Bayesian whole-genome regression (Bayesian alphabet) for GWAS

Malachy Campbell

6/25/2019

Overview

- Intro to whole genome regression
  - Dealing with $n << p$
- Bayesian flavors
  - Bayesian Ridge Regression
  - Bayesian LASSO (Park and Casella 2008)
  - Bayes A (Meuweissen et al 2001)
  - Bayes B (Meuweissen et al 2001)
  - Bayes C and Bayes Cπ (Habier et al 2011)

Quantitative genetics review

$$\mathbf{y} = \mathbf{g} + \mathbf{e}$$

Quantitative genetics review

$$\mathbf{y} = \mathbf{g} + \mathbf{e}$$

- What is $\mathbf{g}$?

Quantitative genetics review

$$\mathbf{y} = \mathbf{g} + \mathbf{e}$$

$$\mathbf{y} = \sum_{j=1}^{n} a_j w_j + e$$

- $\mathbf{g}$ is the summation of QTL effects across all QTL for each individual

Quantitative genetics review

$$\mathbf{y} = \mathbf{g} + \mathbf{e}$$

$$\mathbf{y} = \sum_{j=1}^{n} a_j w_j + e$$

- $\mathbf{g}$ is the summation of QTL effects across all QTL for each individual

Quantitative genetics review

- With GWAS we are interested in **estimating** a_j
- With genomic prediction we are interested in **predicting g**

Quantitative genetics review

- With GWAS we are interested in **estimating** a_j
- With genomic prediction we are interested in **predicting g**
- **Why not use a model that does both?**

Whole genome regression

$$\mathbf{y} = \mathbf{g} + \mathbf{e}$$

$$\mathbf{y} = \sum_{j=1}^{n} a_j w_j + e$$

- Goal is to fit all markers and obtain all a's simultaneously

Ordinary Least Squares (OLS)

$$\mathbf{y} = \sum_{j=1}^{n} a_j w_j + e$$

- Objective function for OLS

$$\hat{a} = \operatorname{argmin} \left\{ \sum_i (y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2 \right\}$$

Ordinary Least Squares (OLS)

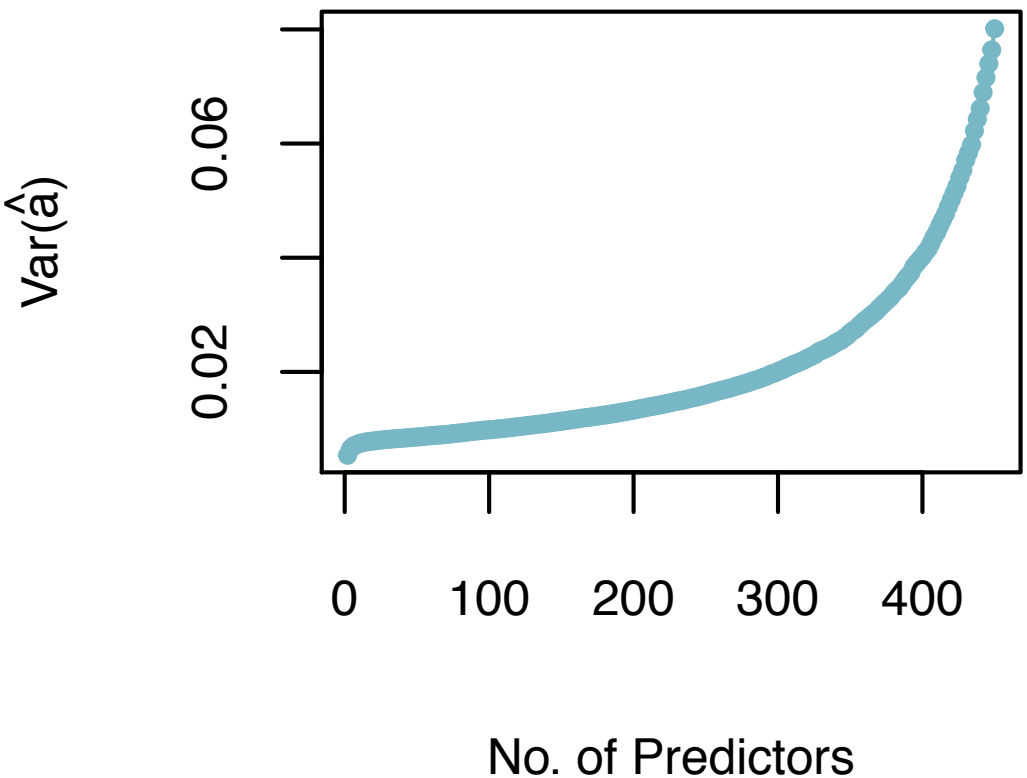
$$\mathbf{y} = \sum_{j=1}^{n} a_j w_j + e$$

- Objective function for OLS

$$\hat{a} = \operatorname{argmin} \left\{ \underbrace{\sum_i (y_i - \mu - \sum_{j=1}^{p} w_{ij} a_j)^2}_{\text{residual sum of squares (RSS)}} \right\}$$

OLS for whole genome regression

- Population size (n) = 500
- Number of markers (p) = 2 - 450
- $\text{Var}(\hat{a}) = [\mathbf{W}'\mathbf{W}]^{-1}\sigma_e^2$
  - $\mathbf{W}$: $n \times p$
  - $\sigma_e^2 = 1$



OLS for whole genome regression

$$\hat{\mathbf{a}} = [\mathbf{W}'\mathbf{W}]^{-1}\mathbf{W}'\mathbf{y}$$

- If $p >> n$

```
p <- 2000 #no of markers
w <- rbinom(prob = 0.5, n = p, size = 1) # simulate genotypes
W <- matrix(nrow = 500, ncol = p, data = w)

WW <- t(W) %*% W
try(solve(WW))
```

```
## Error in solve.default(WW) :
##   Lapack routine dgesv: system is exactly singular: U[5,5] = 0
```

```
is.singular.matrix(WW)
```

```
## [1] TRUE
```

Dealing with $p >> n$

- **Problem:** Variance increases with p; $\mathbf{W}'\mathbf{W}$ is singular
- **Solution**: Treat markers as random; shrink estimates towards 0
  - Penalization methods
  - Bayesian methods

Penalized methods

- **Goal is to balance goodness-of-fit and model complexity**
  - Goodness-of-fit: minimize the residual sum of squares

$$\sum_i (y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2$$

  - Model complexity: function of model unknowns

$$J(a)$$

Penalized methods

- How do we obtain estimates for marker effects?
- Solve the optimization function

$$(\hat{\mu}, \hat{a}) = \text{argmin} \left\{ \sum_i (y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2 + \lambda J(a) \right\}$$

- λ is regularization parameter; $J(a)$ is penalty function

- Minimizing the **penalized** sum of squares

Ridge regression BLUP

$$(\hat{\mu}, \hat{a}) = \operatorname{argmin} \left\{ \sum_i (y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2 + \frac{\sigma_e^2}{\sigma_a^2} \sum_{j=1}^{p} a_j^2 \right\}$$

- Penalty function: $J(a) = \sum_{j=1}^{p} a_j^2$
  - L2 Norm
- Regularization parameter: $\lambda = \frac{\sigma_e^2}{\sigma_a^2}$ -$\mathrm{Var}(a) = \sigma_a^2 \mathbf{I}$

Ridge regression BLUP

$$(\hat{\mu}, \hat{a}) = \operatorname{argmin} \left\{ \sum_i (y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2 + \frac{\sigma_e^2}{\sigma_a^2} \sum_{j=1}^{p} a_j^2 \right\}$$

$$\hat{a} = [\mathbf{W}'\mathbf{W} + \lambda \mathbf{I}]^{-1} \mathbf{W}'\mathbf{y}$$

- Shrinks $\hat{a}$ towards zero (increases bias, but reduces variance)

Ridge regression BLUP: Summary

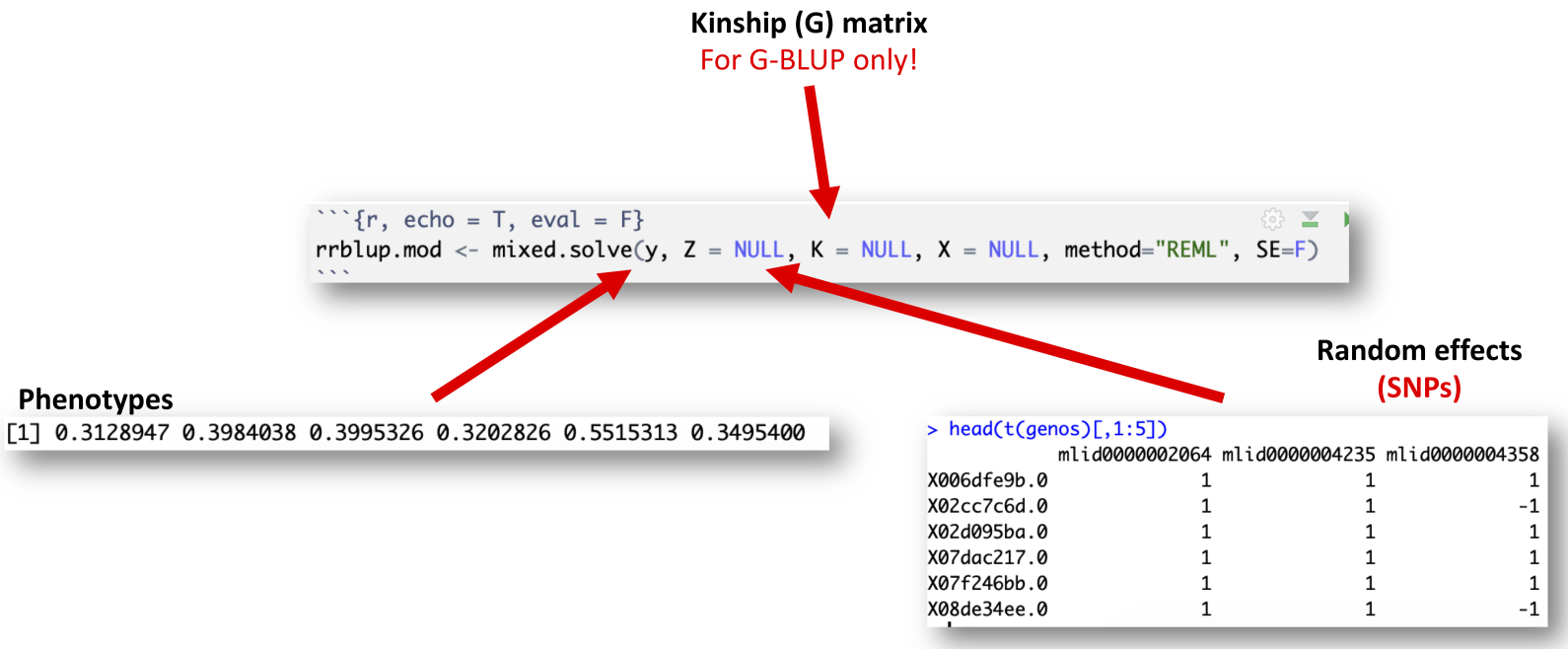
$$y = \mu + \mathbf{Wa} + e$$

$$a \sim \mathrm{N}(0, \mathbf{I}\sigma_a^2)\ ; e \sim \mathrm{N}(0, \mathbf{I}\sigma_e^2)$$

$$\hat{a} = [\mathbf{W}'\mathbf{W} + \lambda\mathbf{I}]^{-1}\mathbf{W}'\mathbf{y}$$

- **All markers** have a small, non-zero contribution to the phenotype (infinitesimal model)

Ridge regression using rrBLUP



Thinking like a Bayesian

- Frequentist: "What is the best estimate for model unknowns given the data?"
- Bayesian: "What is the posterior density of the model unknowns given the data and hyperparameters?"

$$p(\mu, \boldsymbol{a}, \sigma^2 | \boldsymbol{y}, \boldsymbol{\omega})$$

- This is proportional to the conditional density of the data given the unknowns and the joint prior density of model unknowns

$$p(\mu, \boldsymbol{a}, \sigma^2 | \boldsymbol{y}, \boldsymbol{\omega}) \propto p(\mathbf{y} | \mu, \boldsymbol{a}, \sigma^2)\ p(\mu, \boldsymbol{a}, \sigma^2 | \boldsymbol{\omega})$$

Bayesian Ridge regression

- Optimization function for RR:

$$(\hat{\mu}, \hat{a}) = \operatorname{argmin} \left\{ \sum_i (y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2 + \frac{\sigma_e^2}{\sigma_a^2} \sum_{j=1}^{p} a_j^2 \right\}$$

- After some fancy transformations:

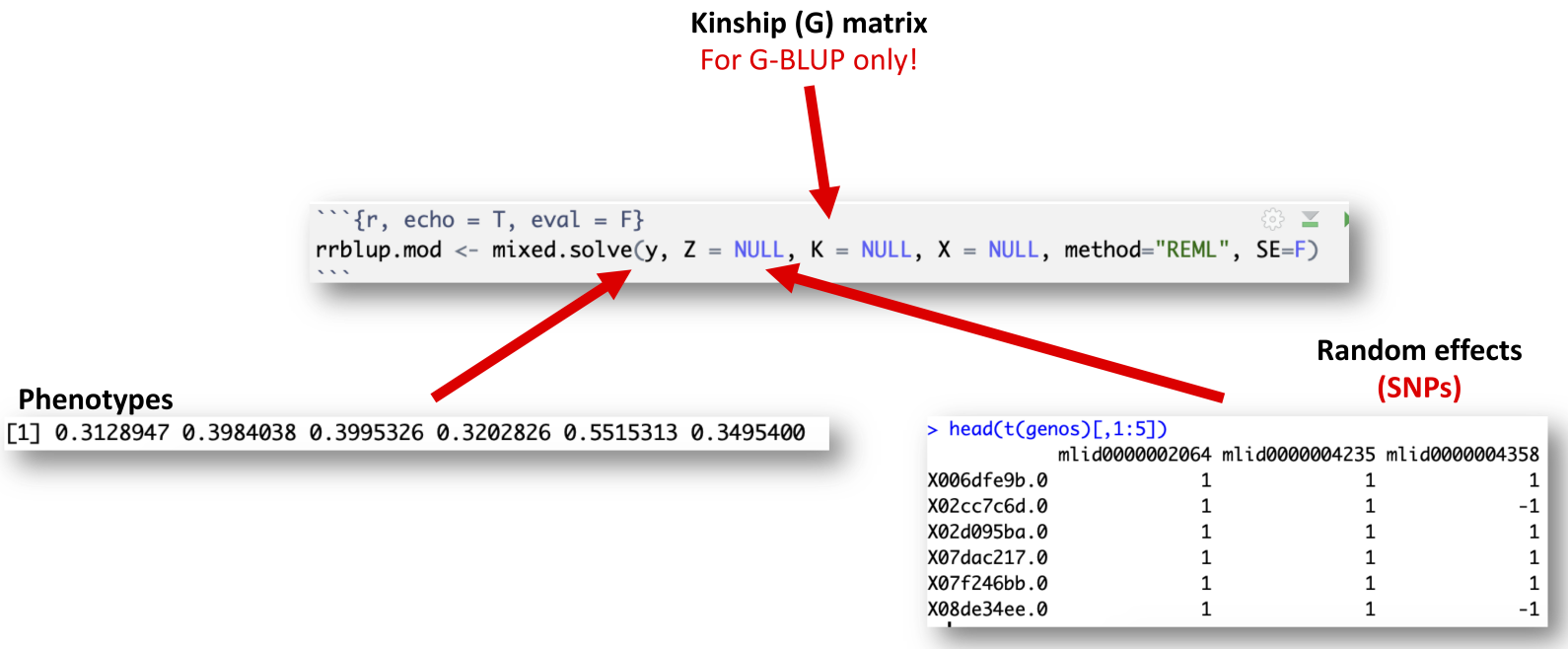
$$(\hat{\mu}, \hat{a}) = \operatorname{argmin} \left\{ \prod_i^n \exp\left(\frac{(y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2}{2\sigma^2} \right) \right\} \left\{ \prod_{j=1}^{p} \exp\left(-\frac{a_j^2}{2\sigma^2} \right) \right\}$$

Bayesian Ridge Regression (BRR)

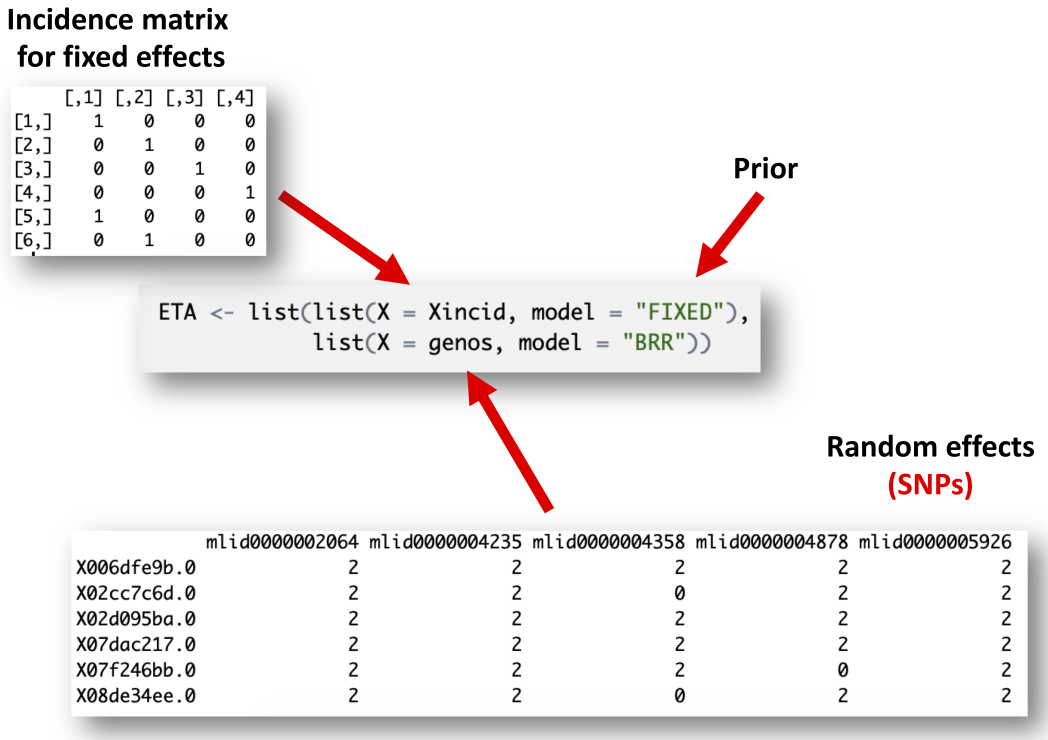
$$(\hat{\mu}, \hat{a}) = \operatorname{argmin} \left\{ \prod_{i}^{n} \exp\left(\frac{(y_i - \mu - \sum_{j=1}^{p} a_j x_{ij})^2}{2\sigma^2} \right) \right\} \left\{ \prod_{j=1}^{p} \exp\left(-\frac{a_j^2}{2\sigma^2} \right) \right\}$$

- Probability density function for normal distribution
  $p(x|a,b) = \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{(x-a)^2}{2b} \right)$
- The mode of the posterior of the Bayesian model is equivalent to the RR solution
  - Bayesian Ridge Regression = Bayesian WGR with Gaussian prior

BRR using the BGLR package



BRR using the BGLR package



Straying from the infinitesimal genetic architecture

- Will all traits follow an infinitesimal model?
  - infinitesimal model = QTL effects are normally distributed
- How can we modify the Bayesian approach to fit different genetic architectures?

Straying from the infinitesimal genetic architecture

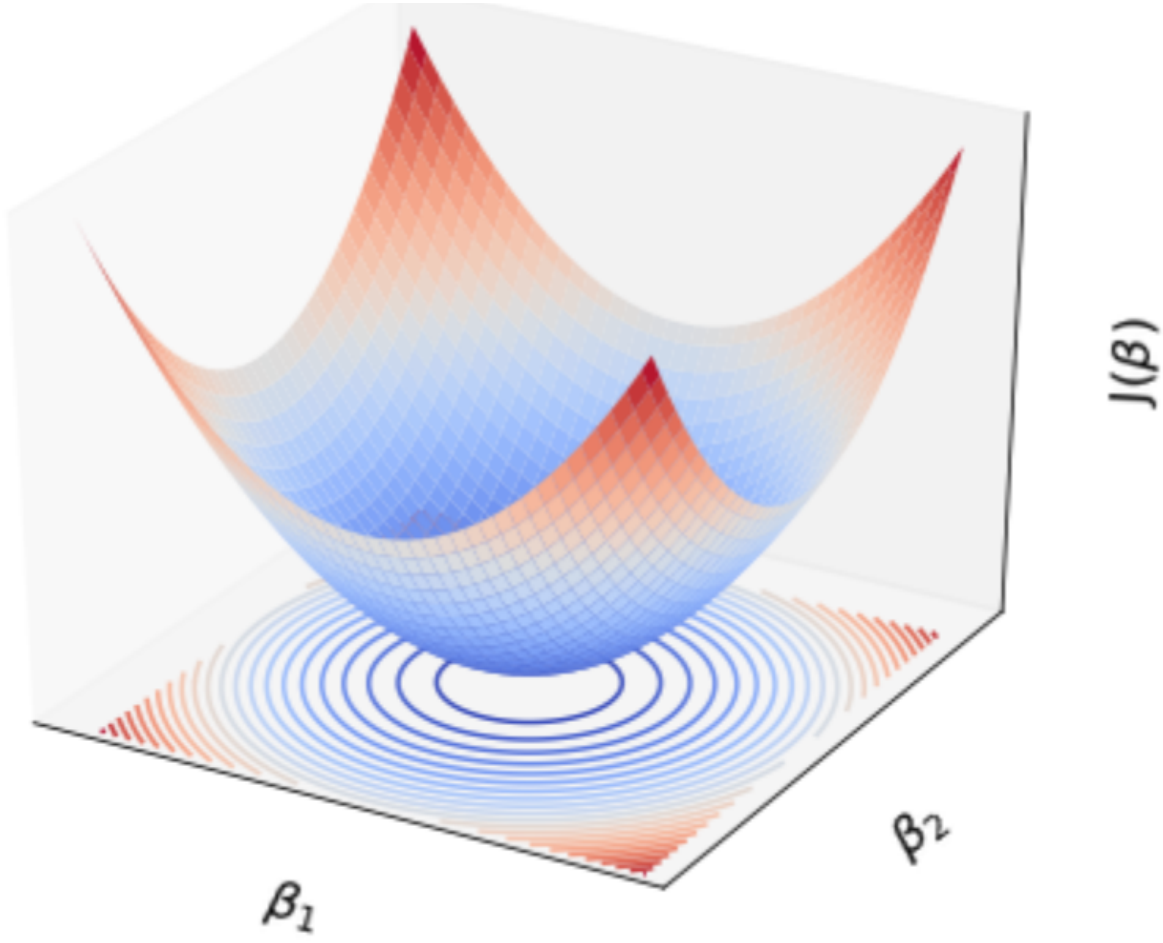
- Will all traits follow an infinitesimal model? probably not
  - infinitesimal model = QTL effects are normally distributed
- How can we modify the Bayesian approach to fit different genetic architectures? choose a different prior

LASSO

- **L**east **A**bsolute **S**hrinkage and **S**election **O**perator
  - **Variable selection**: Some marker effects will be 0, others non-0
  - Like RR, LASSO can be interpreted from both a frequentist and Bayesian perspective
- Frequentist LASSO (L1 norm)

$$(\hat{\mu}, \hat{a}) = \text{argmin} \left\{ \sum_i (y_i - \mu - \sum_{j=1}^{p} w_{ij} a_j)^2 + \frac{\sigma_e^2}{\sigma_a^2} \sum_{j=1}^{p} |a_j| \right\}$$

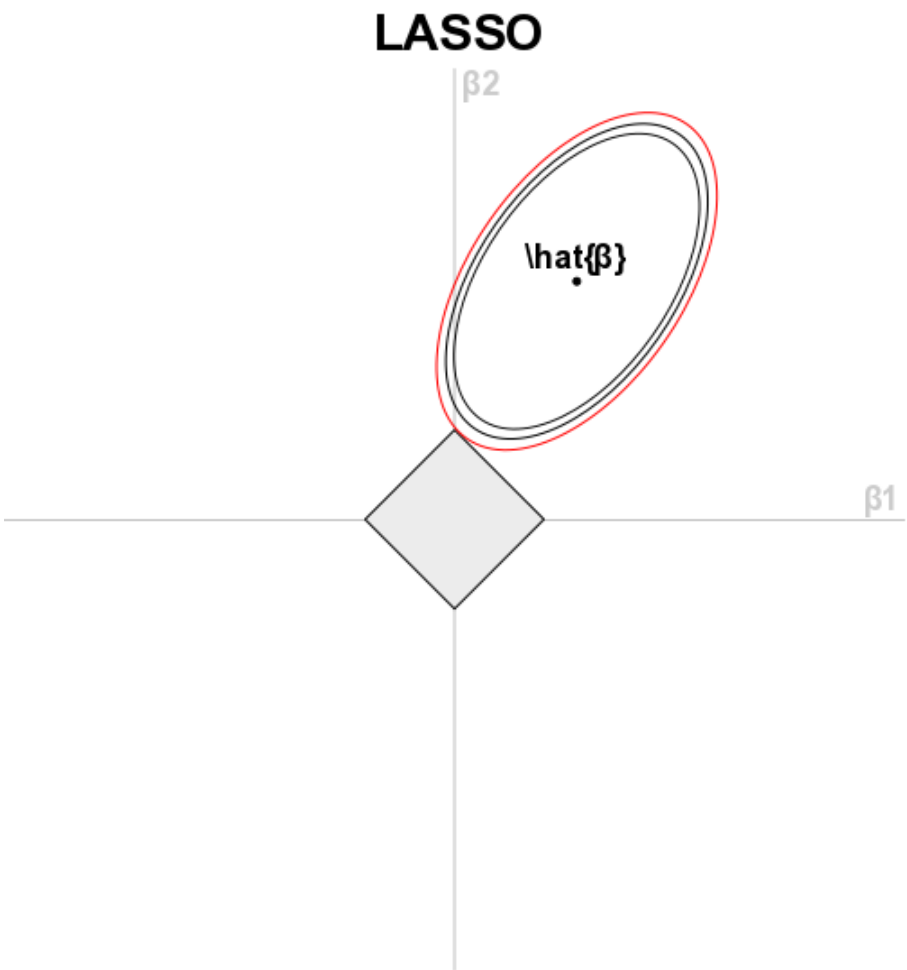
Variable selection with LASSO

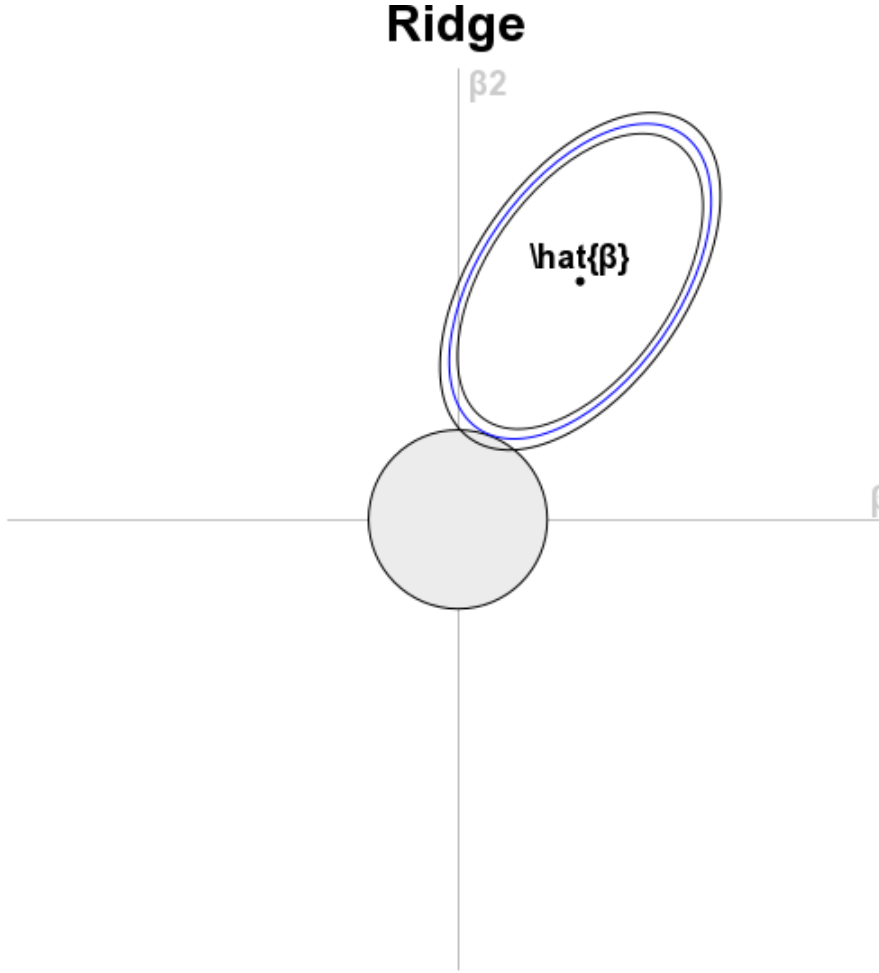


- x and y-axis are the coefficients for the predictors; z-axis is the error
  - OLS will find solutions to minimize the error

Variable selection with LASSO

- Goal: find first point where the elliptical contours intersect the constraint region ($\sum_{j=1}^{p} |a_j|^q$)





- Blue ellipse is the RR solution, red is LASSO solution

Bayesian LASSO

$$(\hat{\mu}, \hat{a}) = \text{argmin} \left\{ \sum_i (y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2 + \frac{\sigma_e^2}{\sigma_a^2} \sum_{j=1}^{p} |a_j| \right\}$$

$$(\hat{\mu}, \hat{a}) = \text{argmin} \left\{ \prod_i^n \exp\left(\frac{(y_i - \mu - \sum_{j=1}^{p} a_j w_{ij})^2}{2\sigma^2} \right) \right\} \left\{ \prod_{j=1}^{p} \exp\left(-\frac{\lambda |a_j|}{2\sigma^2} \right) \right\}$$

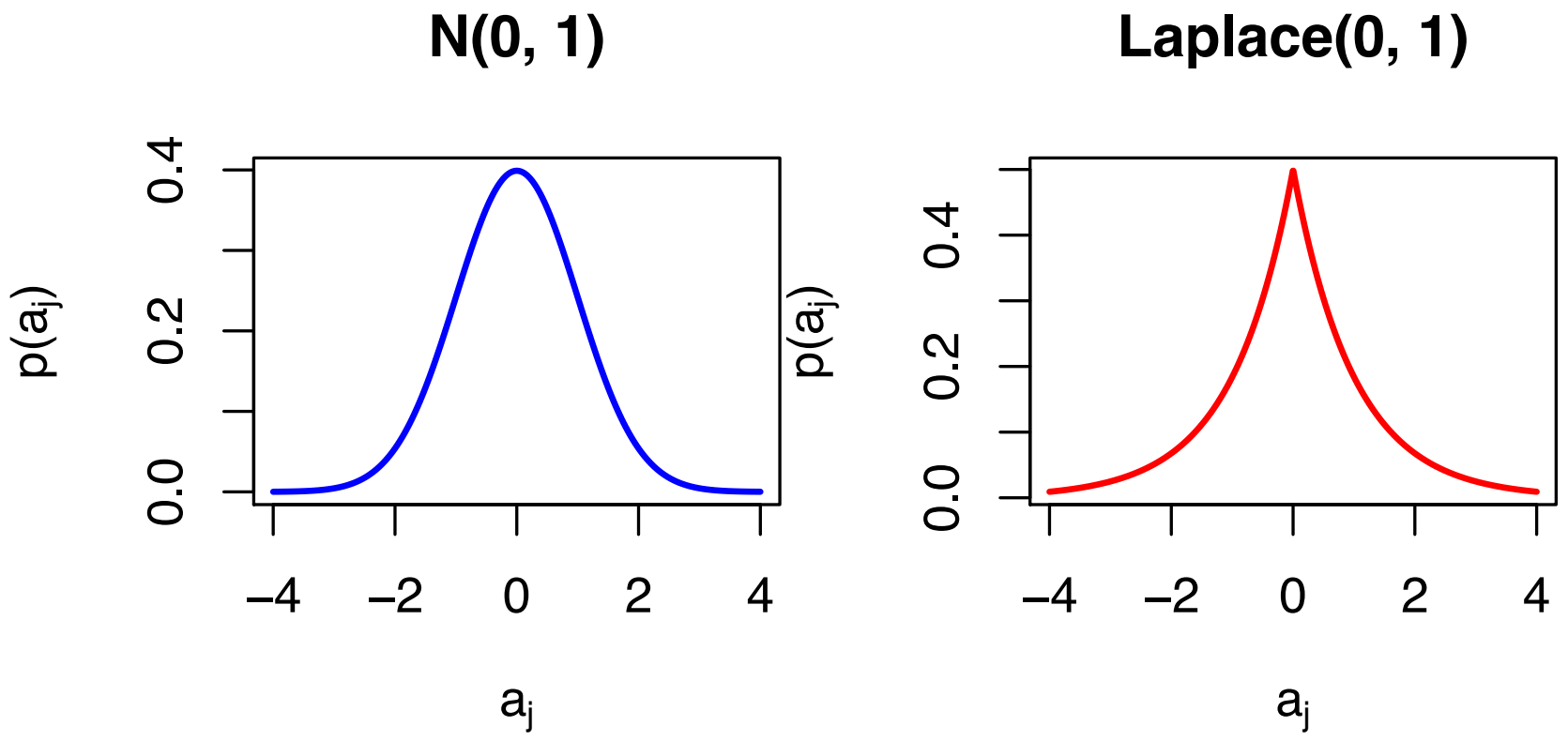
- The posterior is the same as RR, but the **prior is different**

Bayesian LASSO

$$(\hat{\mu}, \hat{a}) = \operatorname{argmin} \left\{ \prod_{i}^{n} \exp\left(\frac{(y_i - \mu - \sum_{j=1}^{p} x_{ij} a_j)^2}{2\sigma^2} \right) \right\} \left\{ \prod_{j=1}^{p} \exp\left(-\frac{\lambda |a_j|}{2\sigma^2} \right) \right\}$$

- Probability density function for Laplace distribution: $f(x|a, b) = \frac{1}{2b} exp(\frac{-|x-a|}{b})$

Priors for Bayesian ridge and LASSO

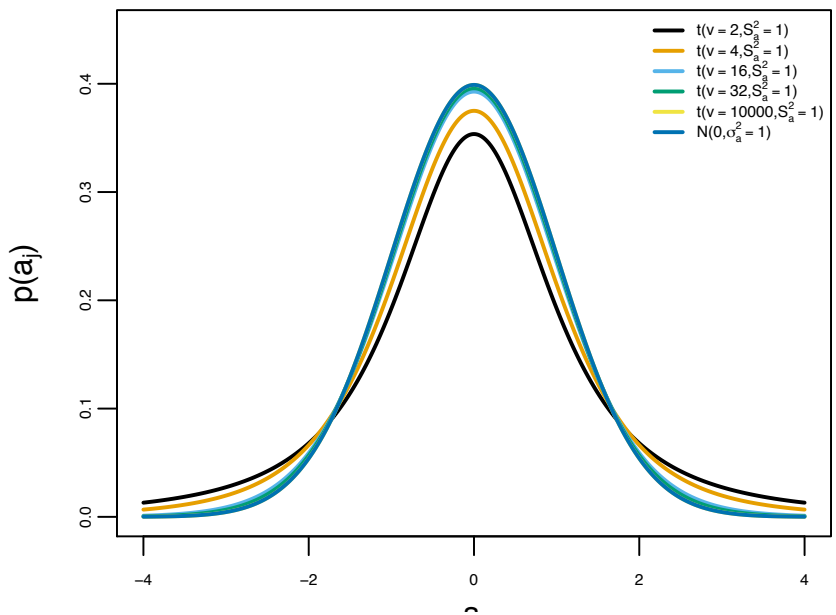


Different Bayesian flavors

- ~~Bayesian Ridge Regression~~
- ~~Bayesian LASSO~~
- BayesA
- BayesB
- BayesC
- BayesCπ

BayesA: scaled t

- Bayes A: Each marker has a marker-specific variance $a_j \sim \mathrm{N}(0, \sigma^2_{a_j})$
- Prior: Infinite mixture of normal distributions = scaled t $(t(a_j|v, S_a^2))$
  - Scale (S_a^2); degrees of freedom (v)

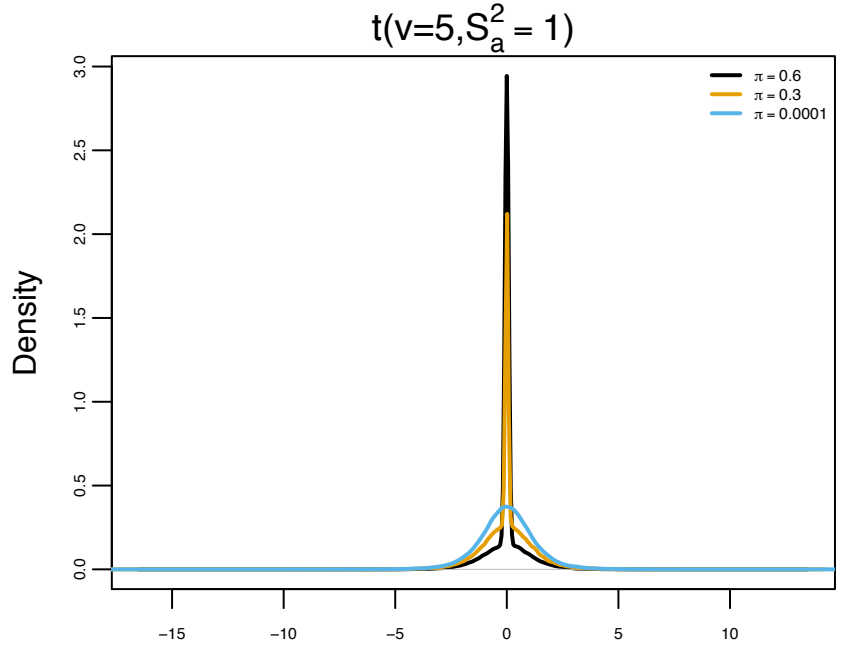


BayesA vs ridge regression

- Small-effect loci: Shrunk closer to 0 with BayesA compared to RR
- Large-effect loci: Less shrinkage with BayesA compared to RR
- How is shrinkage affected by degrees of freedom (v)?

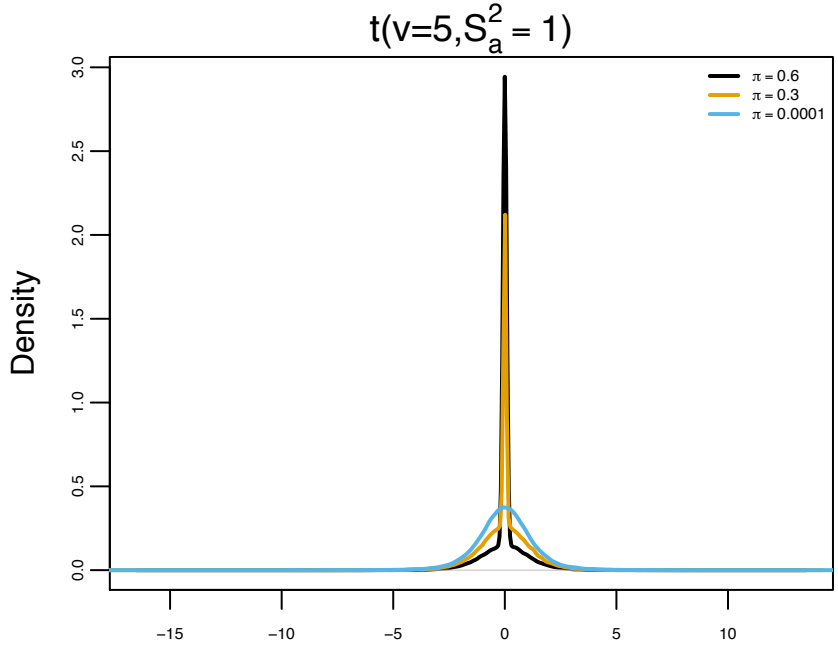
BayesB: spike-slab density

- Like BayesA, with BayesB each marker has a marker-specific variance
- Prior: Spike-slab
  - Spike: marker effect will have point mass at 0 with a probability π
  - Slab: marker effect will have t distribution with scale (S_a^2) and degrees of freedom (v), and **a probability** $1 - \pi$



BayesB vs Bayes A

- How does shrinkage compare between BayesA and BayesB for small-effect loci?
  - Assume $\pi = 0.3$



Bayes C and BayesCπ

- BayesC:
  - Nearly the same as BayesB, however the prior is slightly different
  - Prior: Spike-slab
    - Spike: marker effect will have point mass at 0 with a probability π
    - Slab: marker effect will follow a normal distribution ($N(0, \sigma_a^2 | v, S_a^2)$) and **a probability** $1 - \pi$
  - Assumes π is known
- BayesCπ: π is unknown
  - Uniform (flat, uninformative) prior is assigned to π

Summary of Bayesian WGR methods

- **BRR:** a's assigned N(0, σ_a^2); variance parameter (σ_a^2) assigned $\chi^{-2}(\sigma_a^2|df_a, S_a)$
- **BL:** a's assigned N(0, $\tau_{jk}^2\sigma_e^2$); marker-specific variance parameter (τ_j^2) assigned $DE(\tau_{jk}^2|\frac{\lambda^2}{2})$
- **BayesA:** a's assigned N(0, $\sigma_{a_j}^2$); each marker-specific variance ($\sigma_{a_j}^2$) is assigned $\chi^{-2}(\sigma_{a_j}^2|df_a, S_a)$ which is the same as a scaled t dist.
- **BayesB:** a's assigned 0 with probability π and N(0, $\sigma_{a_j}^2$) with probability (1 - π); each marker-specific variance ($\sigma_{a_j}^2$) is assigned $\chi^{-2}(\sigma_{a_j}^2|df_a, S_a)$ which is the same as a scaled t dist.
- **BayesC:** a's assigned 0 with probability π and N(0, $\sigma_a^2|v, S_a^2$) with probability (π).
  - **BayesCπ**: BayesC with a flat prior on π

Inference on Bayesian models

What is the genetic architecture for trait X?

1. Define the model: Prior knowledge of unknown parameters, data generating process (likelihood)

- $p(\mu, a, \sigma^2 | y, \omega) \propto$
  $\prod_{i=1}^{n} \mathrm{N}(y_i | \mu + \sum_{j=1}^{p} w_{ij} a_j, \sigma^2) \prod j = 1^{p} p(a_j | \omega) p(\sigma^2)$

2. Given some data estimate the unknown parameters via Markov chain Monte Carlo method

- Monte Carlo: Random sampling
- Markov Chain: Sequence where value for current iteration depends on previous iteration

Gibbs sampling (very brief)

0. Set initial values for all unknown parameters
1. For iteration i in 1:nIter

i. For predictor p in 1:nPred, sample variable p conditional on all others
ii. Repeat i. until $p ==$ nPred

2. Repeat 1.

Visualization of simple distribution

P-values for marker effects

- For frequentist approaches calculating p-values is simple, for Bayesian... not so much
- p-values for ridge regression
  1. Define t-statistic: $t = \frac{\hat{a}}{\sqrt{Var(\hat{a})}}$
  2. Get p-value: $p - value = 2(1 - \Phi(|t|))$; Φ is the CDF for the Normal dist.

P-values for RR-BLUP

- In rrBLUP

```
foo <- mixed.solve(y = y, Z = t(geno[4:ncol(geno)]),
                   K = NULL, X = NULL, SE = T)
str(foo)
```

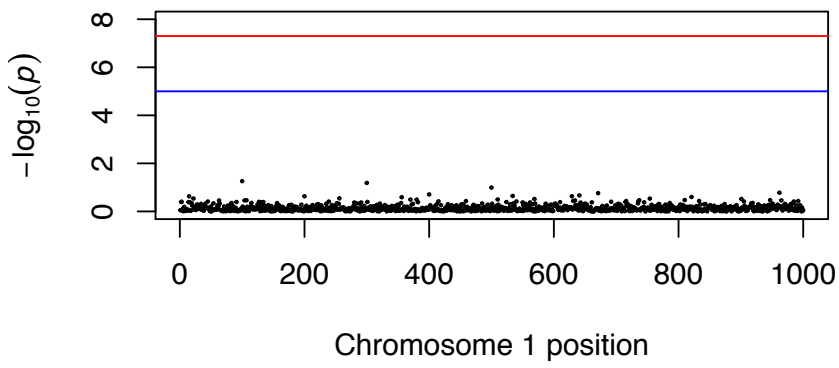
```
## List of 7
##  $ Vu     : num 0.00579
##  $ Ve     : num 1.45
##  $ beta   : num [1(1d)] 0.17
##  $ beta.SE: num [1(1d)] 0.177
##  $ u      : num [1:1000(1d)] 0.00975 0.058 0.05923 0.00558 -0.02076
##  $ u.SE   : num [1:1000(1d)] 0.0705 0.0695 0.0699 0.07 0.0705 ...
##  $ LL     : num -473
```

P-values for RR-BLUP

- In rrBLUP

```
foo <- mixed.solve(y = y, Z = t(geno[4:ncol(geno)]),
                   K = NULL, X = NULL, SE = T)

SNPe_ad <- foo$u / foo$u.SE
pvals <- 2*(1-pnorm(abs(SNPe_ad)))
```

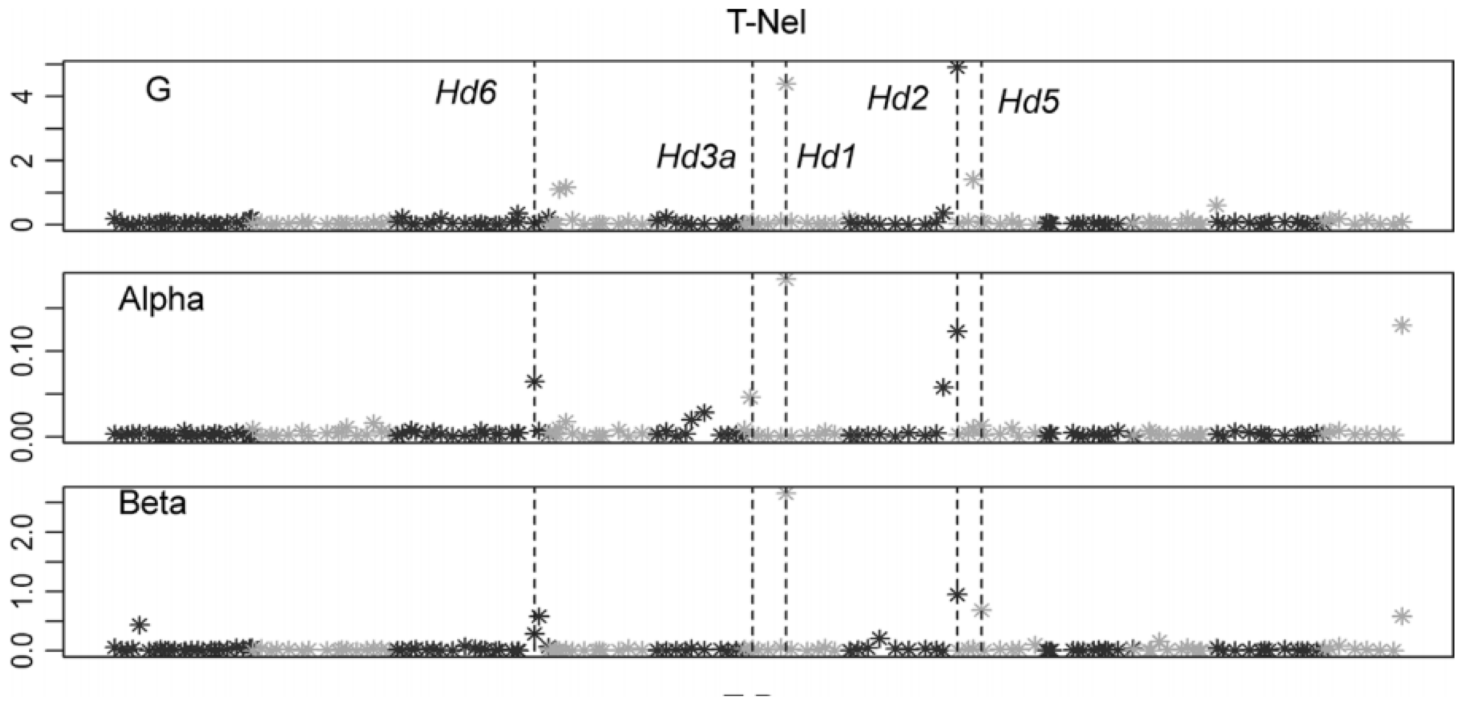


Finding informative SNPs with Bayesian approaches

- Ranking marker effects
- Bayesian model frequency: proportion of samples in which *a* has a non-zero effect
- Window-based genetic variance

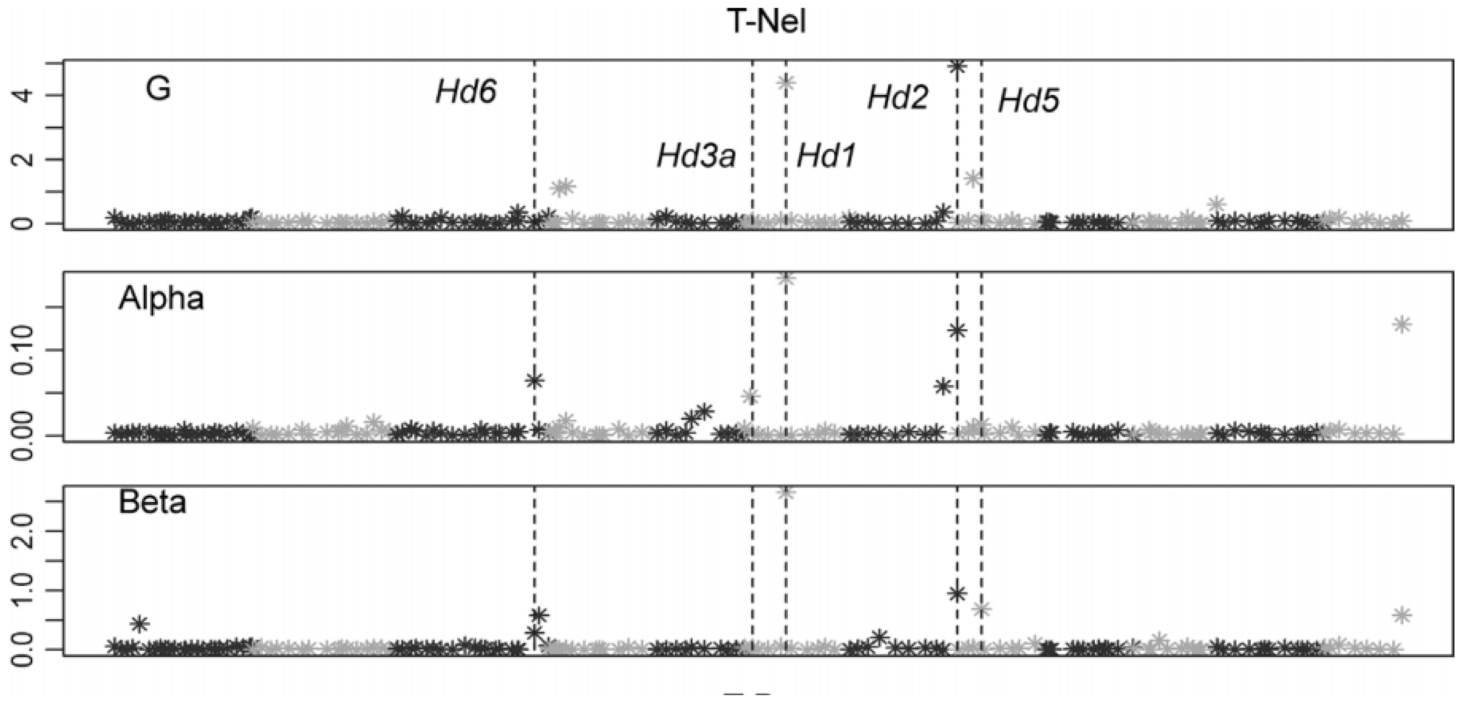
Ranking marker effects

- Rank effect sizes and select just the top SNPs



Ranking marker effects

- Rank effect sizes and select just the top SNPs



- Problems??

Window-based genetic variance

1. Define window
2. Calculate the proportion of genetic variance explained markers in a window relative to the total genetic variance

Window-based genetic variance

- Genetic value for each window $(_w)$

$$g_w = W_w a_w$$

- Total genetic value

$$g = Wa$$

Window-based genetic variance

- Genetic variance explained by window

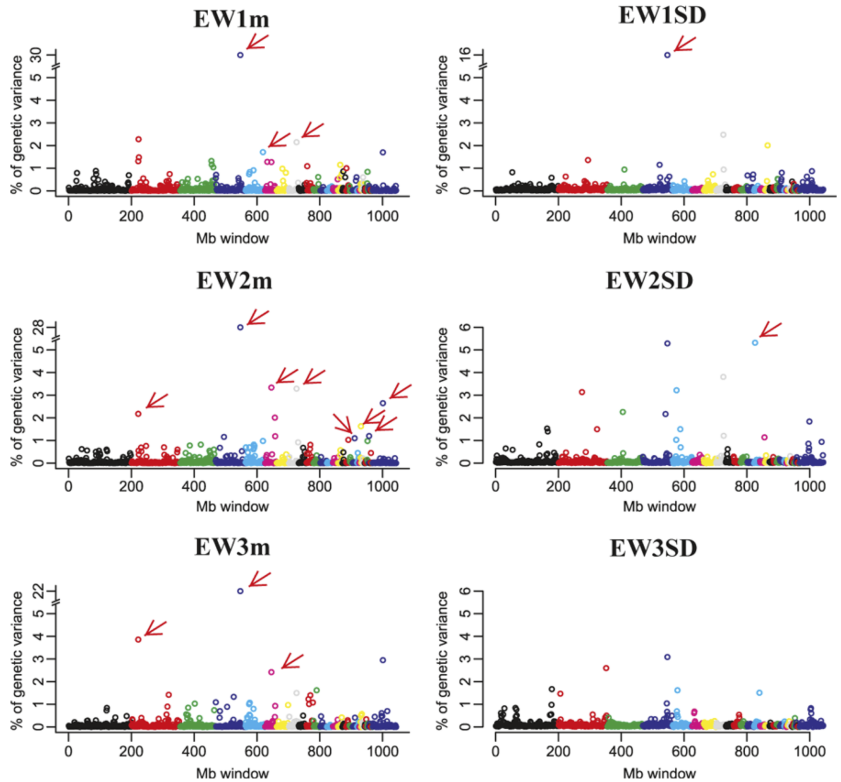
$$\sigma^2_{gw} = \frac{\sum_{j=1}^{n} g^2_{w_j}}{n} - (\frac{\sum_{j=1}^{n} g^2_{w_j}}{n})^2$$

- Total genetic variance

$$\sigma^2_{g} = \frac{\sum_{j=1}^{n} g^2_{j}}{n} - (\frac{\sum_{j=1}^{n} g^2_{j}}{n})^2$$

Window-based genetic variance

- Plot $\frac{\sigma^2_{gw}}{\sigma^2_g}$



Animal Genetics, Volume: 43, Issue: s1, Pages: 87-96, First published: 28 June 2012, DOI: (10.1111/j.1365-2052.2012.02381.x)

Bayesian model frequency

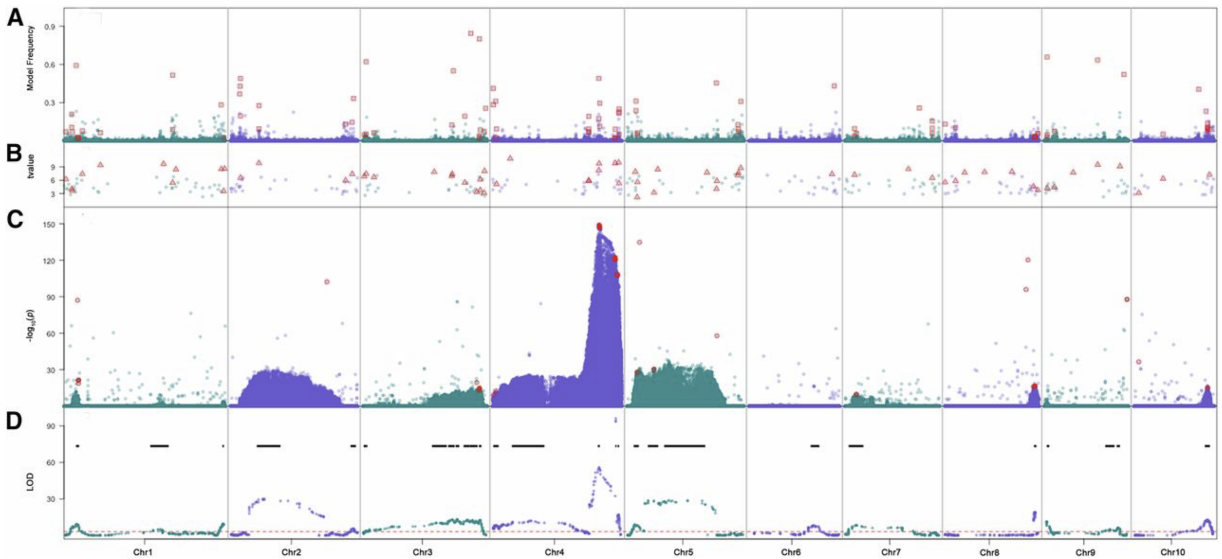
- Bayesian model frequency: **proportion of samples (MCMC chains)** in which *a* has a non-zero effect
  - **Which Bayesian methods can we use for model frequency?**

Bayesian model frequency

- Bayesian model frequency: proportion of samples (MCMC chains) in which *a* has a non-zero effect
  - BayesB
  - BayesC and BayesCπ

Bayesian model frequency in plants

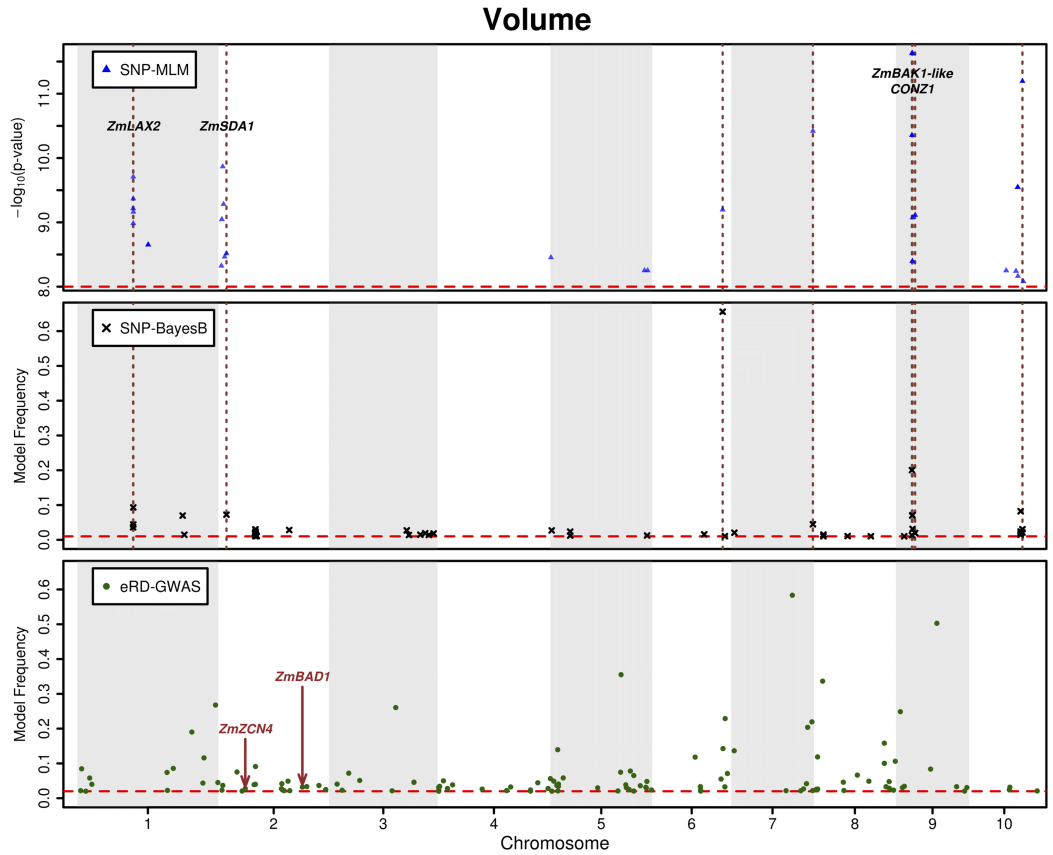
- GWAS for kernel row number in maize



Jinliang Yang et al. G3 2018;8:3567-3575

Bayesian model frequency in plants

- GWAS for shoot apical meristem volume in maize



https://doi.org/10.1186/s13059-017-1328-6

Questions??