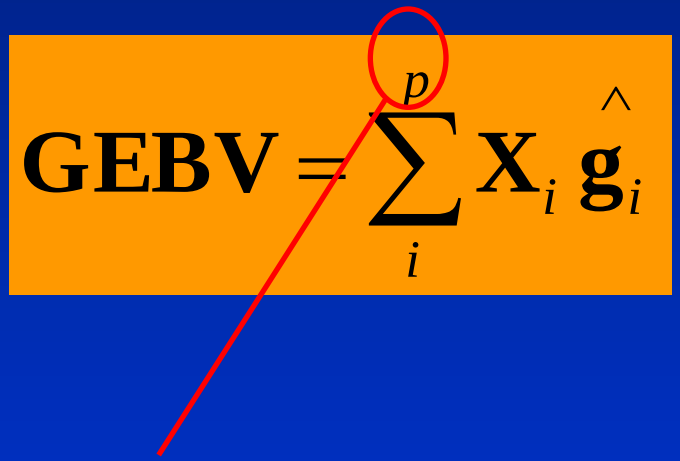
Genomic prediction

- Predict genomic breeding values as sum of effects over *all* SNP

$$\mathbf{GEBV} = \sum_{i}^{p} \mathbf{X}_i \hat{\mathbf{g}}_i$$

Genomic prediction

- Predict genomic breeding values as sum of effects over *all* SNP



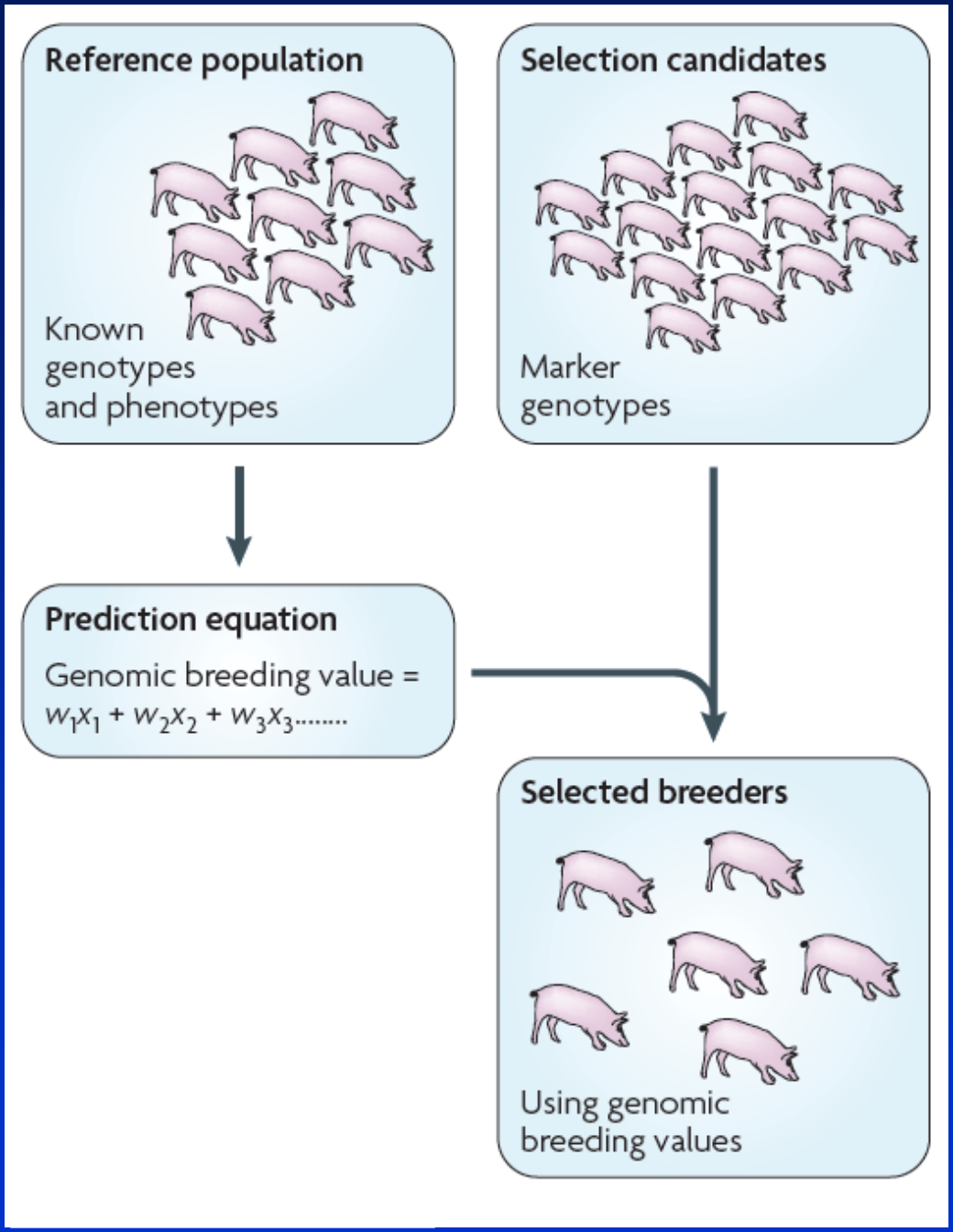
Genomic prediction

- Genomic prediction exploits linkage disequilibrium
  - Assumption is that markers picking up QTL and will have same effect across the whole population
- Possible within dense marker maps now available

Genomic prediction

- Genomic prediction avoids bias in estimation of effects due to multiple testing, as all effects fitted simultaneously

Genomic selection



Genomic prediction

- First step is to predict the chromosome segment effects in a reference population
- Number of effects >>> than number of records
- Eg. 50,000 SNPs
- From ~ 2000 records?
- Need methods that can deal with this

Genomic prediction

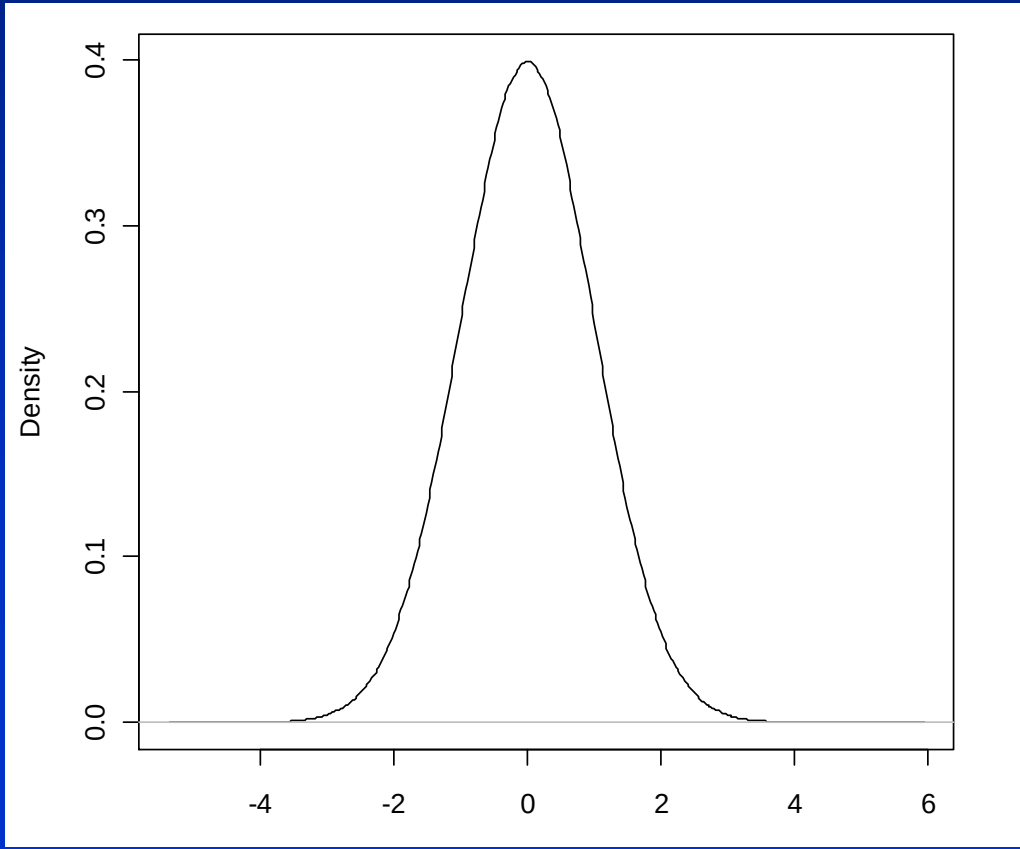
- BLUP = best linear unbiased prediction
- Model:

$$\mathbf{y} = \mu \mathbf{1_n} + \sum_{i=1}^{p} \mathbf{X_i g_i} + \mathbf{e}$$

- In BLUP we assume SNP effects come from normal distribution with same variance $E(\mathbf{g}) \sim N(0, \sigma_g^2)$

Genomic prediction with BLUP

- BLUP assumes normal distribution of SNP effects



Genomic prediction with BLUP

- **BLUP** = best linear unbiased prediction
- Then we can estimate segment effects as:

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

- $\lambda = \sigma_e^2 / \sigma_g^2$

Genomic prediction with BLUP

- Example
- A “simulated” data set
- Single chromosome, with 10 markers
- Phenotypes “simulated”
  - overall mean of 1
  - an effect for SNP 1 of 2 allele of 1
  - normally distributed error term with mean 0 and variance 1.

Genomic prediction with BLUP

- Example

		X									
Animal	Y		1	2	3	4	5	6	7	8	9	10
	1	0.19	0	0	0	0	0	0	1	2	0	2
	2	1.23	1	0	0	1	1	1	2	1	0	1
	3	0.86	1	0	0	1	0	0	1	1	1	1
	4	1.23	1	1	1	1	0	1	2	1	1	1
	5	0.45	0	1	1	1	1	1	2	1	0	1

- 10 SNPs
- Only 5 phenotypic records.

Genomic prediction with BLUP

- Example

		X								
Animal	Y		1	2	3	4	5	6	7	8	9	10
1		0.19	0	0	0	0	0	0	1	2	0	2
2		1.23	1	0	0	1	1	1	2	1	0	1
3		0.86	1	0	0	1	0	0	1	1	1	1
4		1.23	1	1	1	1	0	1	2	1	1	1
5		0.45	0	1	1	1	1	1	2	1	0	1

- Assume value of 1 for λ
- 1_n = [1 1 1 1 1]

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

Genomic prediction with BLUP

- Example

Mean	0.47
SNP1	0.29
SNP2	-0.05
SNP3	-0.05
SNP4	0.08
SNP5	-0.02
SNP6	0.13
SNP7	0.13
SNP8	-0.08
SNP9	0.11
SNP10	-0.08

Genomic prediction with BLUP

- Now we want to predict GEBV for a group of young animals without phenotypes.

$$\mathbf{GEBV} = \mathbf{X}\hat{\mathbf{g}}$$

- We have the g_hat, and we can get **X** from their haplotypes (after genotyping)……

Progeny	X								
1	1	1	1	1	1	1	2	1	0	1
2	1	0	0	1	1	1	2	1	0	1
3	1	0	0	1	1	1	2	1	0	1
4	1	0	0	1	1	1	2	1	0	1
5	0	0	0	0	0	0	1	2	0	2

Genomic prediction with BLUP

- GEBV

$$\mathbf{GEBV} = \mathbf{X}\hat{\mathbf{g}}$$

X									$\hat{\mathbf{g}}$	GEBV
1	1	1	1	1	1	2	1	0	1	0.29	0.47
1	0	0	1	1	1	2	1	0	1	-0.05	0.58
1	0	0	1	1	1	2	1	0	1	-0.05	0.58
1	0	0	1	1	1	2	1	0	1	0.08	0.58
0	0	0	0	0	0	1	2	0	2	-0.02	-0.20
										0.13	
										0.13	
										-0.08	
										0.11	
										-0.08	

Genomic prediction with BLUP

- Where do we get σ_g^2 from?
- Can estimate total additive genetic variance and divide by number of segments, eg $\sigma_g^2 = \sigma_a^2 / p$
- If using single markers take account of heterozygosity

$$\sigma_g^2 = \sigma_a^2 / 2\sum_{i=1}^{p} q_i(1-q_i)$$

- Ridge regression (Bayesian approach)
- Cross validation

Genomic prediction with BLUP

- An equivalent model
- If there are many QTLs whose effects are normally distributed with constant variance,
- Then genomic selection equivalent to replacing the expected relationship matrix with the realised or genomic relationship matrix (**G**) estimated from DNA markers in normal BLUP equations.
  - G_{ij} = proportion of genome that is IBD between animals i and j

Genomic prediction with BLUP

- An equivalent model
- Rescale X to account for allele frequencies
  - $w_{ij} = x_{ij} - 2p_j$
- Then breeding values are
  - **v = Wg** ($\mathbf{GEBV} = \mathbf{X}\hat{\mathbf{g}}$)
- And

$$\mathbf{G} = \mathbf{WW}' / 2\sum_{j=1}^{p} p_j(1-p_j)$$

- Then

$$V(\mathbf{v}) = \mathbf{G}\sigma_a^2$$

Genomic prediction with BLUP

- An equivalent model

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{Zv} + \mathbf{e}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'Z} \\ \mathbf{Z'1_n} & \mathbf{Z'Z} + \mathbf{G}^{-1}\frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{Z'y} \end{bmatrix}$$

Genomic prediction with BLUP

- An equivalent model
  - Model 1.

$$\mathbf{y} = \mathbf{1_n}\mu + \sum_{i=1}^{p} \mathbf{X_i g_i} + \mathbf{e}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'X} \\ \mathbf{X'1_n} & \mathbf{X'X} + \mathbf{I}\frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{X'y} \end{bmatrix}$$

$$\mathbf{GEBV} = \mathbf{X}\hat{\mathbf{g}}$$

  - Model 2.

Genomic prediction with BLUP

- An equivalent model
  - Model 1.

$$\mathbf{y} = \mathbf{1_n}\mu + \sum_{i=1}^{p} \mathbf{X_i g_i} + \mathbf{e}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'X} \\ \mathbf{X'1_n} & \mathbf{X'X} + \mathbf{I}\frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{X'y} \end{bmatrix}$$

$$\mathbf{GEBV} = \mathbf{X}\hat{\mathbf{g}}$$

  - Model 2.

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{Zv} + \mathbf{e}$$

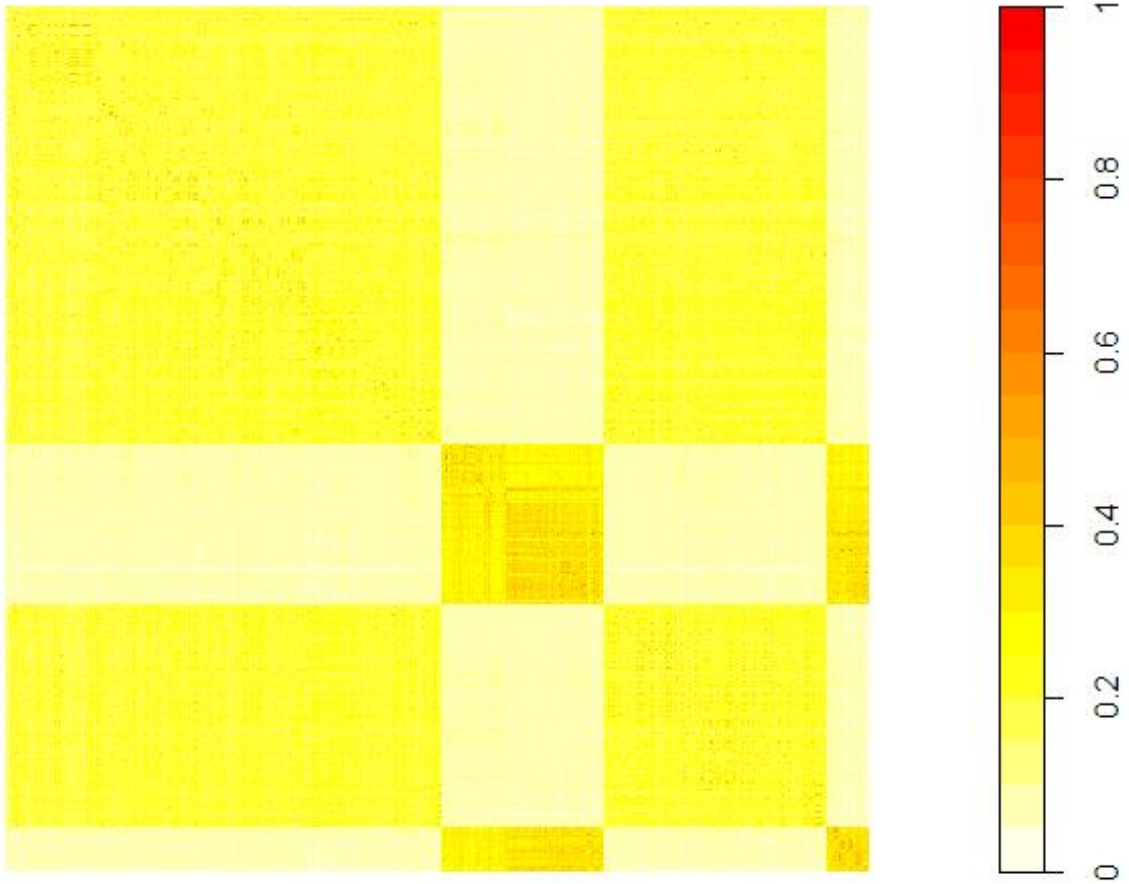
$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_v'Z} \\ \mathbf{Z'1_n} & \mathbf{Z'Z} + \mathbf{G}^{-1}\frac{\sigma_e^2}{\sigma_v^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1_n'y} \\ \mathbf{Z'y} \end{bmatrix}$$

Holstein reference *n = 781*

Jersey reference *n = 287*

Holstein validation *n = 400*

Jersey validation *n = 77*



Genomic prediction with BLUP

- An equivalent model
- Why use model 2 (GBLUP).
  - If number of markers >>> large than number of animals, more computationally efficient
  - Can be integrated into national evaluations more readily?
  - Calculate accuracy of GEBV from inverse coefficient matrix

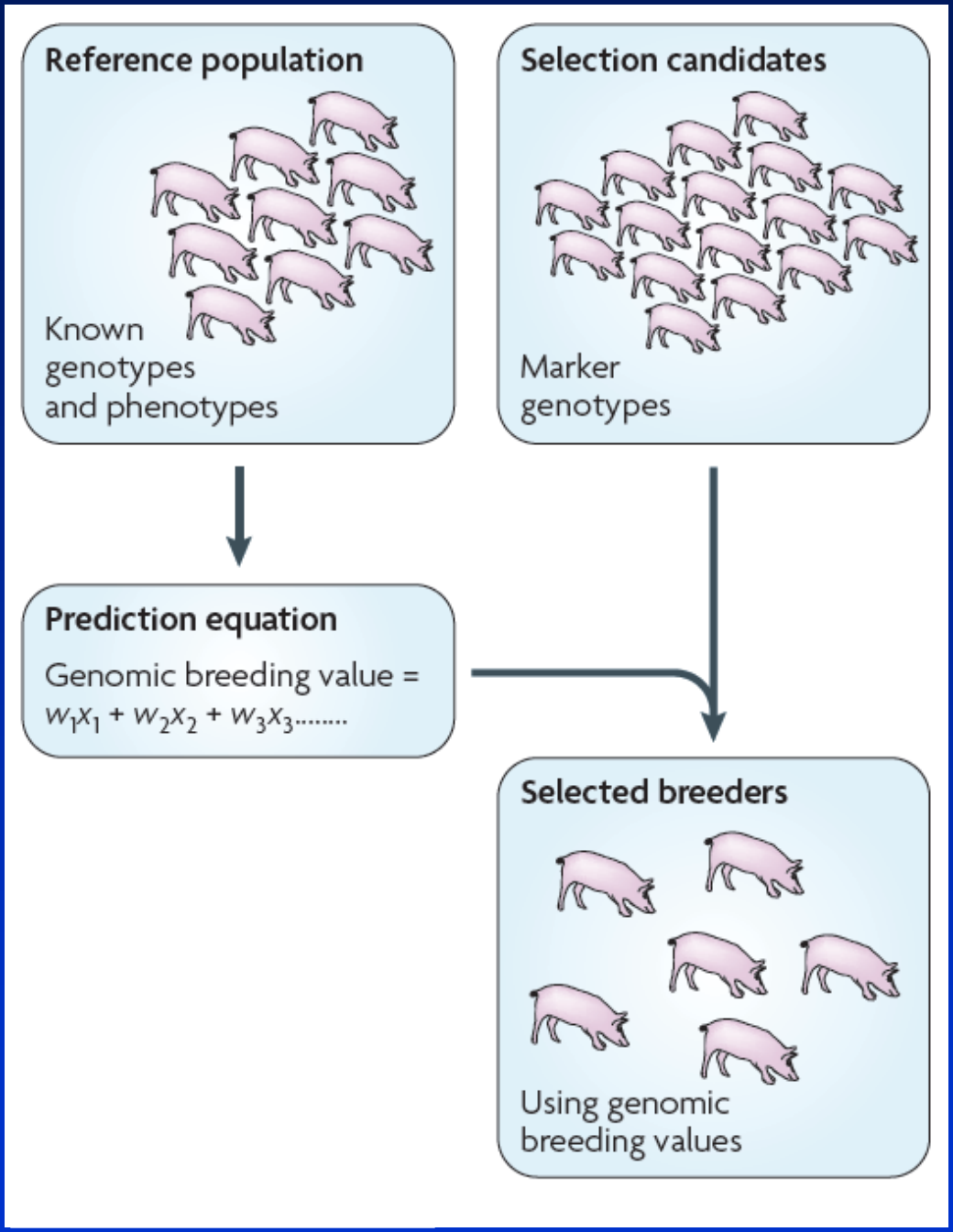
Genomic prediction

- Introduction to genomic selection
- Genomic prediction with BLUP
- Genomic prediction with Bayesian methods
- Examples in real data

Genomic prediction

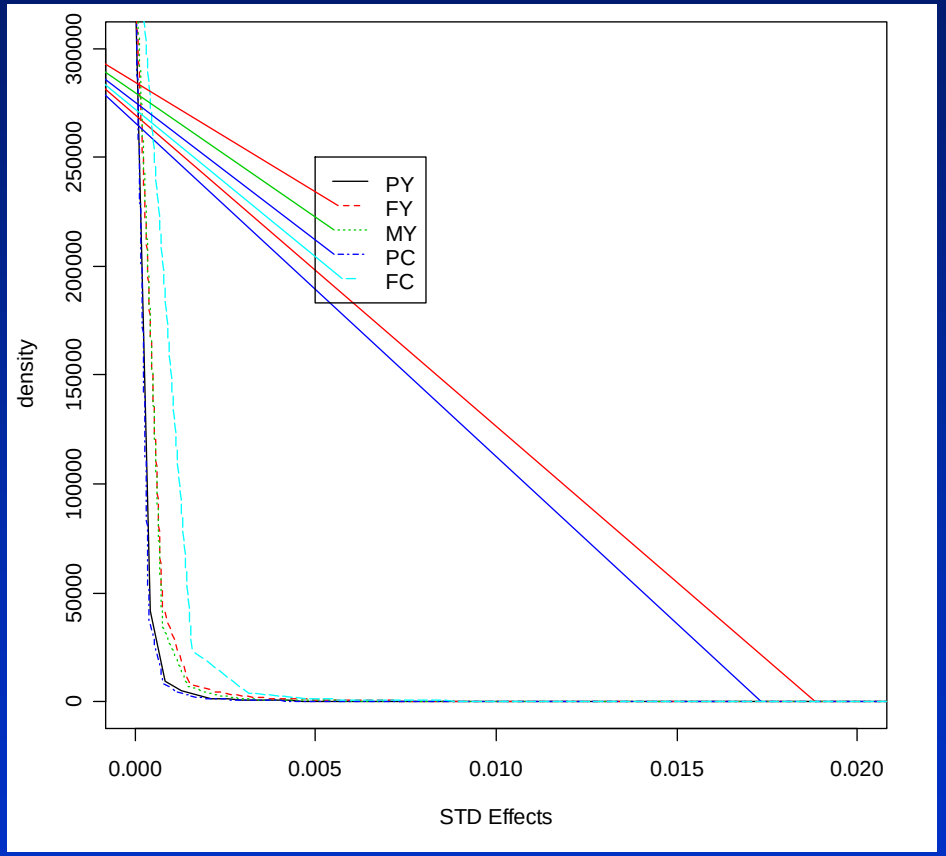
- Alternative assumptions regarding the distribution of SNP effects
- Introduction to Bayesian methods
- Genomic prediction with Bayesian methods
- Comparison of accuracy of methods

Genomic selection



Alternative prior assumptions for SNP effects

- BLUP assumes normally distributed QTL effects
- Does not match prior knowledge of distributions of QTL effects for some traits



Alternative prior assumptions for SNP effects

- Students t distribution?
  - BayesA
- Many zero effects and proportion Students t distribution?
  - BayesB
- Many zero effect and rest normal distribution
  - BayesCpi
- Double exponential effects
  - BayesianLASSO
- Multiple normal distributions
  - BayesMulti, BayesR

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)

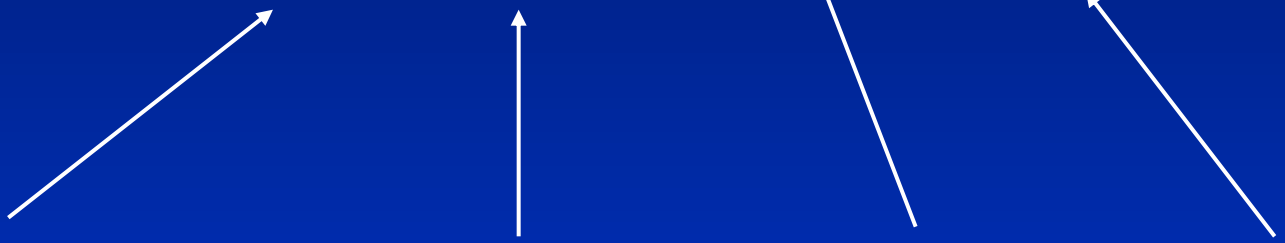
Is proportional to

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

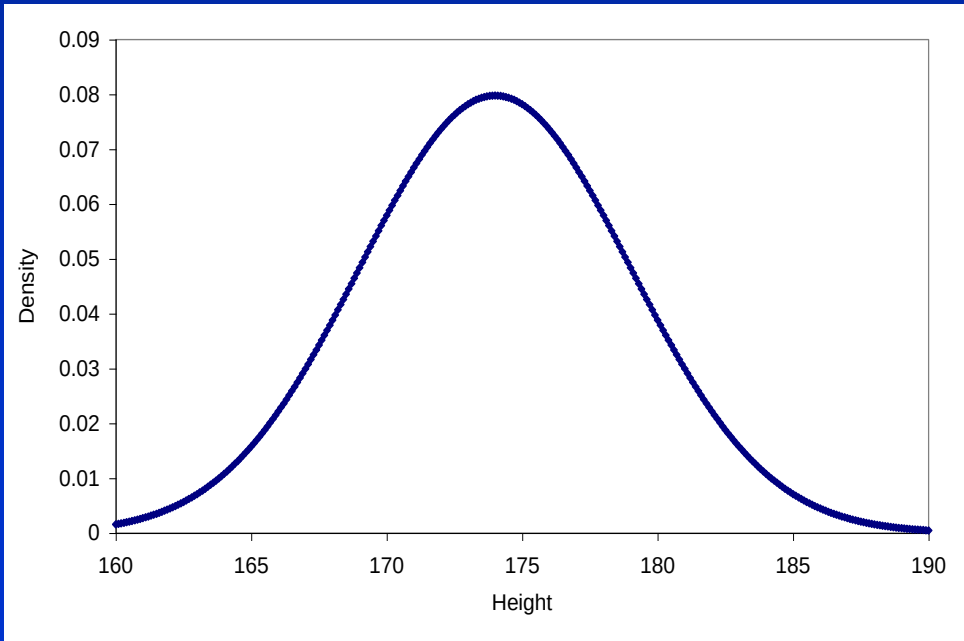
y=average height + e

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

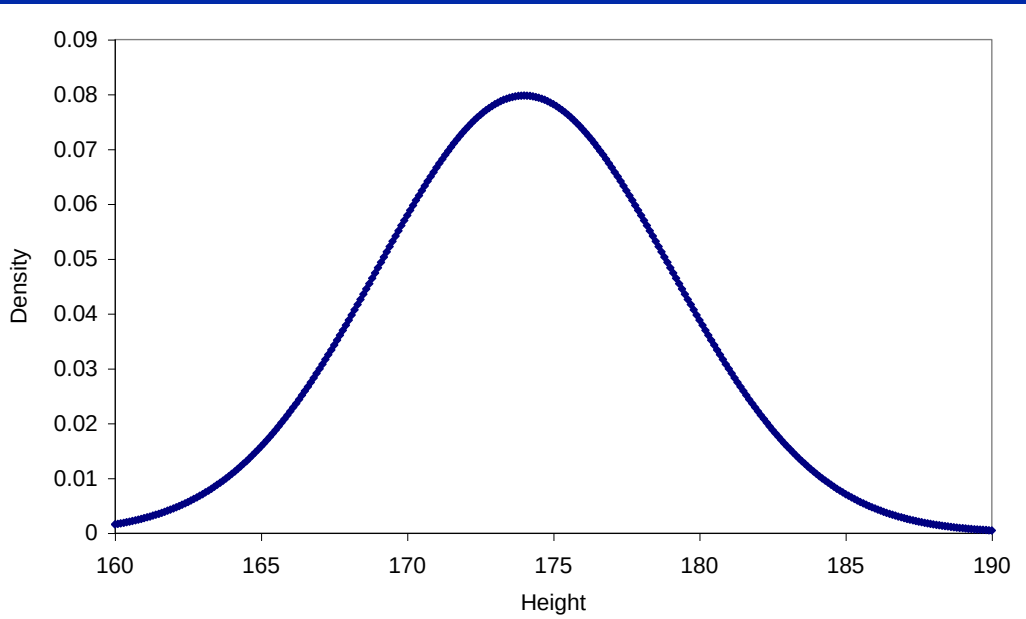
$$P(x \mid y) \propto P(y \mid x)P(x)$$

Prior probability of x (average height)

From the data……

$$\bar{x} = 178$$

$$s.e = 5$$



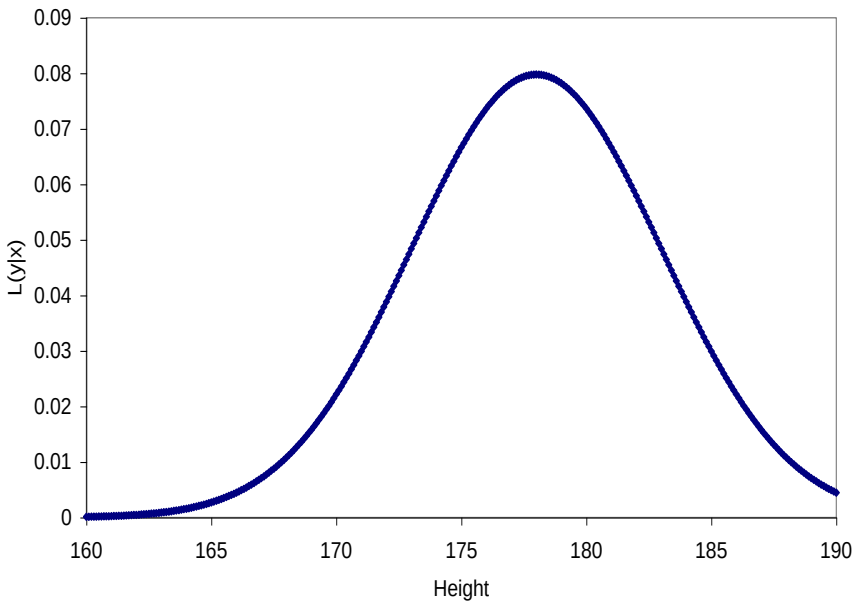
Bayesian methods

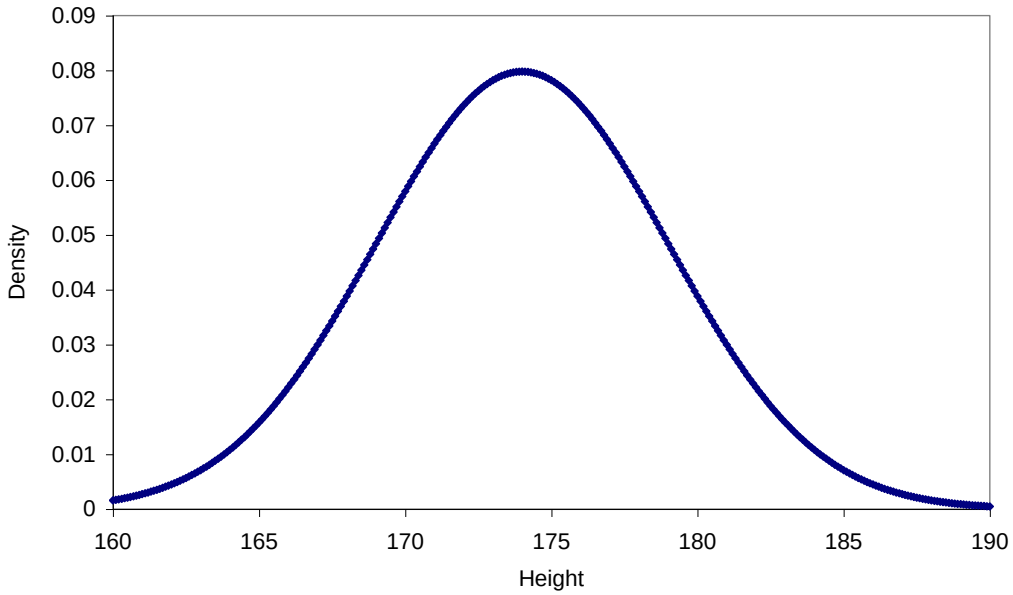
- Bayes theorem

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

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

Prior probability of x (average height)





Bayesian methods

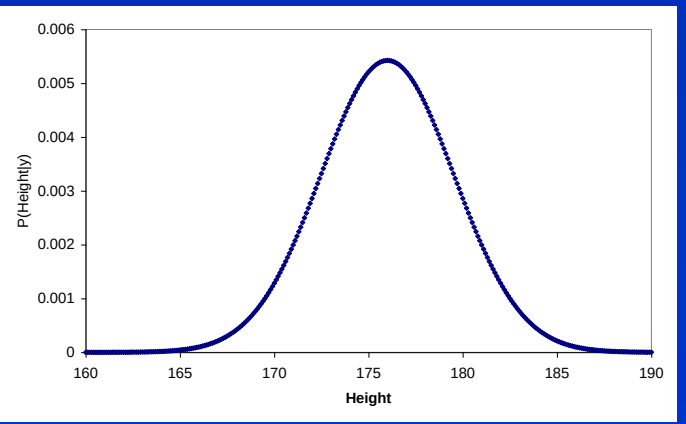
- Bayes theorem

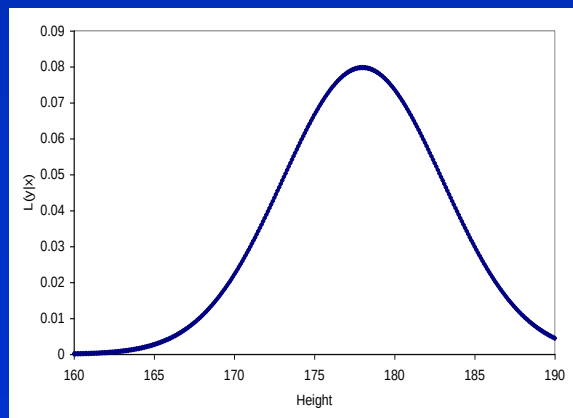
$$P(x \mid y) \propto P(y \mid x)P(x)$$

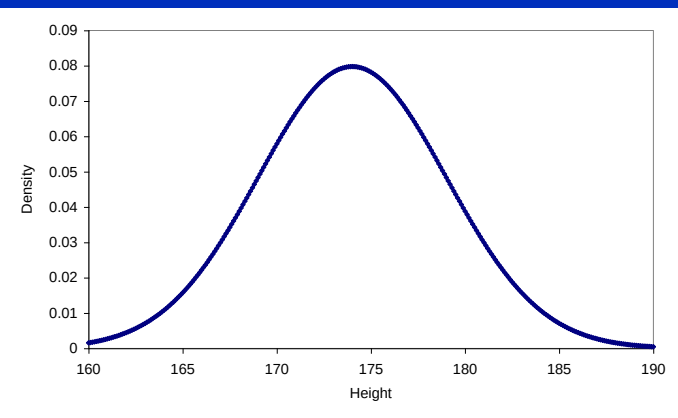
P(x|y) mean = 176cm

L(y|x)

P(x)







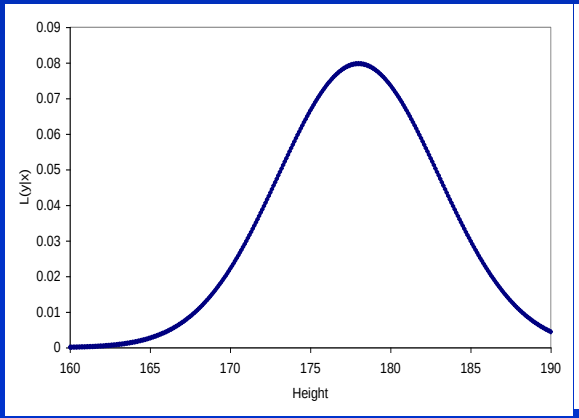
Bayesian methods

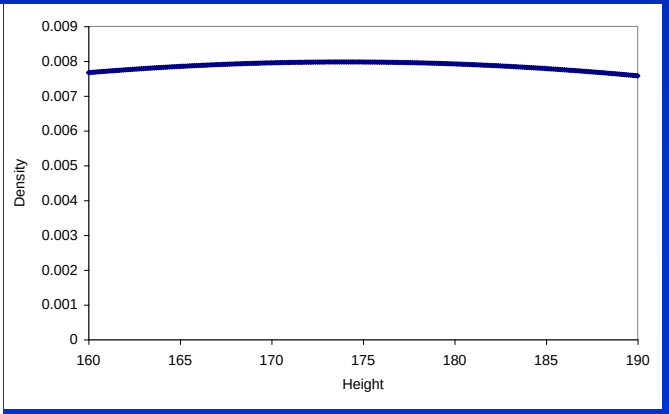
- Bayes theorem
- Less certainty about prior information? Use *less* informative (flat) prior

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

L(y|x)

P(x)





Bayesian methods

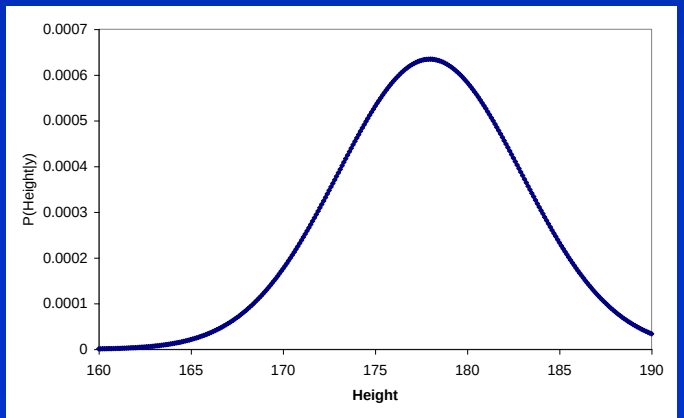
- Bayes theorem
- Less certainty about prior information? Use *less* informative (flat) prior

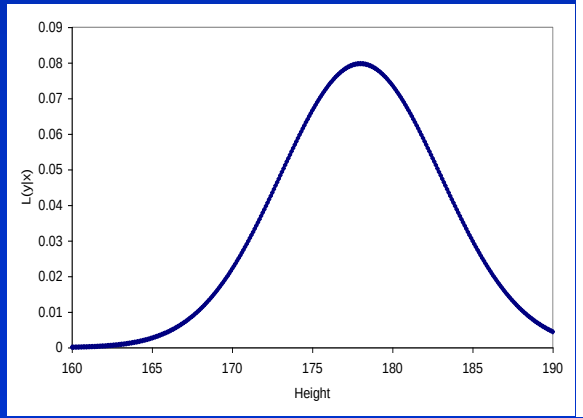
$$P(x \mid y) \propto P(y \mid x)P(x)$$

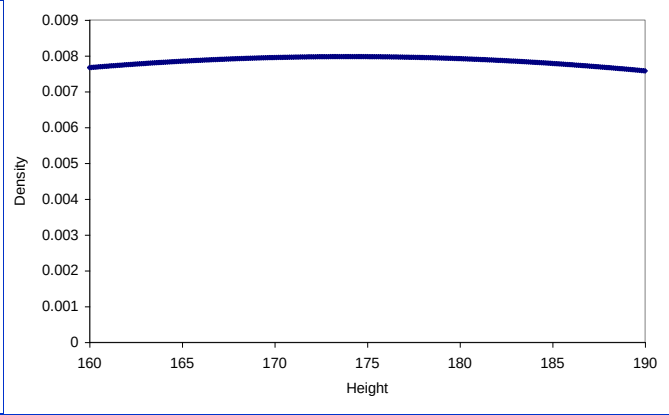
P(x|y) mean = 178cm

L(y|x)

P(x)







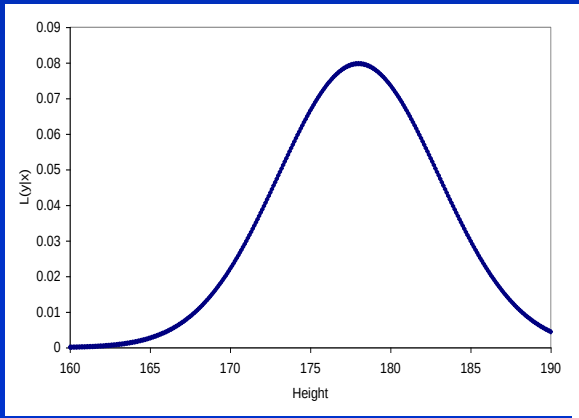
Bayesian methods

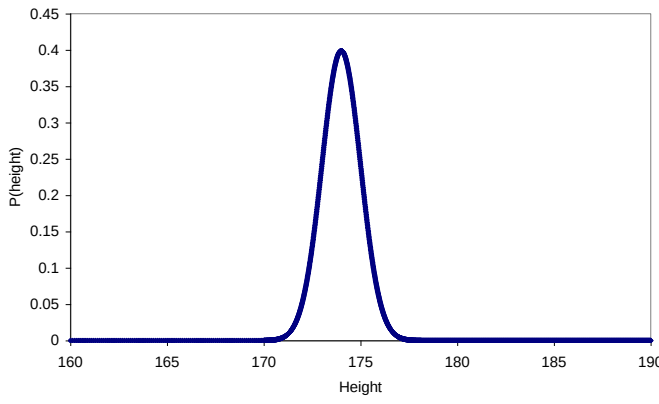
- Bayes theorem
- More certainty about prior information? Use *more* informative prior

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

L(y|x)

P(x)



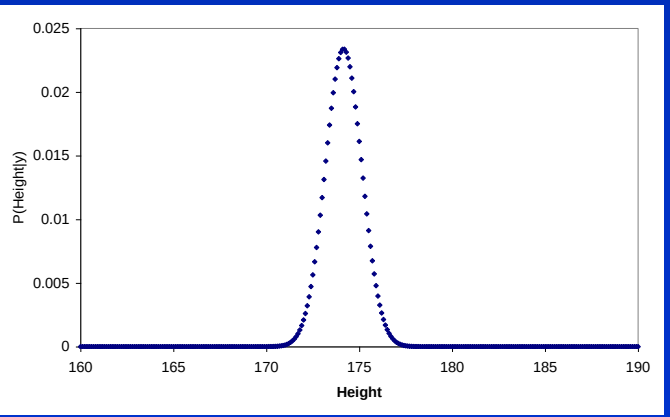


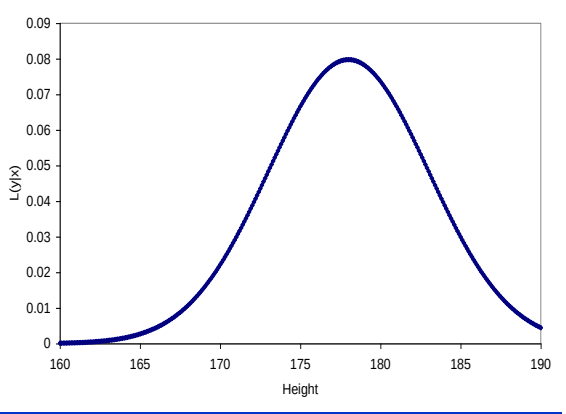
Bayesian methods

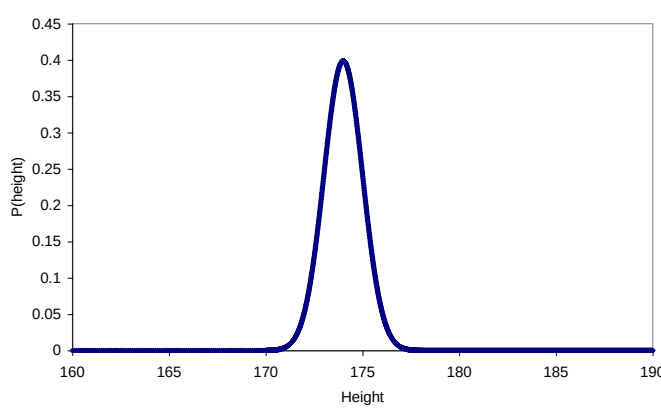
- Bayes theorem
- More certainty about prior information? Use *more* informative prior

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

P(x|y) mean = 174.5cm L(y|x) P(x)





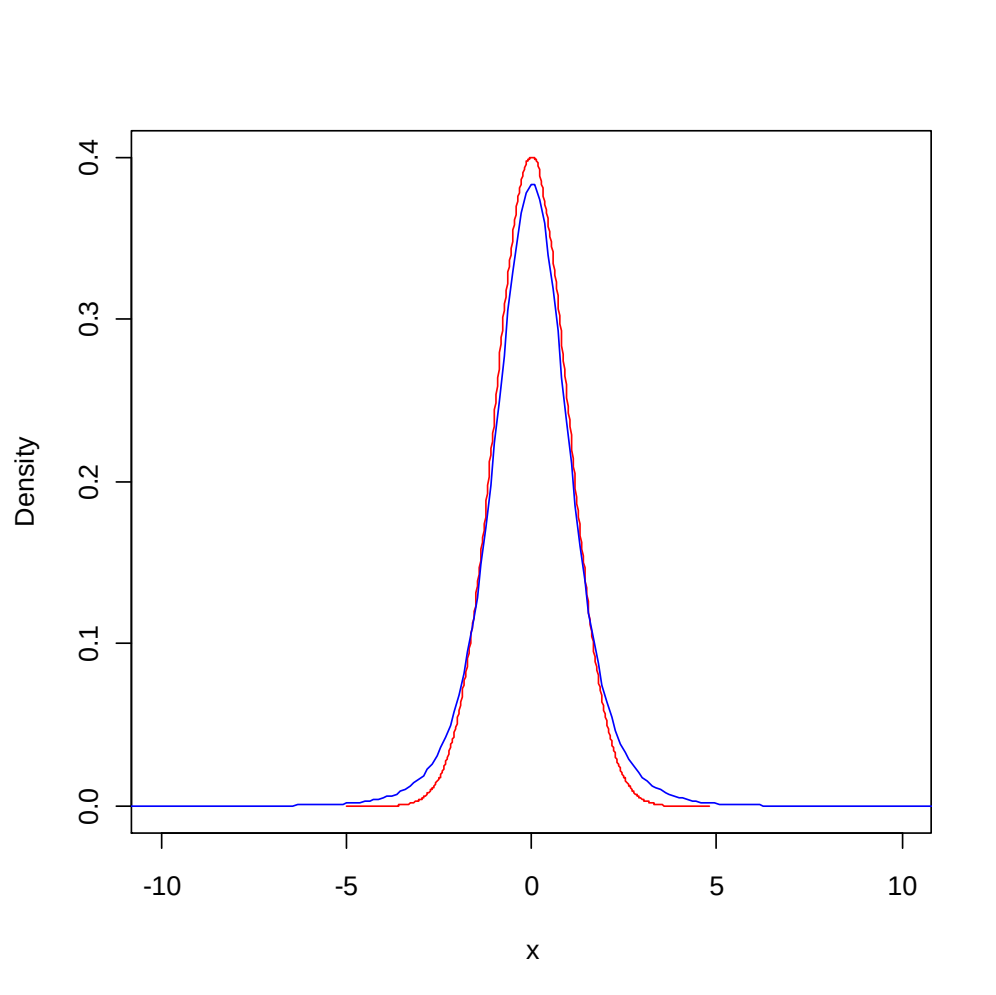


Genomic prediction

- Alternative assumptions regarding the distribution of SNP effects
- Introduction to Bayesian methods
- Genomic prediction with Bayesian methods
- Comparison of accuracy of methods

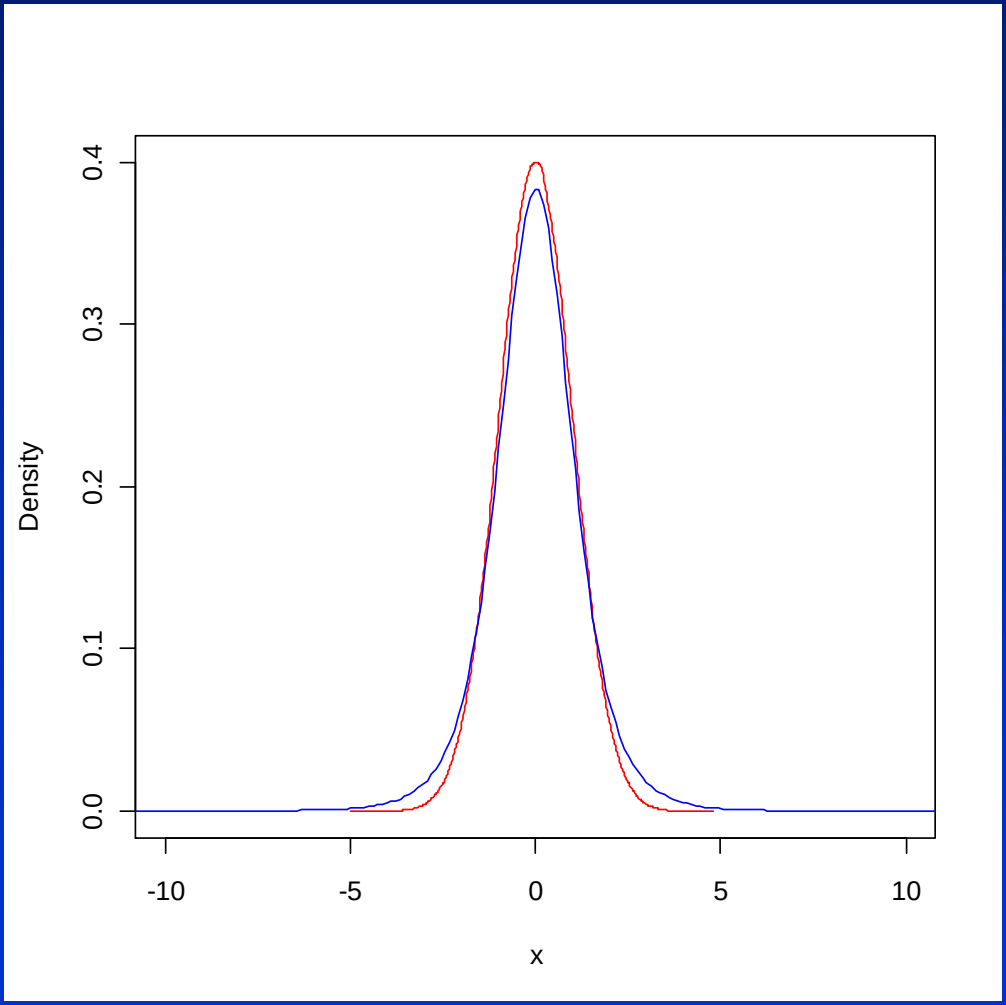
Genomic selection

- For some traits prior knowledge suggests t-distribution of effects
- How to incorporate this into our predictions?



Genomic selection

- The **t distribution** can be presented as a two level hierarchical model
- Allow different variances between chromosome segments
- Assume a distribution of these variances
- Computationally easier to deal with than original form

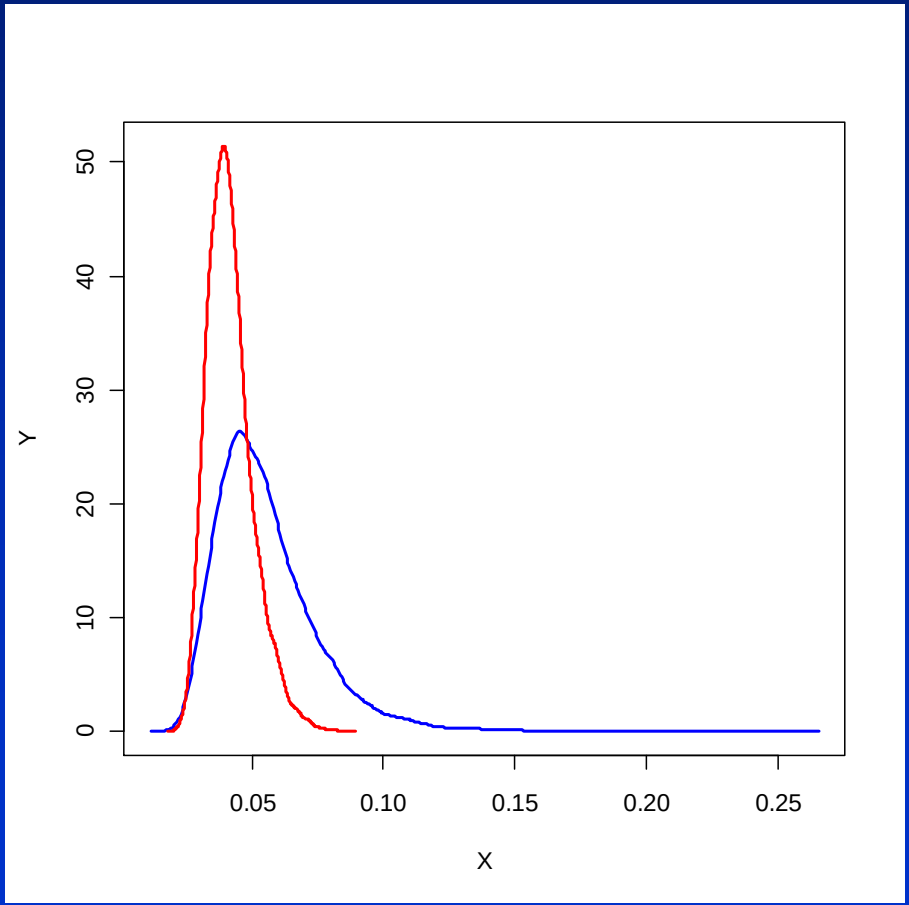


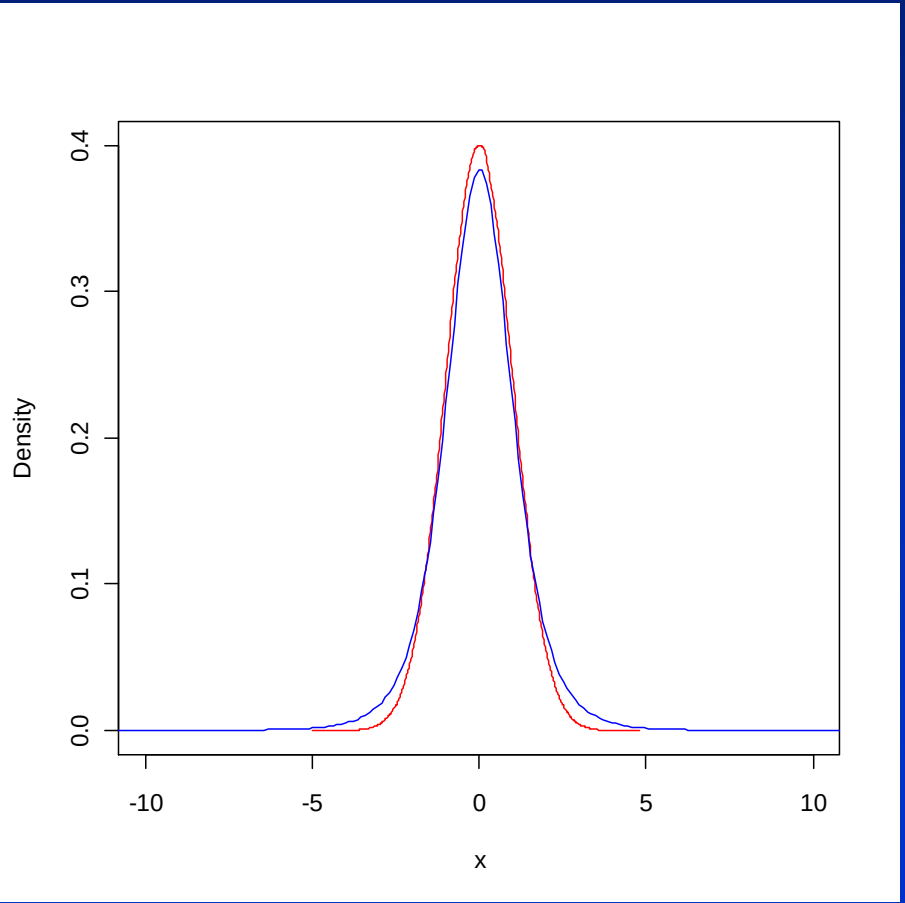
Bayesian methods

- Now lets allow different variances of chromosome segment effects

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}}_1 \\ . \\ \hat{\mathbf{g}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'X_1} & . & \mathbf{1_n'X_p} \\ \mathbf{X_1'1_n} & \mathbf{X_1'X_1} + \mathbf{I}\frac{\sigma_e^2}{\sigma_{g1}^2} & . & \mathbf{X_1'X_p} \\ . & . & . & . \\ \mathbf{X_p'1_n} & \mathbf{X_p'X_1} & . & \mathbf{X_p'X_p} + \mathbf{I}\frac{\sigma_e^2}{\sigma_{gp}^2} \end{bmatrix}^{-1} \begin{bmatrix} 1_n'y \\ X_1'y \\ . \\ X_p'y \end{bmatrix}$$

Distribution of σ_{gj}^2 --→ Distribution of g_j





Bayesian methods

- Now lets allow different variances of chromosome segment effects
- Need two levels of models
  - Data

$$P(\mathbf{g}, \mu \mid y) \propto P(y \mid \mathbf{g}, \mu) P(\mathbf{g}, \mu)$$

  - Variances of chromosome segment effects

$$P(\sigma^2_{gi} \mid g_i) \propto P(g_i \mid \sigma^2_{gi}) P(\sigma^2_{gi})$$

Bayesian methods

- Now lets allow different variances of chromosome segment effects
- Data

$$P(\mathbf{g}, \mu \mid y) \propto P(y \mid \mathbf{g}, \mu) P(\mathbf{g}, \mu)$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}}_1 \\ . \\ \hat{\mathbf{g}}_p \end{bmatrix} = \begin{bmatrix} \mathbf{1_n'1_n} & \mathbf{1_n'X_1} & . & \mathbf{1_n'X_p} \\ \mathbf{X_1'1_n} & \mathbf{X_1'X_1} + \mathbf{I}\frac{\sigma_e^2}{\sigma_{g1}^2} & . & \mathbf{X_1'X_p} \\ . & . & . & . \\ \mathbf{X_p'1_n} & \mathbf{X_p'X_1} & . & \mathbf{X_p'X_p} + \mathbf{I}\frac{\sigma_e^2}{\sigma_{gp}^2} \end{bmatrix}^{-1} \begin{bmatrix} 1_n'y \\ X_1'y \\ . \\ X_p'y \end{bmatrix}$$

Bayesian methods

- Variances of chromosome segments

$$P(\sigma^2_{gi} \mid g_i) \propto P(g_i \mid \sigma^2_{gi}) P(\sigma^2_{gi})$$

- Note that these variance components are not the parameters of interest
- However they are useful intermediates to arrive at better inferences for the g_i
- Amount of shrinkage of effects varies between segments

Bayesian methods

- Variances of chromosome segments

$$P(\sigma^2_{gi} \mid g_i) \propto P(g_i \mid \sigma^2_{gi}) P(\sigma^2_{gi})$$

- Prior?
  - Inverted chi square convenient for variances

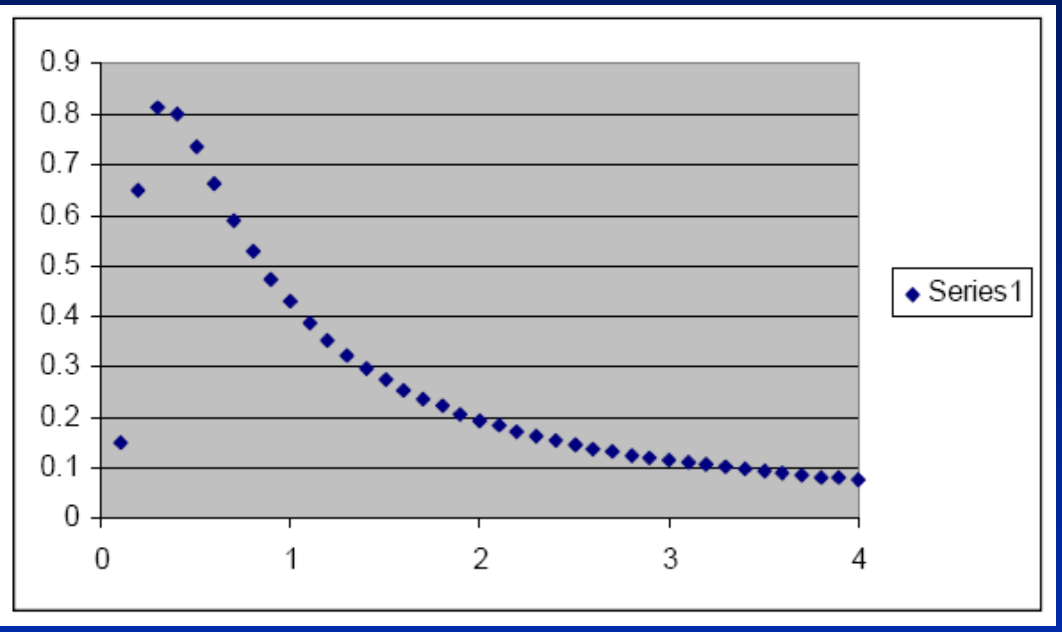
Bayesian methods

- Prior?
  - Inverted chi square convenient for variances
  - An inverted chi square with v degrees of freedom and scaled by S^2, eg.

$$S^2 / \chi_v^2$$

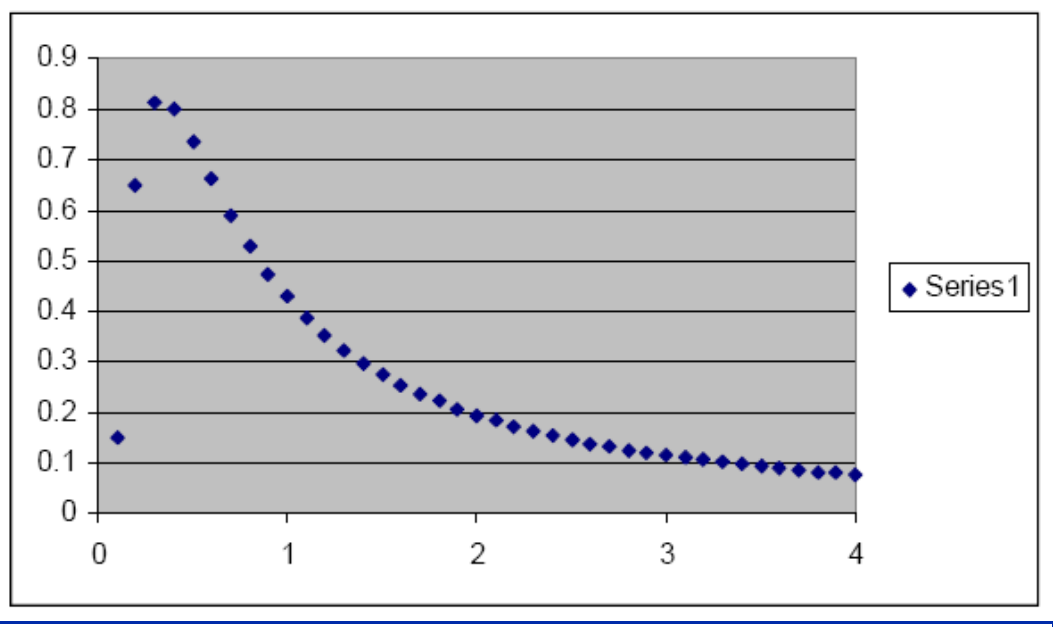
  - Describes a distribution with
    - mean $$vS^2/(v-2)$$
    - variance $$\frac{2v^2S^4}{(v-2)^2(v-4)}$$
  - Larger v, more informative prior = more belief about variance

Bayesian methods

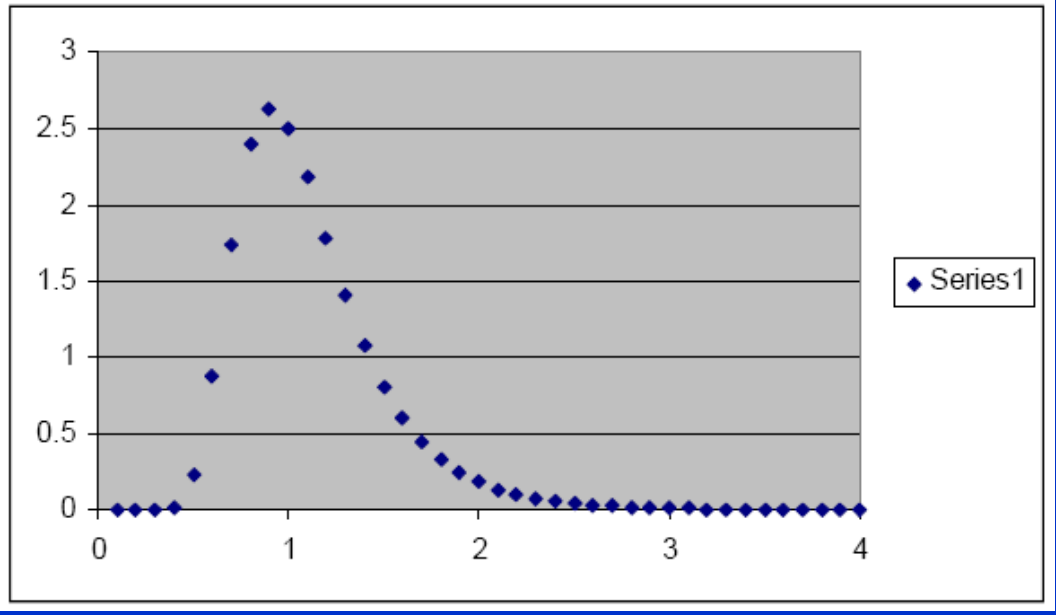


v=2

Bayesian methods



ν=2



ν=20

Bayesian methods

- Variances of chromosome segments

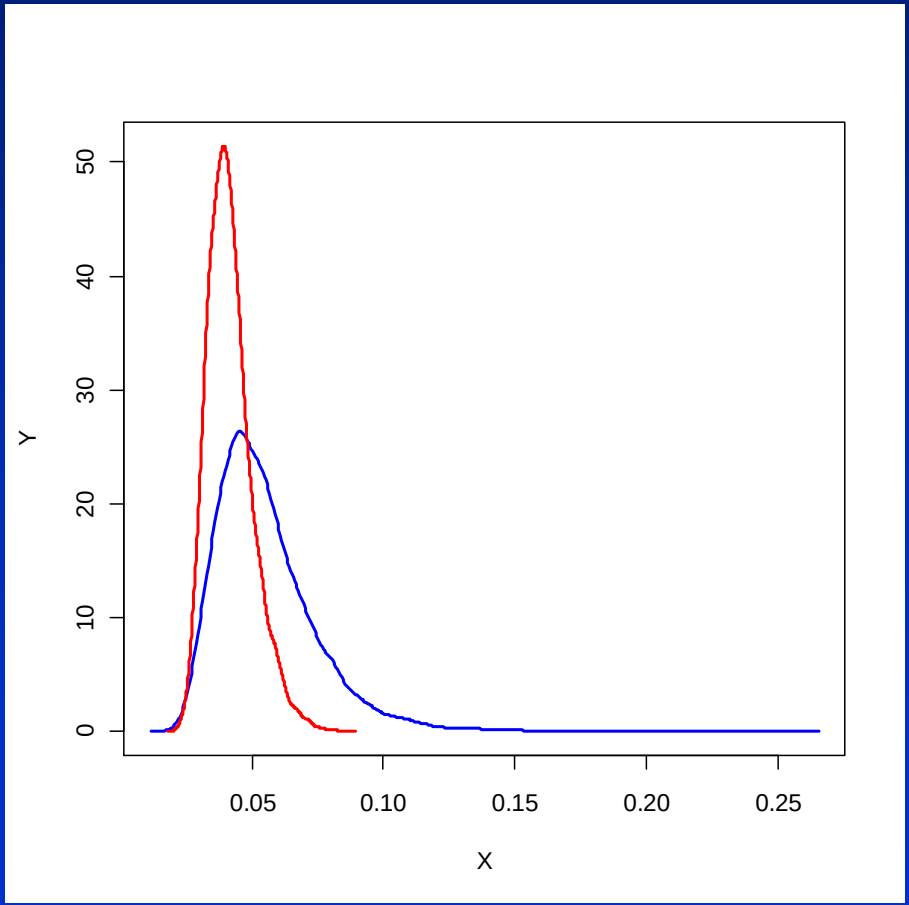
$$P(\sigma^2_{gi} \mid g_i) \propto P(g_i \mid \sigma^2_{gi}) P(\sigma^2_{gi})$$

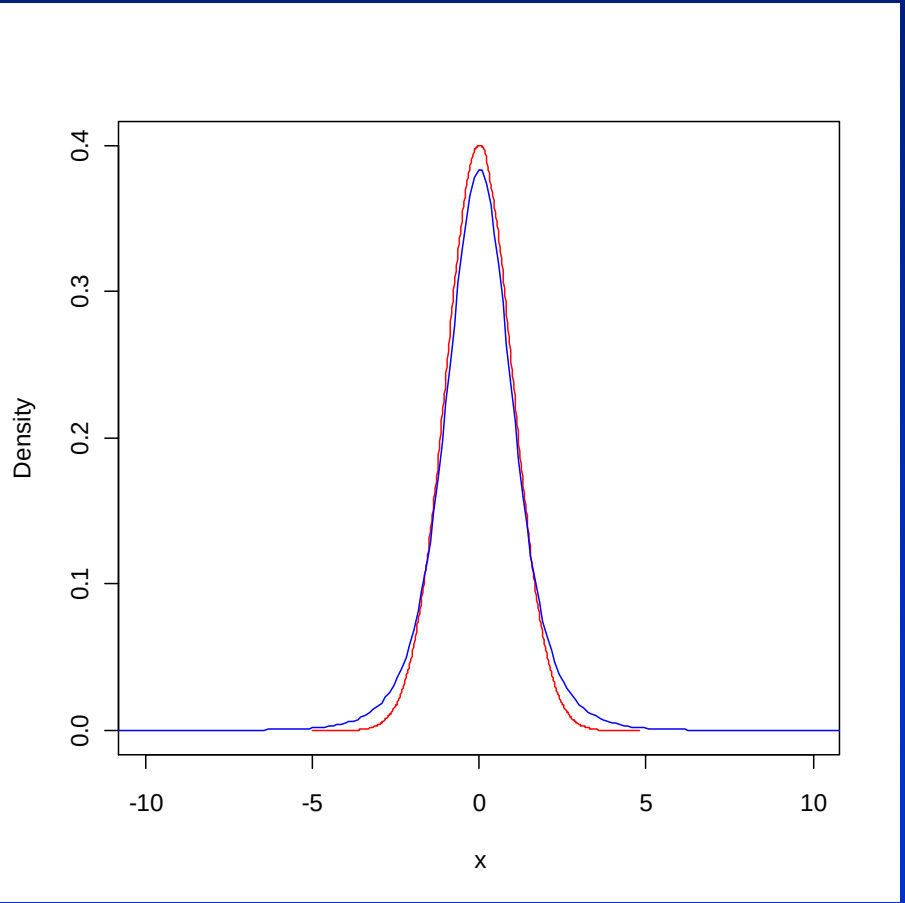
- Prior?

$$S^2 / \chi^2_v$$

- We can choose v and S^2 so that the prior reflects our knowledge that there are many QTL of small effect and few of large effect

Distribution of σ_{gj}^2 --→ Distribution of g_j



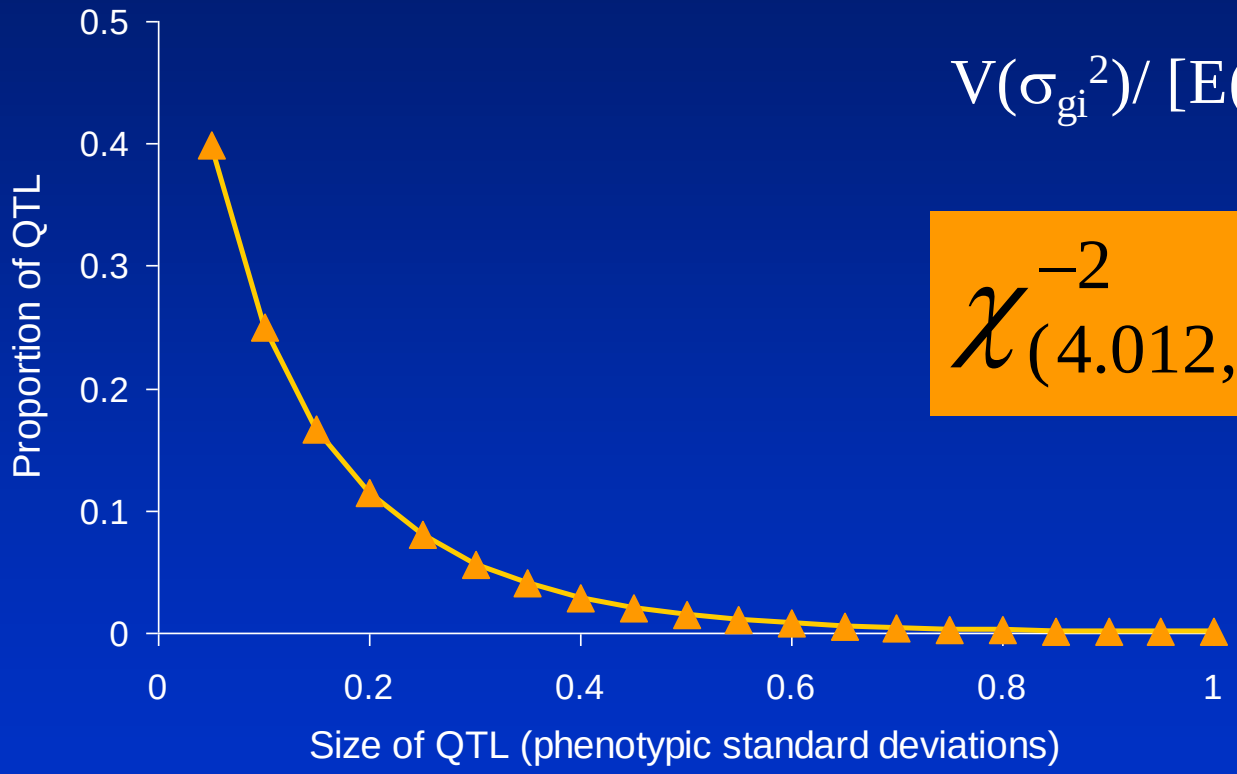


Bayesian methods

$E(\sigma_{gi}^2)=S/(v-2)$

$V(\sigma_{gi}^2)/\ [E(\sigma gi2)]^2=2/(v-4)$

$$\chi^{-2}_{(4.012,0.002)}$$



Bayesian methods

- Variances of chromosome segments

$$P(\sigma^2_{gi} \mid \mathbf{g_i}) \propto P(\mathbf{g_i} \mid \sigma^2_{gi}) P(\sigma^2_{gi})$$

- Posterior?
  - An advantage of choosing the inverse chi-square distribution for the prior is that the posterior will also be an inverse chi-square distribution
    - Degrees of freedom = prior + data
    - Scaling factor = sums of squares prior (S^2) + sums of squares from data

Bayesian methods

- Variances of chromosome segments

$$P(\sigma^2_{gi} \mid \mathbf{g_i}) \propto P(\mathbf{g_i} \mid \sigma^2_{gi}) P(\sigma^2_{gi})$$

- Posterior?
  - n_i = number of haplotype effects

$$\chi^{-2}_{(v+n_i,\, S^2+\mathbf{g_i}'\mathbf{g_i})}$$

Bayesian methods

- Variances of chromosome segments

$$P(\sigma^2_{gi} \mid \mathbf{g_i}) \propto P(\mathbf{g_i} \mid \sigma^2_{gi}) P(\sigma^2_{gi})$$

- Posterior?

$$\chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$$

- But posterior cannot be estimated directly, dependent on $\mathbf{g_i}$!!

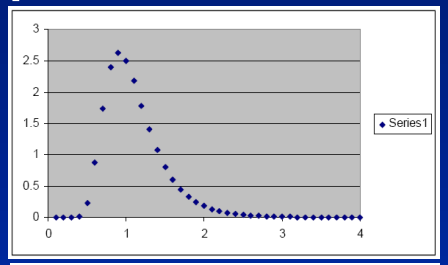
Bayesian methods

- Solution is to use Gibbs sampling
  - Draw samples from the posterior distributions of parameters conditional on all other effects
  - The average of these samples can be used as the estimates of the parameters

Bayesian methods

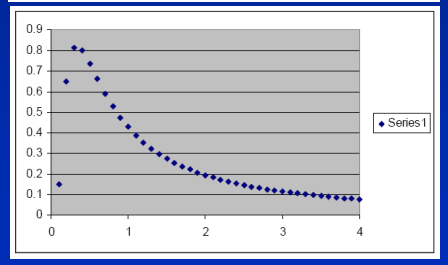
- Gibbs sampling scheme
  - Parameters to estimate and their posteriors
    - $P(\sigma_{gi}^2|g_i)$

$$\chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$$



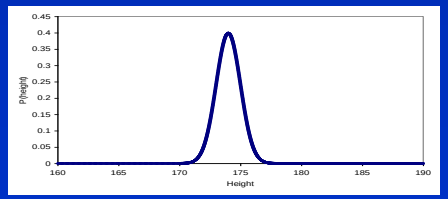
    - $P(\sigma_e^2|\mathbf{e})$

$$\chi^{-2}_{(n-2,\,\mathbf{e}'\mathbf{e})}$$



    - $P(\mu|\mathbf{y},\mathbf{e},\mathbf{g},\ \sigma_e$

$$N\left(\frac{1}{n}\left(\mathbf{1_n}'\mathbf{y}-\mathbf{1_n}'\mathbf{Xg}\right),\sigma_e^2/n\right)$$



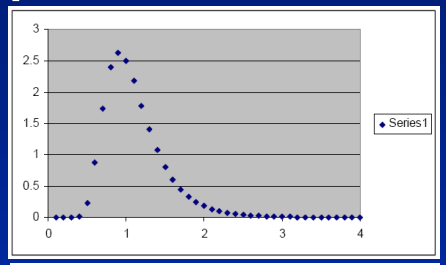
    - $P(g_{ij}|\mathbf{y},\mu,\mathbf{g}\neq ij,\sigma_{gi}^2,\sigma_e^2)$

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

Bayesian methods

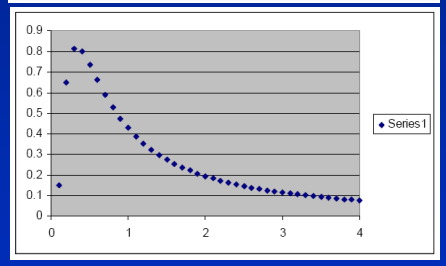
- Gibbs sampling scheme
  - Parameters to estimate and their posteriors
    - $P(\sigma_{gi}^2|g_i)$

$$\chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$$



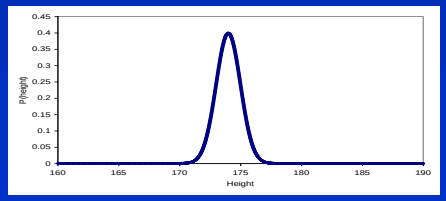
    - $P(\sigma_e^2|\mathbf{e})$

$$\chi^{-2}_{(n-2,\,\mathbf{e}'\mathbf{e})}$$



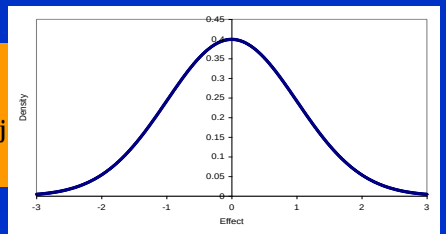
    - $P(\mu|\mathbf{y},\mathbf{e},\mathbf{g},\ \sigma_e$

$$N\left(\frac{1}{n}\left(\mathbf{1_n}'\mathbf{y}-\mathbf{1_n}'\mathbf{Xg}\right),\sigma_e^2/n\right)$$



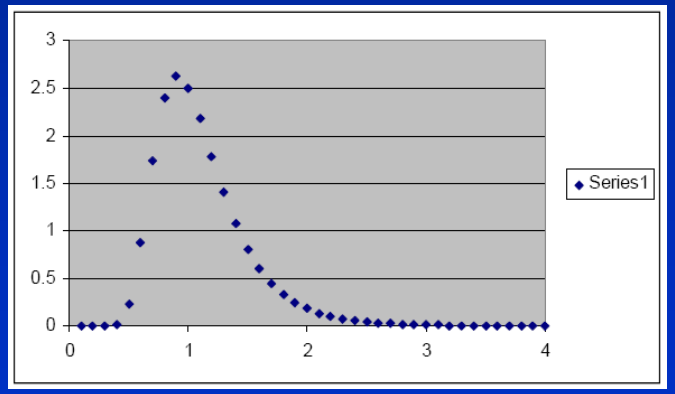
    - $P(g_{ij}|\mathbf{y},\mu,\mathbf{g}\neq ij,\sigma_{gi}^2,\sigma_e^2)$

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



Bayesian methods

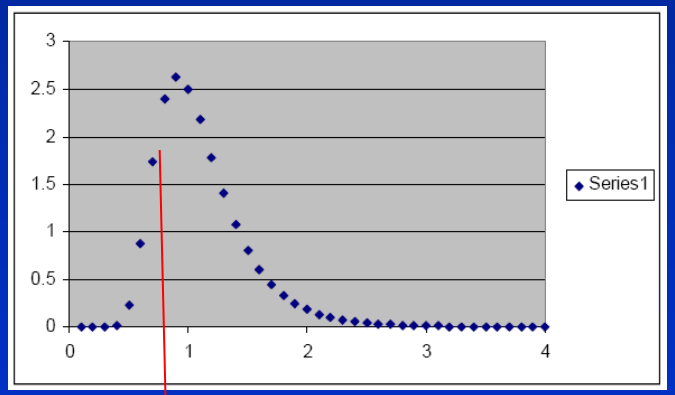
- The Gibbs chain
  - Step 1. Initialise value of **g**, eg. **g**=0.01 and μ, eg μ=0.01
  - Step 2. For each *i*, draw from $P(\sigma_{gi}^2|g_i)$



$$\chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, eg. **g**=0.01 and μ, eg μ=0.01
  - Step 2. For each *i*, draw from $P(\sigma_{gi}^2|g_i)$



$$\chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$$

- σ_{g1}^2=0.95

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, eg. **g**=0.01 and μ, eg μ=0.01
  - Step 2. For each *i*, draw from $P(\sigma_{gi}^2|g_i)$
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$ First calculate the **e** as

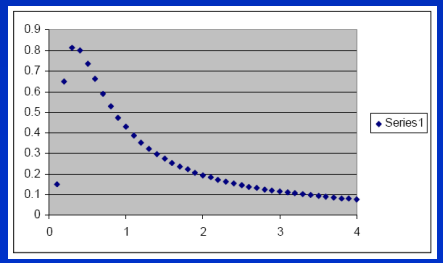
$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{g} - 1_n' \mu$$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, eg. **g**=0.01 and μ, eg μ=0.01
  - Step 2. For each *i*, draw from $P(\sigma_{gi}^2|g_i)$
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$ First calculate the **e** as

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{g} - 1_n' \mu$$

  - Then sample…



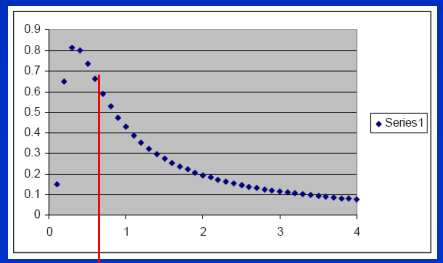
$$\chi^{-2}_{(n-2,\mathbf{e}'\mathbf{e})}$$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, eg. **g**=0.01 and μ, eg μ=0.01
  - Step 2. For each *i*, draw from $P(\sigma_{gi}^2|g_i)$
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$ First calculate the **e** as

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{g} - 1_n \mu$$

  - Then sample…



$$\chi^{-2}_{(n-2,\mathbf{e}'\mathbf{e})}$$

  - $\sigma_e^2 = 0.5$

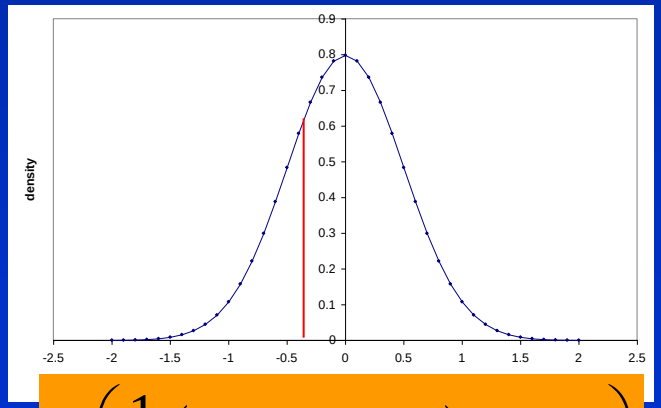
Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, eg. **g**=0.01 and μ, eg μ=0.01
  - Step 2. For each *i*, draw from $P(\sigma_{gi}^2|g_i)$
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$
  - Step 4. Draw a sample from $P(\mu|y,g,\sigma_e^2)$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, eg. **g**=0.01 and μ, eg μ=0.01
  - Step 2. For each *i*, draw from $P(\sigma_{gi}^2|g_i)$
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$
  - Step 4. Draw a sample from $P(\mu|y,g,\sigma_e^2)$

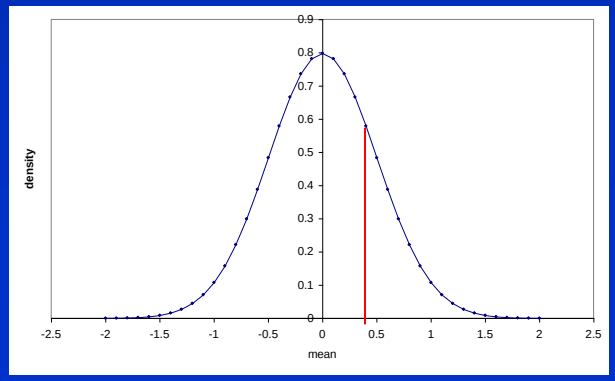
- μ=-0.1



$$N\left(\frac{1}{n}\left(\mathbf{1_n'y}-\mathbf{1_n'Xg}\right),\sigma_e^2/n\right)$$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, eg. **g**=0.01 and μ, eg μ=0.01
  - Step 2. For each *i*, draw from $P(\sigma_{gi}^2|g_i)$
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$
  - Step 4. Draw a sample from $P(\mu|y,g,\sigma_e^2)$
  - Step 5. For each g_{ij}, draw from $P(g_{ij}|y,\mu,g,\sigma_{gi}^2,\sigma_e^2)$
    - g_{11} = 0.5



Bayesian methods

- The Gibbs chain
  - Repeat steps 2-5 many times to build up samples from posterior distributions of the parameters

Bayesian methods

- The Gibbs chain
  - Repeat steps 2-5 many times to build up samples from posterior distributions of the parameters
  - Finally, take estimates of parameters as average over many cycles
  - Discard first ~ 100 cycles as dependent on starting values

Bayesian methods

- Example
  - Consider a data set with three markers. The data set was simulated as:
  - the effect of a 2 allele at the first marker is 3, the effect of a 2 allele at the second marker is 0, and the effect of a 2 allele at the third marker was -2.
  - the μ was 3
  - σ_e^2 was 0.23. The data set was:

Bayesian methods

- Example

Animal	Phenotype	Marker1 allele 1	Marker1 allele 2	Marker2 allele 1	Marker 2 allele 2	Marker3 allele 1	Marker 3 allele 2
1	9.68	2	2	2	1	1	1
2	5.69	2	2	2	2	2	2
3	2.29	1	2	2	2	2	2
4	3.42	1	1	2	1	1	1
5	5.92	2	1	1	1	1	1
6	2.82	2	1	2	1	2	2
7	5.07	2	2	2	1	2	2
8	8.92	2	2	2	2	1	1
9	2.4	1	1	2	2	1	2
10	9.01	2	2	2	2	1	1
11	4.24	1	2	1	2	2	1
12	6.35	2	2	1	1	1	2
13	8.92	2	2	1	2	1	1
14	-0.64	1	1	2	2	2	2
15	5.95	2	1	1	1	1	1
16	6.13	1	2	2	1	1	1
17	6.72	2	1	2	1	1	1
18	4.86	1	2	2	1	1	2
19	6.36	2	2	2	2	2	2
20	0.81	1	1	2	1	1	2
21	9.67	2	2	1	2	1	1
22	7.74	2	2	2	1	1	2
23	1.45	1	1	2	2	2	1
24	1.22	1	1	2	1	2	1
25	-0.52	1	1	2	2	2	2

Bayesian methods

- Example
  - The Bayesian approach was applied, fitting single marker effects
  - X matrix
    - Number of copies of two allele for each animal, eg. 2 1 0 for animal 1.

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, μ
    - g1=0.01, g2=0.01,g3=0.01, μ=0.1

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, μ
    - g1=0.01, g2=0.01,g3=0.01, μ=0.1
  - Step 2. For *i=1,2,3*, draw from $P(\sigma_{gi}^2|g_i)$

$$\chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, μ
    - g1=0.01, g2=0.01,g3=0.01, μ=0.1
  - Step 2. For *i=1,2,3*, draw from $P(\sigma_{gi}^2|g_i)$

$$\chi^{-2}_{(4.012+1,0.002+0.0001)}$$

- σ_{g1}^2=0.002, σ_{g2}^2=0.06, σ_{g3}^2=0.009

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, μ
    - g1=0.01, g2=0.01,g3=0.01, μ=0.1
  - Step 2. For *i=1,2,3*, draw from $P(\sigma_{gi}^2|g_i)$
    - σ_{g1}^2=0.002, σ_{g2}^2=0.06, σ_{g3}^2=0.009
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$

$$\chi^{-2}_{(n-2,\mathbf{e'e})}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{Xg} - 1_n \mu$$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, μ
    - g1=0.01, g2=0.01,g3=0.01, μ=0.1
  - Step 2. For *i=1,2,3*, draw from $P(\sigma_{gi}^2|g_i)$
    - σ_{g1}^2=0.002, σ_{g2}^2=0.06, σ_{g3}^2=0.009
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$

$$\chi^{-2}_{(23,812.031)}$$

- σ_e^2= 53.38

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, μ
    - g1=0.01, g2=0.01,g3=0.01, μ=0.1
  - Step 2. For *i=1,2,3*, draw from $P(\sigma_{gi}^2|g_i)$
    - σ_{g1}^2=0.002, σ_{g2}^2=0.06, σ_{g3}^2=0.009
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$
    - σ_e^2= 53.38
  - Step 4. Draw a sample from $P(\mu|y,g,\sigma_e^2)$
    - μ=3.25

$$N\left(\frac{1}{n}\left(\mathbf{1}_{\mathbf{n}}'\mathbf{y}-\mathbf{1}_{\mathbf{n}}'\mathbf{X}\mathbf{g}\right),\sigma_e^2/n\right)$$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, μ
    - g1=0.01, g2=0.01,g3=0.01, μ=0.1
  - Step 2. For *i=1,2,3*, draw from $P(\sigma_{gi}^2|g_i)$
    - σ_{g1}^2=0.002, σ_{g2}^2=0.06, σ_{g3}^2=0.009
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$
    - σ_e^2= 53.38
  - Step 4. Draw a sample from $P(\mu|y,g,\sigma_e^2)$
    - μ=3.25
  - Step 5. Draw a sample from $P(g_{ij}|y,\mu,\mathbf{g}\neq ij,\sigma_{gi}^2,\sigma_e^2)$

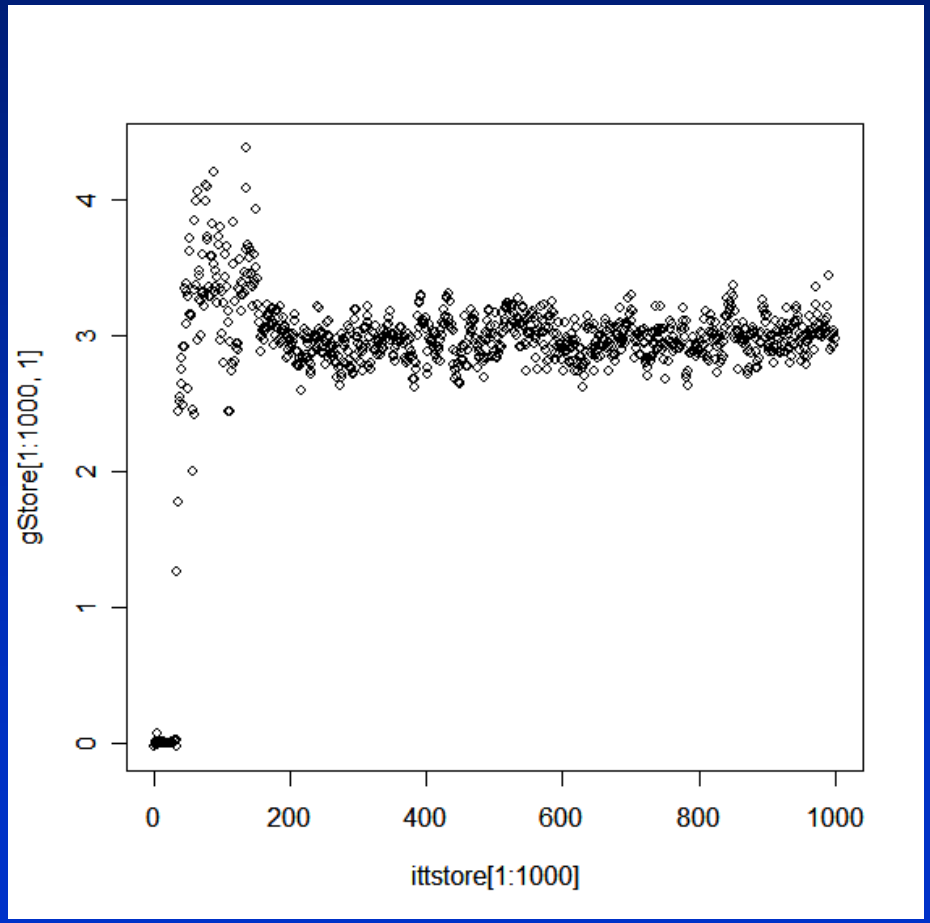
$$N\left(\frac{\mathbf{X}_{ij}'\mathbf{y}-\mathbf{X}_{ij}'\mathbf{X}\mathbf{g}_{(ij=0)}-\mathbf{X}_{ij}'\mathbf{1}_n\mu}{\mathbf{X}_{ij}'\mathbf{X}_{ij}+\sigma_e^2/\sigma_{gi}^2},\sigma_e^2/\left(\mathbf{X}_{ij}'\mathbf{X}_{ij}+\sigma_e^2/\sigma_{gi}^2\right)\right)$$

Bayesian methods

- The Gibbs chain
  - Step 1. Initialise value of **g**, μ
    - g1=0.01, g2=0.01,g3=0.01, μ=0.1
  - Step 2. For *i=1,2,3*, draw from $P(\sigma_{gi}^2|g_i)$
    - σ_{g1}^2=0.002, σ_{g2}^2=0.06, σ_{g3}^2=0.009
  - Step 3. Draw a sample from $P(\sigma_e^2|e)$
    - σ_e^2= 53.38
  - Step 4. Draw a sample from $P(\mu|y,g, \sigma_e^2,e)$
    - μ=3.25
  - Step 5. Draw a sample from $P(g_{ij}|y,\mu,\mathbf{g}\neq ij,\sigma_{gi}^2,\sigma_e^2)$
    - g1=-0.02, g2=-0.81,g3=-0.005

Bayesian methods

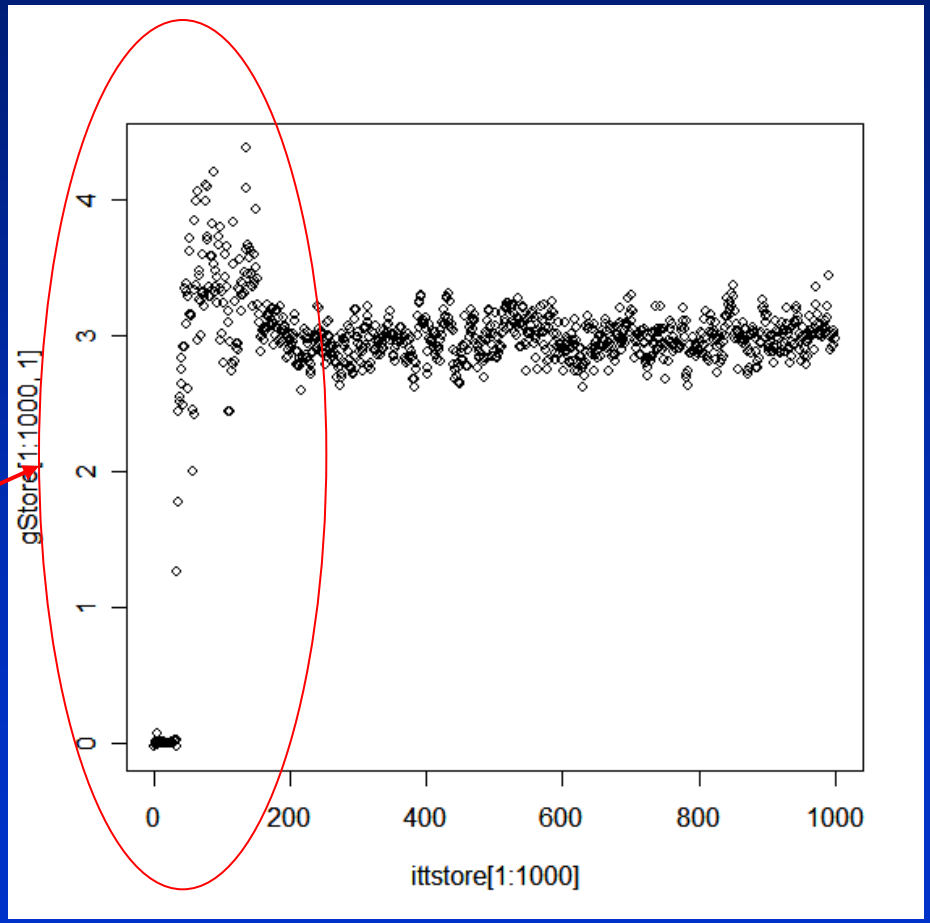
- Gibbs chain for 1000 cycles
  - $P(g_1|y,\mu,\mathbf{g}\neq 1,\sigma_{g1}^2,\sigma_e^2)$



Bayesian methods

- Gibbs chain for 1000 cycles
  - $P(g_1|y,\mu,\mathbf{g}\neq 1,\sigma_{g1}^2,\sigma_e^2)$

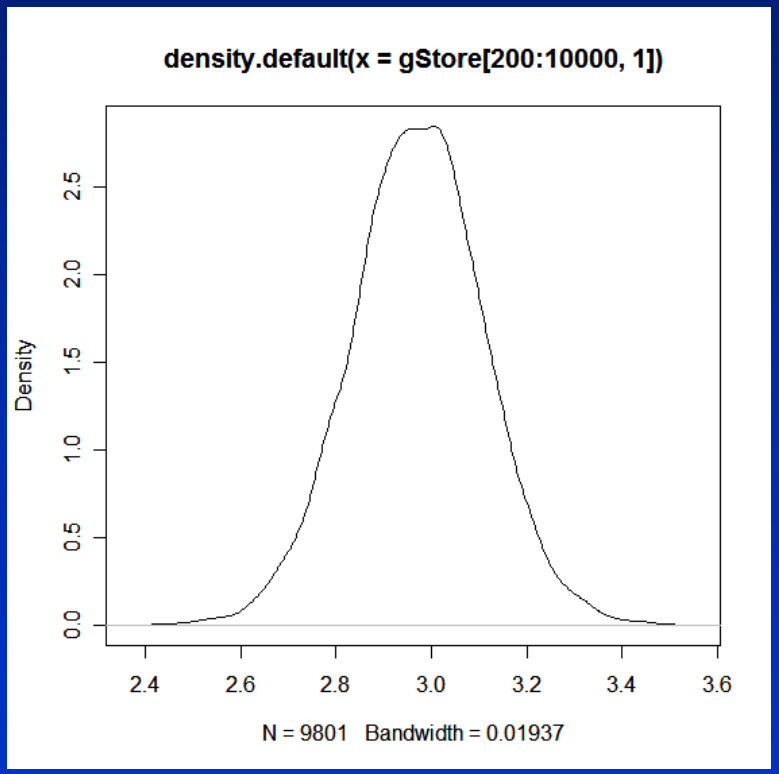
"Burn in"



Bayesian methods

- Gibbs chain for 1000 cycles
  - $P(g_1|y,\mu,\mathbf{g}\neq 1,\sigma_{g1}^2,\sigma_e^2)$

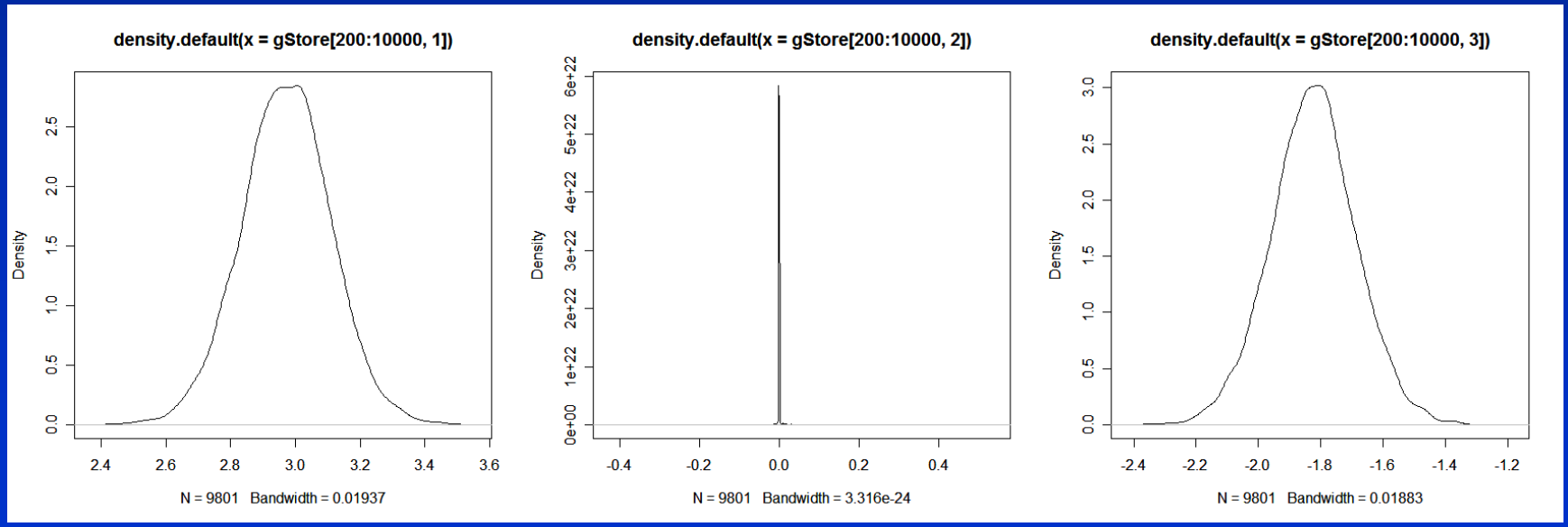
$$\hat{g}_1 = 2.97$$



Bayesian methods

- Gibbs chain for 1000 cycles

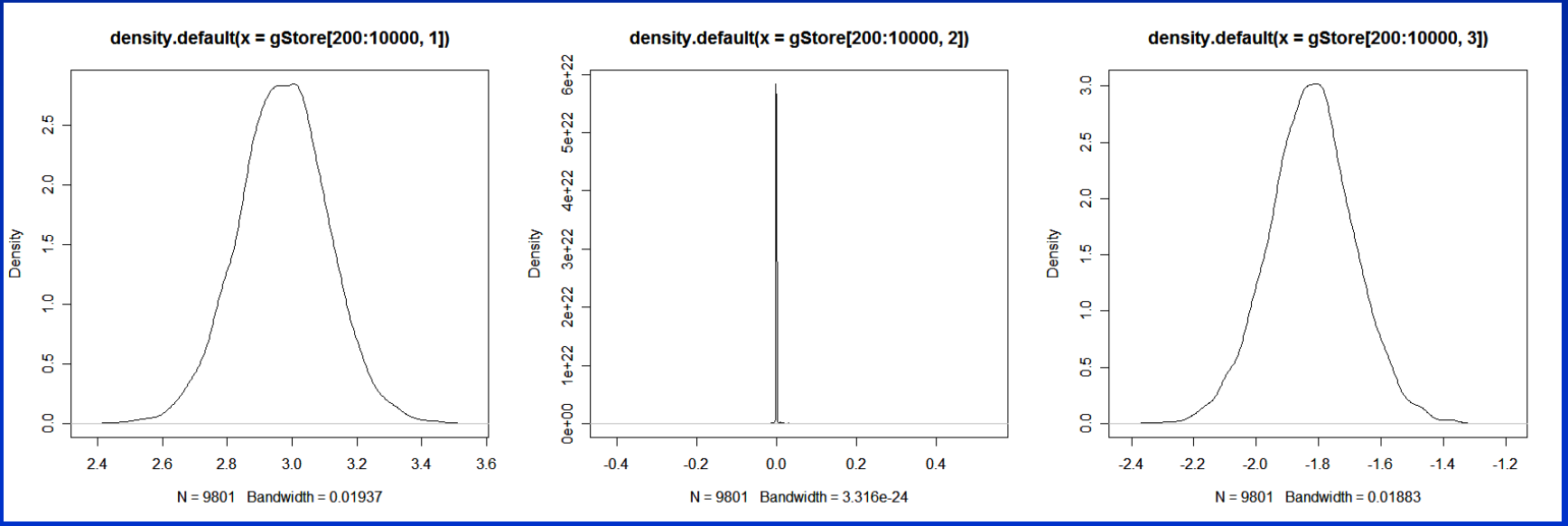
$\hat{g}_1 = 2.97$ $\hat{g}_2 = 0.002$ $\hat{g}_1 = -1.81$



Bayesian methods

Vector of SNP effects for calculating GEBV

$$\hat{g}_1 = 2.97 \qquad \hat{g}_2 = 0.002 \qquad \hat{g}_1 = -1.81$$



Bayesian methods

- Alternative priors for variance of segment haplotype/snp effects
  - Meuwissen BayesA

$$\chi^{-2}_{(4.012,0.002)} \qquad \chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$$

  - Xu (2003)
    - Uninformative

$$\chi^{-2}_{(0,0)} \qquad \chi^{-2}_{(1,\mathbf{g}'\mathbf{g})}$$

  - Ter Braak (2006)

$$p(\sigma^2_{gi}) \propto (\sigma^2_{gi})^{-1+\alpha} \qquad g_i' g_i / \chi^{-2}_{1-2a}$$

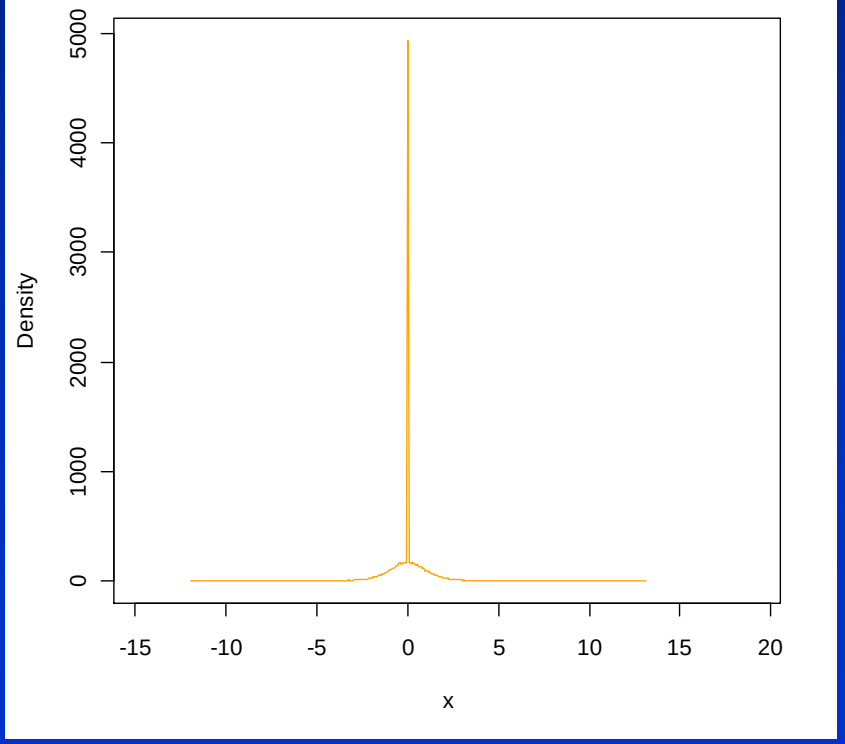
  - Meuwissen BayesB

$$\begin{aligned} \sigma^2_{gi} &= 0 \text{ with probability } \pi, \\ \sigma^2_{gi} &\sim \chi^{-2}(\nu,\ S) \text{ with probability } (1-\pi), \end{aligned}$$

Bayesian methods

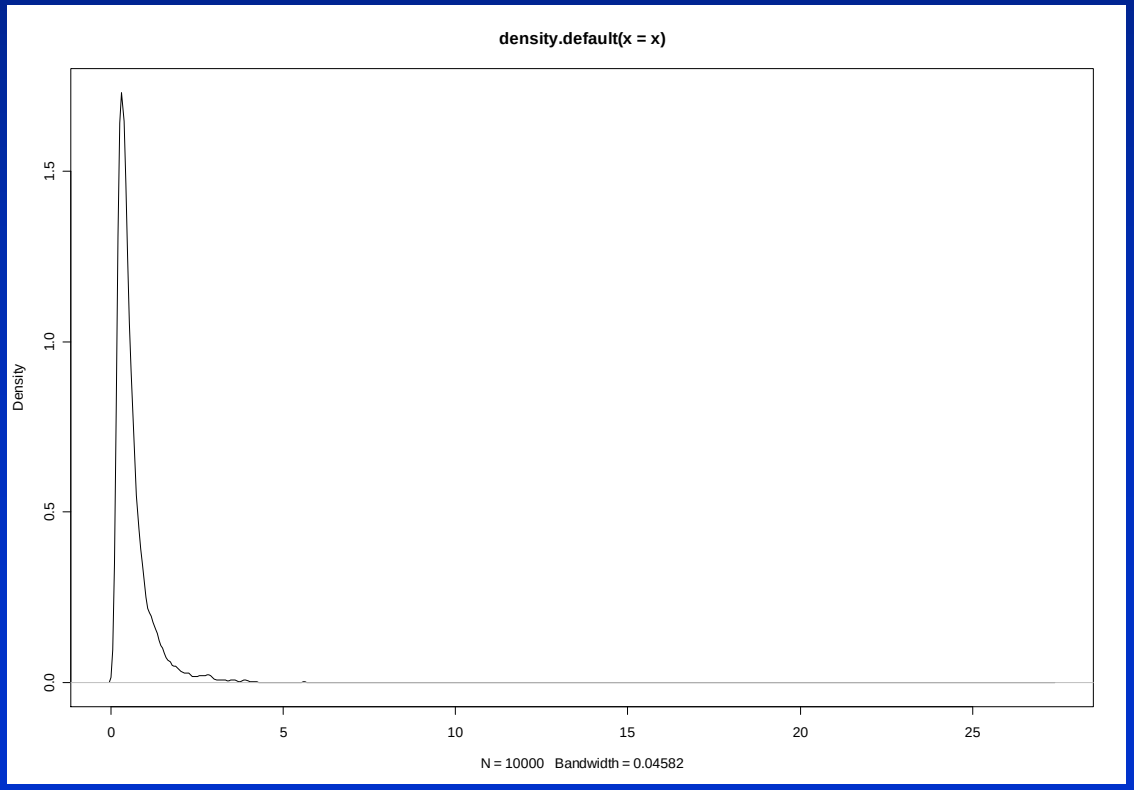
- Meuwissen BayesB
  - BayesA prior information is many QTL with small effects and few with moderate effects
  - But we have more prior knowledge than this – some chromosome segments will have no effect at all (*contain no QTL*)
    - $\sigma_{gi}^2=0, g_i=0$
  - How to sample from the posterior?

$$\sigma_{gi}^2 = 0 \text{ with probability } \pi,$$
$$\sigma_{gi}^2 \sim \chi^{-2}(\nu, S) \text{ with probability } (1-\pi),$$



Bayesian methods

- Meuwissen BayesB
  - If we sample σ_{gi}^2 from $\chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$
  - We will never sample 0, as the distribution has no mass at zero.



Bayesian methods

- Meuwissen BayesB
  - If we sample σ_{gi}^2 from $\chi^{-2}_{(4.012+n_i,\,0.002+\mathbf{g_i}'\mathbf{g_i})}$
  - We will never sample 0 if $g_i'g_i>0$, as the distribution has no mass at zero.
  - But if $\sigma_{gi}^2 >0$, then sampling $g_i = 0$ has infinitesimal (basically zero) probability

Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

We want to sample from this

Can do it by sampling from these two distributions

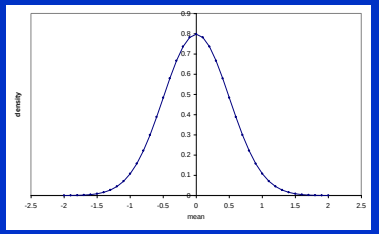
Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

We want to sample from this

$P(g_i|y,\mu,g,\sigma_{gi}^2,\sigma_e^2)$



Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

??

Sample σ_{gi}^2 without conditioning on g_i

Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

  - Cannot be expressed as a known distribution = cannot use Gibbs for this bit
  - Use a Metropolis Hastings algorithm

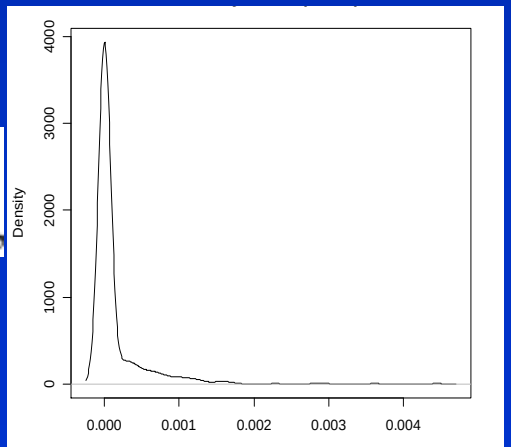
Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

  - Step1 Sample $\sigma_{g\_new}^2$,from prior($\sigma_{g\_new}^2$)

$$\sigma_{gi}^2 = 0 \text{ with probability } \pi,$$
$$\sigma_{gi}^2 \sim \chi^{-2}(\nu, S) \text{ with probability } (1-\pi),$$

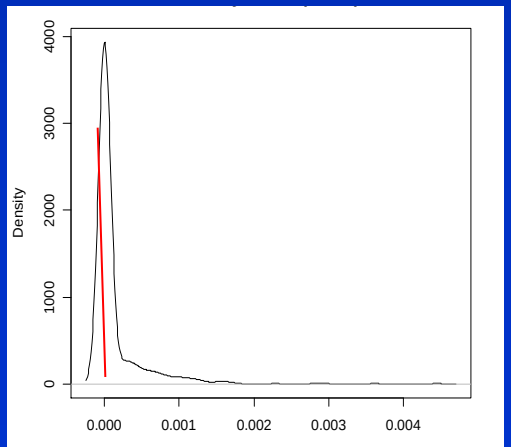


Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

  - Step1 Sample $\sigma_{g\_new}^2$,from prior($\sigma_{g\_new}^2$)
    - $\sigma_{g\_new}^2 = 0$

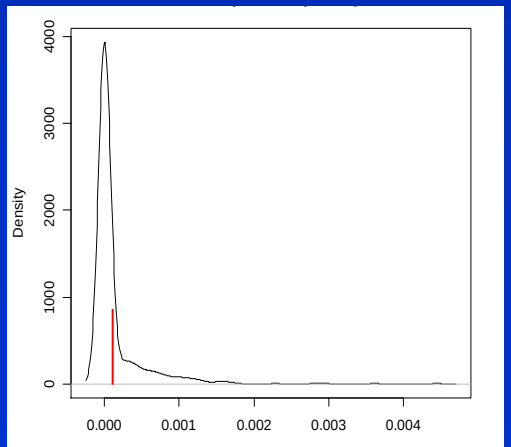


Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

  - Step1 Sample $\sigma_{g\_new}^2$,from prior($\sigma_{g\_new}^2$)
    - $\sigma_{g\_new}^2 = 0.5$



Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

  - Step 1 Sample $\sigma_{g\_new}^2$,from prior($\sigma_{g\_new}^2$)
  - Step 2 Evaluate p(y*| $\sigma_{g\_new}^2$) (Likelihood)

$$L(\mathbf{y^*} \mid \sigma_{ginew}^2 = \frac{1}{2\pi^{1/2n} \mid \mathbf{V} \mid^{1/2}} e(-0.5 * (\mathbf{y^{*'} V^{-1} y^*}))$$

$$\mathbf{V} = \mathbf{X}(\mathbf{I}\sigma_{ignew}^2)\mathbf{X'} + \mathbf{I}\boldsymbol{\sigma}_{\mathbf{e}}^{\mathbf{2}})$$

Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

  - Step 1 Sample $\sigma_{g\_new}^2$,from prior($\sigma_{g\_new}^2$)
  - Step 2 Evaluate p(y*| $\sigma_{g\_new}^2$) (Likelihood)
  - Step 3 Replace σ_{gi}^2 with $\sigma_{g\_new}^2$ probability min[p(y*| $\sigma_{g\_new}^2$)/ p(y*| σ_{gi}^2):1]

Bayesian methods

- Meuwissen BayesB
  - Solution: sample σ_{gi}^2, g_i simultaneously from the distribution:

$$p(\sigma_{gi}^2, g_i \mid y^*) = p(\sigma_{gi}^2 \mid y^*) \times p(g_i \mid \sigma_{gi}^2, y^*)$$

  - Step 1 Sample $\sigma_{g\_new}^2$, from prior($\sigma_{g\_new}^2$)
  - Step 2 Evaluate p(y*| $\sigma_{g\_new}^2$) (Likelihood)
  - Step 3 Replace σ_{gi}^2 with $\sigma_{g\_new}^2$ probability min[p(y*| $\sigma_{g\_new}^2$)/ p(y*| σ_{gi}^2):1]
  - Step 4 Repeat ~ 100 cycles

Genomic prediction

- Comparison of accuracy of methods (Meuwissen et al. 2001)
  - Genome of 1000 cM simulated, marker spacing of 1 cM.
  - Markers surrounding each 1-cM region combined into haplotypes.
  - Due to finite population size (Ne = 100), marker haplotypes were in linkage disequilibrium with QTL between markers.
  - Effects of haplotypes predicted in one generation of 2000 animals
  - Breeding values for progeny of these animals predicted based on marker genotypes

Genomic prediction

- Comparison of accuracy of methods (Meuwissen et al. 2001)

	$r_{TBV;EBV}$ + SE	$b_{TBV.EBV}$ + SE
LS	0.318 ± 0.018	0.285 ± 0.024
BLUP	0.732 ± 0.030	0.896 ± 0.045
BayesA	0.798	0.827
BayesB	0.848 + 0.012	0.946 + 0.018

Genomic prediction

- Comparison of accuracy of methods (Meuwissen et al. 2001)
  - The least squares method does very poorly, primarily because the haplotype effects are over-estimated.

Genomic prediction

- Comparison of accuracy of methods (Meuwissen et al. 2001)
  - The least squares method does very poorly, primarily because the haplotype effects are over-estimated.
  - Increased accuracy of the Bayesian approach because method sets many of the effects of the chromosome segments close to zero in BayesA, or zero in BayesB

Genomic prediction

- Comparison of accuracy of methods (Meuwissen et al. 2001)
  - The least squares method does very poorly, primarily because the haplotype effects are over-estimated.
  - Increased accuracy of the Bayesian approach because method sets many of the effects of the chromosome segments close to zero in BayesA, or zero in BayesB
  - Also “shrinks” estimates of effects of other chromosome segments based on a prior distribution of QTL effects.

Genomic prediction

- Comparison of accuracy of methods (Meuwissen et al. 2001)
  - The least squares method does very poorly, primarily because the haplotype effects are over-estimated.
  - Increased accuracy of the Bayesian approach because method sets many of the effects of the chromosome segments close to zero in BayesA, or zero in BayesB
  - Also "shrinks" estimates of effects of other chromosome segments based on a prior distribution of QTL effects.
  - Accuracies were very high, as high as following progeny testing for example

Genomic prediction

- Introduction to genomic selection
- Genomic prediction with BLUP
- Genomic prediction with Bayesian methods
- Examples in real data

In real data

- 1500 Australian dairy bulls
- genotyped for 56000 genome wide SNPs
- Phenotypes average of daughters milk production

In real data

- Split data into two sub-populations
  - Reference: Bulls born < 2003
  - Validation: Bulls born >= 2003

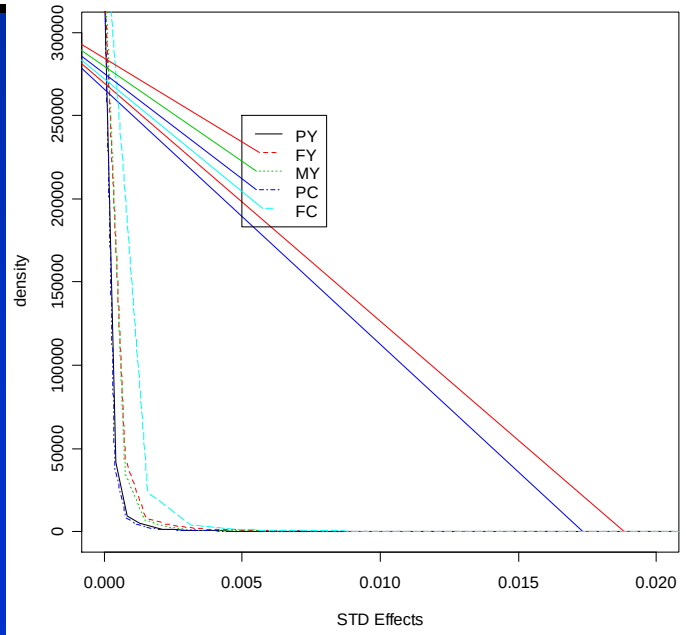
In real data

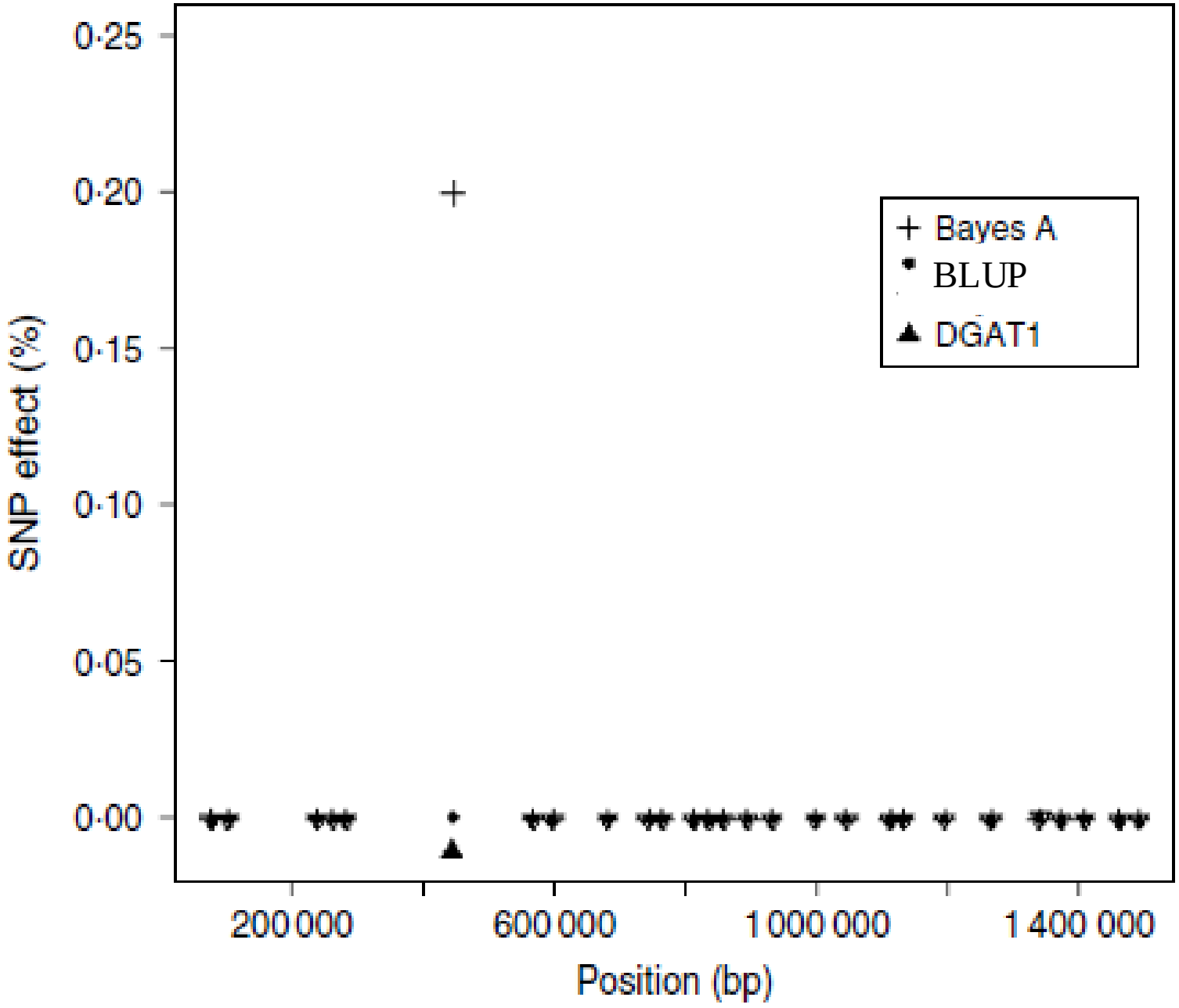
- Split data into two sub-populations
  - Reference: Bulls born < 2003
  - Validation: Bulls born >= 2003
- Accuracy
  - Correlation of genomic breeding values with EBVs (which include daughter information) in validation set

In real data

Table 3 MEBV- Correlation between predicted MEBV and ABV in the validation data set (Bulls proven in years 2005, 2006, 2007)

Method	*Protein kg*	*Fat kg*	*Protein %*	*Fat %*
Bayes B	0.55	0.51	0.68	0.73
Bayes A	0.53	0.48	0.66	0.70
BLUP	0.60	0.48	0.66	0.64




0·25
0·20
0·15
0·10
0·05
0·00
SNP effect (%)
200 000
600 000
1 000 000
1 400 000
Position (bp)
Bayes A
BLUP
DGAT1

Genomic prediction

- Bayesian Cπ (Habier et al 2011)

- Two criticisms of BayesB
  - Posterior of locus-specific variance has only one additional degree of freedom, compared to its prior regardless of the number of genotypes, so
    - Degree of shrinkage of depends strongly on prior
    - Little information coming from data
  - π is treated as known, not estimated from the data

Genomic prediction

- Bayesian Cπ (Habier et al 2011)

- Use a common σ_{gi}^2 across all SNP
  - Many degrees of freedom from data
  - A "BLUP" for SNP in model
- Estimate π from data
  - Sample from
    - Beta(K - m(t) + 1, m(t) + 1).
    - Where K is number of SNP, m(t) is the number of SNP in the model at iteration t (eg. Those not set to zero)

Genomic prediction

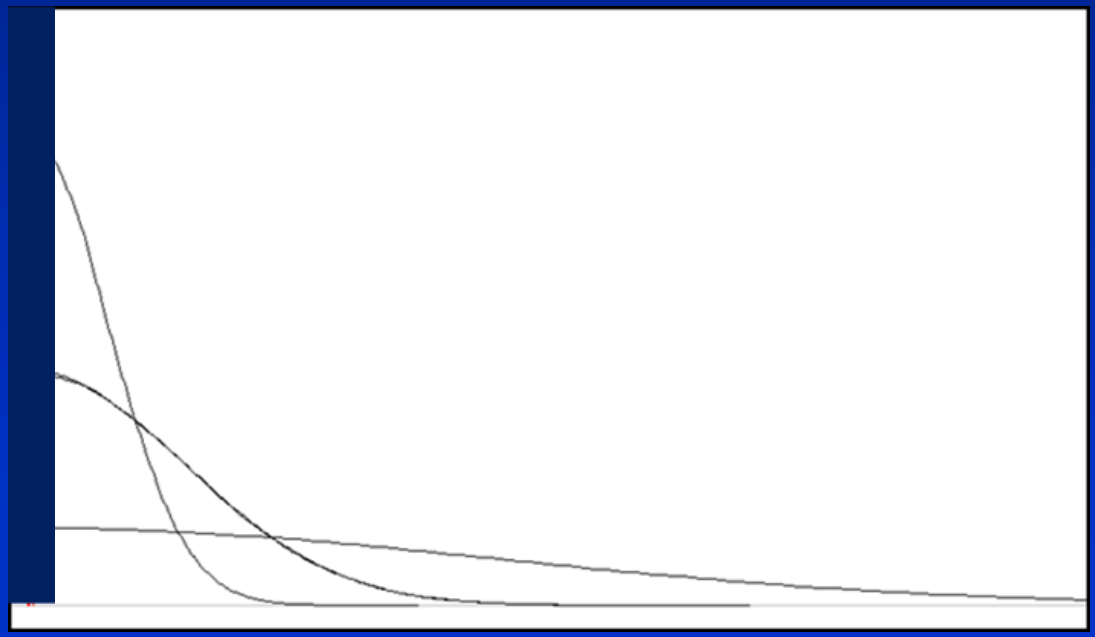
- Bayesian Cπ (Habier et al 2011)
  - Accuracy in German Holstein Friesian data set

Trait	GBLUP	BayesA	BayesB	BayesCpi
Milk Yield	0.48	0.48	0.40	0.43
Fat Yield	0.51	0.56	0.52	0.54
Protein Yield	0.21	0.22	0.17	0.21
Somatic cells	0.17	0.17	0.12	0.14

- Can draw inferences about trait architecture?

BayesR

- BayesR -> variants belong to one of 4 normal distributions, with zero, very small, small, medium variance
- Posterior proportion of variants in each distribution



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**

Genomic prediction

- Methods for deriving prediction equation differ in assumptions about distribution of QTL effects
  - BLUP = normal distribution with known variance
  - Ridge regression = normal distribution with prior assumption about variance
  - BayesA = t-distribution, degree of shrinkage known a-priori, or sampled
  - BayesB = mixture distribution, many effects zero
  - BayesianLASSO, double exponential distribution of effects
  - Bayesian Cπ, estimate π from data, common variance across SNP
  - BayesR = multiple normal distributions

Genomic prediction

- Bayesian methods can have an advantage when:
- QTL of moderate to large effect on the trait (eg Fat%, DGAT1)
- Very large numbers of SNP (eg 800K) (but need large reference sets) – set some SNP effects to zero
- Multi-breed, across breed genomic predictions