Bayesian Group Sparse Multi-Task Regression for Imaging Genomics

Keelin Greenlaw[1], Farouk Nathoo[1], Mary Lesperance[1]
Elena Szefer[2], Jinko Graham[2]

[1]University of Victoria, [2]Simon Fraser University

February 2, 2016

Outline

Introduction: Imaging Genomics

- Imaging genetics: interest in associations between genetic variations and neuroimaging measures as quantitative traits (QTs).
- Compared to case-control status, the QTs derived through neuroimaging may have have increased statistical power, may be closer to the underlying biological etiology of disease, perhaps making it easier to identify underlying genes.

Introduction: Imaging Genomics

- Statistically, interested in a multivariate regression analysis, where the response vector comprises potentially interlinked brain imaging phenotypes that we relate to high-throughput single nucleotide polymorphism (SNP) data.
- We focus here on multivariate phenotypes (volumetric and cortical thickness values) of moderate dimension (e.g. $10 - 30$) derived from MRI for certain ROIs.
- The SNPs are naturally grouped by their belonging genes, and multiple SNPs from a given gene may jointly carry out genetic functionalities. Would like to explot this group structure in the regression analysis.

Introduction: Imaging Genomics

- We develop a Bayesian approach based on a continuous shrinkage prior that encourages sparsity and induces dependence in the regression coefficients corresponding to SNPs within the same gene, and across different components of the imaging phenotypes.
- Our approach is related to the Bayesian group lasso (Park and Casella, 2008; Kyung et al., 2010) but adapted for multivariate phenotypes.
- Primarily motivated by the Group-Sparse Multi-task regression and feature selection estimator (somewhat) recently proposed by Wang et al. [2012].

Outline

Wang et al. Estimator: Data set up

Imaging data

- $\mathbf{y}_\ell = (\mathbf{y}_{\ell 1}, \ldots, \mathbf{y}_{\ell c})^T, \ \ \ell = 1, \ldots, n$
- n subjects; c response variables (QTs)

Genetic data

- $\mathbf{x}_\ell = (\mathbf{x}_{\ell 1}, \ldots, \mathbf{x}_{\ell d})^T, \ \ \ell = 1, \ldots, n$
- $\mathbf{x}_{\ell j} \in \{0, 1, 2\}$ is the number of minor allele for j^{th} SNP.
- d SNPs, which can be grouped into K genes: π_k for $k = 1, 2, \ldots, K$.

Regression coefficients

- $E(\mathbf{y}_\ell) = \mathbf{W}^T \mathbf{x}_\ell, \ \ \ell = 1, \ldots, n$
- $\mathbf{W}$ is a d x c matrix; each w_{ij} is a coefficient.

Wang et al. Estimator: First major component

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \left\{ \sum_{\ell=1}^{n} ||\mathbf{W}^T \mathbf{x}_\ell - \mathbf{y}_\ell||_2^2 + \gamma_1 \sum_{k=1}^{K} \sqrt{\sum_{i \in \pi_k} \sum_{j=1}^{c} w_{ij}^2} + \gamma_2 \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{c} w_{ij}^2} \right\}$$

- Residual sum of squares; element w_{ij} of $\mathbf{W}$ measures the relative importance of the i^{th} SNP to the j^{th} phenotype.

Wang et al. Estimator: Second major component

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \left\{ \sum_{\ell=1}^{n} ||\mathbf{W}^T \mathbf{x}_\ell - \mathbf{y}_\ell||_2^2 + \gamma_1 \sum_{k=1}^{K} \sqrt{\sum_{i \in \pi_k} \sum_{j=1}^{c} w_{ij}^2} + \gamma_2 \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{c} w_{ij}^2} \right\}$$

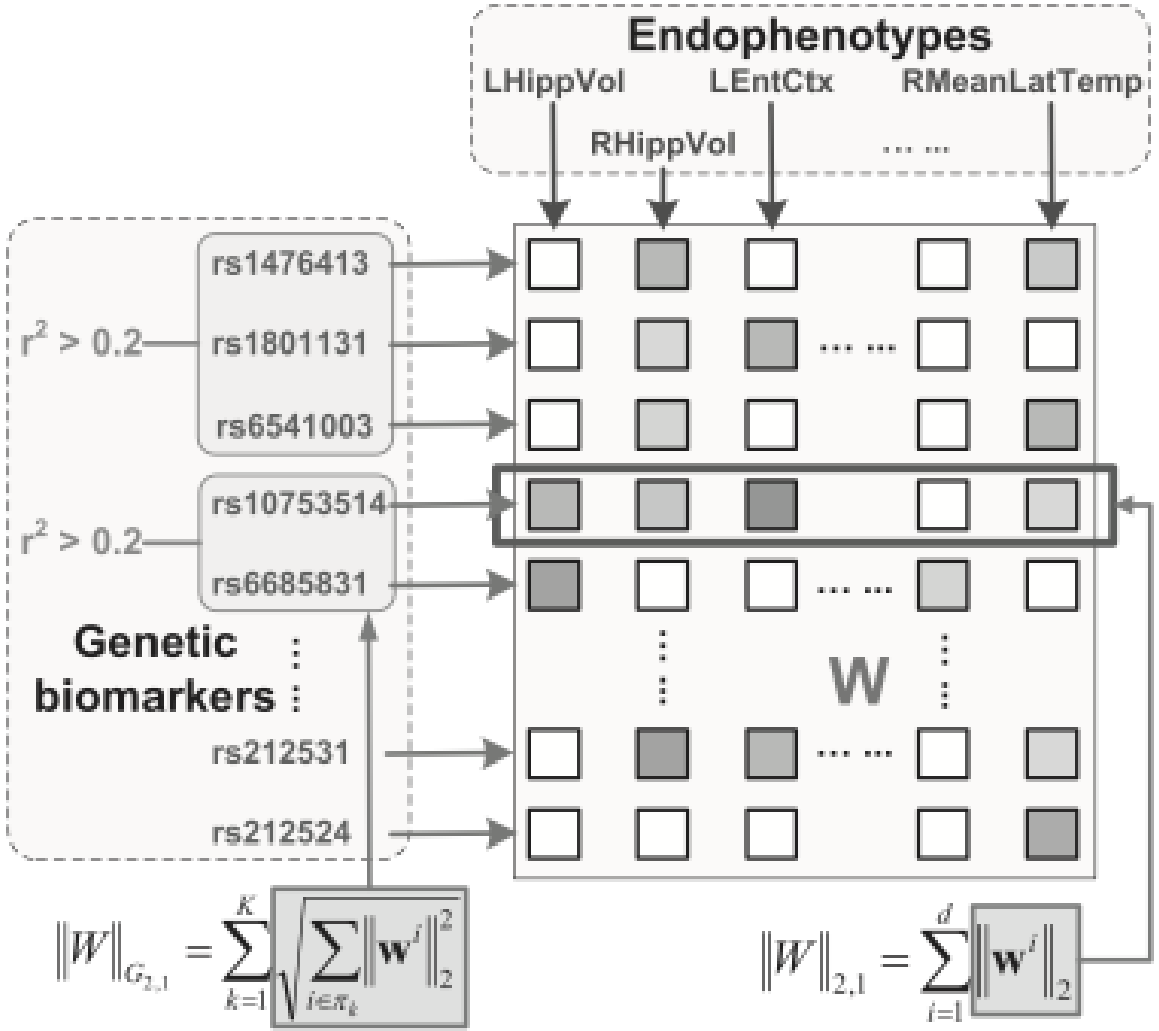
- Inspired by group lasso [Yuan and Lin, 2006], Wang et al. introduce a new form of regularization ($G_{2,1} - norm$) to address group-wise association among SNPs.
- Coefficients within a group, across all QTs, are penalized together via $\ell_2 - norm$ while $\ell_1 - norm$ is used to sum up group-wise penalties to enforce sparsity between groups.
- $G_{2,1} - norm$ regularization differs from group lasso as it penalizes regression coefficients for a group of SNPs across all responses jointly.

Wang et al. Estimator: Third major component

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \left\{ \sum_{\ell=1}^{n} ||\mathbf{W}^T \mathbf{x}_\ell - \mathbf{y}_\ell||_2^2 + \gamma_1 \sum_{k=1}^{K} \sqrt{\sum_{i \in \pi_k} \sum_{j=1}^{c} w_{ij}^2} + \gamma_2 \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{c} w_{ij}^2} \right\}$$

- As an important group may contain irrelevant individual SNPs, or a less important group may contain individually significant SNPs, an additional penalty term is added for individual structured sparsity.
- The second penalty term enforces $\ell_{2,1} - norm$ regularization for individual SNPs.

Wang et al. Estimator: 'G-SMuRFS'



Wang et al. Estimator: 'G-SMuRFS'

Group-sparse multitask regression and feature selection

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \left\{ \sum_{\ell=1}^{n} ||\mathbf{W}^T \mathbf{x}_\ell - \mathbf{y}_\ell||_2^2 + \gamma_1 \sum_{k=1}^{K} \sqrt{\sum_{i \in \pi_k} \sum_{j=1}^{c} w_{ij}^2} + \gamma_2 \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{c} w_{ij}^2} \right\}$$

- The combination of both penalty terms make up the novel method for SNP selection, dubbed 'G-SMuRFS' by the authors.
- Computation of $\hat{\mathbf{W}}$ is based on a simple iterative algorithm that converges to the global optimum.
- Tuning parameters, γ_1 and γ_2, are chosen by standard 5-fold cross-validation in the range of $(10^{-5}, 10^{-4}, \ldots, 10^4, 10^5)$.

Wang et al. Estimator: Limitations

- The proposed method only provides a point estimate of the regression coefficients. A method for computing standard errors is lacking.
- By noting the connection between penalized regression methods and Bayesian models, [Kyung et al., 2010, Park and Casella, 2008] we develop an equivalent hierarchical Bayesian model.
- This allows for inference based on the posterior distributions. As we can validly summarize the spread of the posterior, we have valid measures of variability. Interval estimates can then guide SNP selection.

Outline

Bayesian Model: Priors $\mathbf{W}$

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \left\{ \sum_{\ell=1}^{n} ||\mathbf{W}^T \mathbf{x}_\ell - \mathbf{y}_\ell||_2^2 + \gamma_1 \sum_{k=1}^{K} \sqrt{\sum_{i \in \pi_k} \sum_{j=1}^{c} w_{ij}^2} + \gamma_2 \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{c} w_{ij}^2} \right\} \quad (1)$$

We specify a model hierarchy such that the posterior mode is identical to $\hat{\mathbf{W}}$ in (1).
First level: quantitative imaging traits, conditional on $\mathbf{W}$ and σ^2, are independently distributed as multivariate normal.

$$\mathbf{y}_\ell \,|\mathbf{W}, \sigma^2 \overset{ind}{\sim} MVN_c(\mathbf{W}^T \mathbf{x}_\ell \,,\, \sigma^2 I_c) \quad \ell = 1, \ldots, n$$

Bayesian Model

Let $\mathbf{W}^{(k)} = \{w_{ij} | i \in \pi_k, j = 1, \ldots, c\}$ be submatrix with rows corresponding to the k^{th} gene, $k = 1, \ldots, K$.

We assign conditionally independent priors to each $\mathbf{W}^{(k)}$ to coincide with the penalty terms in (1) as follows:

$$\mathbf{W}^{(k)} | \lambda_1, \lambda_2, \sigma^2 \overset{ind}{\sim} p(\mathbf{W}^{(k)} | \lambda_1, \lambda_2, \sigma^2) \quad k = 1, \ldots, K \tag{2}$$

$$p(\mathbf{W}^{(k)} | \lambda_1, \lambda_2, \sigma^2) \propto \exp\left\{ -\frac{\lambda_1}{\sigma} \sqrt{\sum_{i \in \pi_k} \sum_{j=1}^{c} w_{ij}^2} \right\} \prod_{i \in \pi_k} \exp\left\{ -\frac{\lambda_2}{\sigma} \sqrt{\sum_{j=1}^{c} w_{ij}^2} \right\}. \tag{3}$$

Proposition 1. (Prior Propriety) The prior for $\mathbf{W}$ based on (2) and (3) is proper.

- Density of a product multivariate Laplace distribution induces dependence in coefficients across imaging phenotypes at both the SNP and gene level.
- Given the likelihood and prior the posterior mode is by construction the estimator of Wang et al. [2012].

Proposition 2. (Scale mixture representation) For each $i \in \{1, \ldots, d\}$ let $k(i) \in \{1, \ldots, K\}$ denote the gene associated with the i^{th} SNP. The prior (3) can be obtained through the following scale mixture representation:

$$w_{ij} \mid \sigma^2,\ \tau_1^2, \ldots, \tau_K^2,\ \omega_1^2, \ldots, \omega_d^2 \overset{ind}{\sim} N\left(0,\ \sigma^2\left(\frac{1}{\tau_{k(i)}^2} + \frac{1}{\omega_i^2}\right)^{-1}\right), \tag{4}$$

with continuous scale mixing variables $\tau^{\mathbf{2}} = (\tau_1^2, \ldots, \tau_K^2)'$ and $\omega^{\mathbf{2}} = (\omega_1^2, \ldots, \omega_d^2)'$ distributed according to the density

$$p(\tau^{\mathbf{2}}, \omega^{\mathbf{2}} | \lambda_1^2, \lambda_2^2) \propto \prod_{k=1}^{K} \left(\frac{\lambda_1^2}{2}\right)^{\left(\frac{m_k c+1}{2}\right)} (\tau_k^2)^{\left(\frac{m_k c+1}{2}\right)-1} \exp\left\{-\left(\frac{\lambda_1^2}{2}\right)\tau_k^2\right\}$$
$$\times \left[\prod_{i \in \pi_k} \left(\frac{\lambda_2^2}{2}\right)^{\left(\frac{c+1}{2}\right)} (\omega_i^2)^{\left(\frac{c+1}{2}\right)-1} \exp\left\{-\left(\frac{\lambda_2^2}{2}\right)\omega_i^2\right\} (\tau_k^2 + \omega_i^2)^{-\frac{c}{2}}\right] \tag{5}$$

Outline

Model Fitting: Full Conditionals

The proposed hierarchical model results in standard full conditional distributions (Gaussian, Inverse-Gaussian, Inverse-Gamma).

- $[\text{vec}(\mathbf{W}^{(k)T}) | \mathbf{Y}, \mathbf{W}^{(-k)}, \underset{\sim}{\tau^2}, \underset{\sim}{\omega^2}, \sigma^2, \lambda_1^2, \lambda_2^2] \sim MVN_{m_k c} \quad k = 1, \ldots, K$
- $[\nu_k = \frac{1}{\tau_k^2} \mid \mathbf{Y}, \mathbf{W}, \tau^2_{(-k)}, \underset{\sim}{\omega^2}, \sigma^2, \lambda_1^2, \lambda_2^2] \sim \textit{Inverse-Gaussian} \quad \text{for } k = 1, \ldots, K$
- $[\eta_i = \frac{1}{\omega_i^2} \mid \mathbf{Y}, \mathbf{W}, \underset{\sim}{\tau^2}, \omega^2_{(-i)}, \sigma^2, \lambda_1^2, \lambda_2^2] \sim \textit{Inverse-Gaussian} \quad \text{for } i = 1, \ldots, d$
- $[\sigma^2 | \mathbf{Y}, \mathbf{W}, \underset{\sim}{\tau^2}, \underset{\sim}{\omega^2}, \lambda_1^2, \lambda_2^2] \sim \textit{Inv} - \textit{Gamma}$

Past work on Bayesian lassos [Park and Casella, 2008, Kyung et al., 2010] have discussed two methods for estimation of tuning parameters $(\lambda_1^2, \lambda_2^2)$.

Model Fitting: Estimation of λ_1^2 and λ_2^2

Fully Bayesian model

Assign conditionally conjugate gamma priors for λ_1^2 and λ_2^2.

$$\lambda_1^2 \sim Gamma(r_1, \delta_1)$$

$$\lambda_2^2 \sim Gamma(r_2, \delta_2)$$

The full conditional distributions can be derived in closed form.

λ_1^2 and λ_2^2 can be included as unknown parameters in the Gibbs Sampling algorithm.

Model Fitting: Estimation of λ_1^2 and λ_2^2

Empirical Bayes framework

An alternative approach is to estimate the tuning parameters by maximizing the marginal likelihood.

$$\hat{\lambda}_1^2,\ \hat{\lambda}_2^2 \ = \ \underset{\lambda_1^2,\lambda_2^2}{\arg\max} \int_{\Theta} p\left(\mathbf{Y}, \Theta \mid \lambda_1^2,\ \lambda_2^2\right) d\Theta$$

$$\textit{where } \Theta = (\mathbf{W}, \underset{\sim}{\tau^2}, \underset{\sim}{\omega^2}, \sigma^2)$$

This can be implemented using a Monte Carlo EM algorithm.

Model Fitting: Issues with λ_1^2 and λ_2^2 Estimation

We begin investigating the behaviour of our MCMC algorithm by simulating data from the model, where the underlying true $\mathbf{W}$ is known.

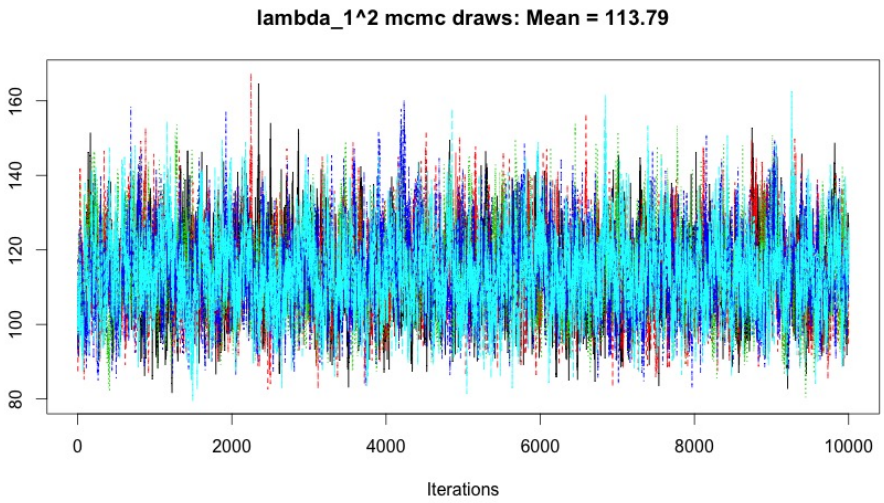
The behaviour changes drastically in two different settings.

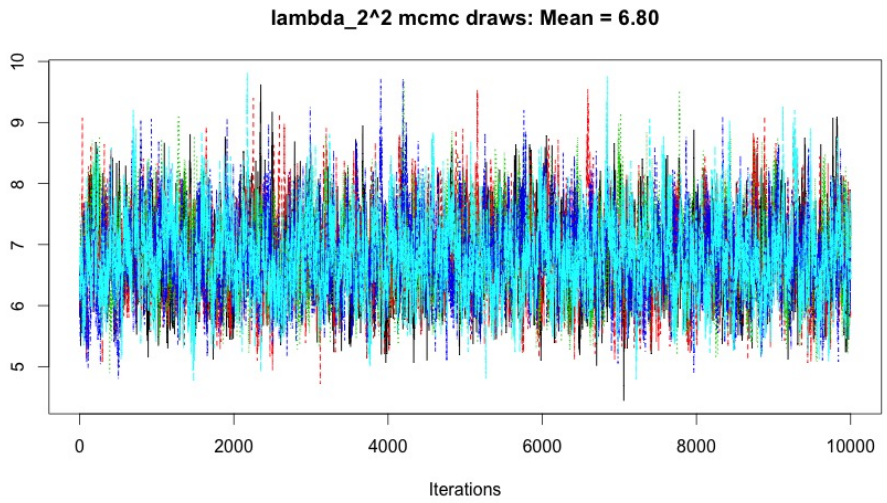
Case 1

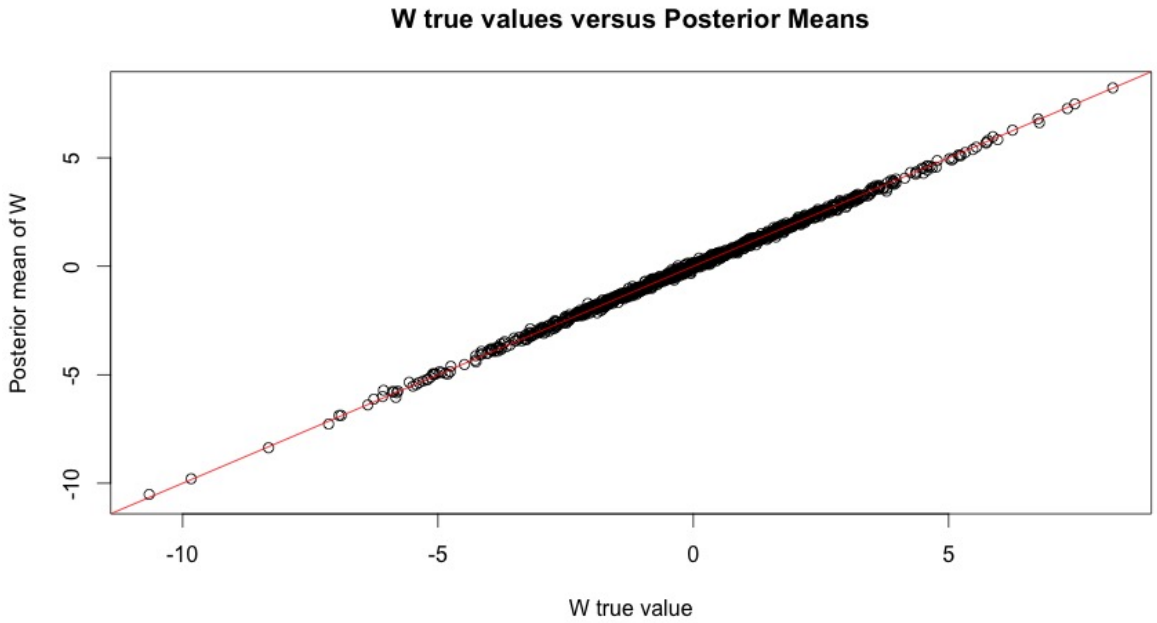
number of SNPs (d) $\ll$ number of simulated observations (n)

Behaviour:
Everything works fine!
Gibbs sampling λ_1^2 and λ_2^2 estimates converge to reasonable values.
MCEM converges.
Resulting posterior means of $\mathbf{W}$ are good estimates.

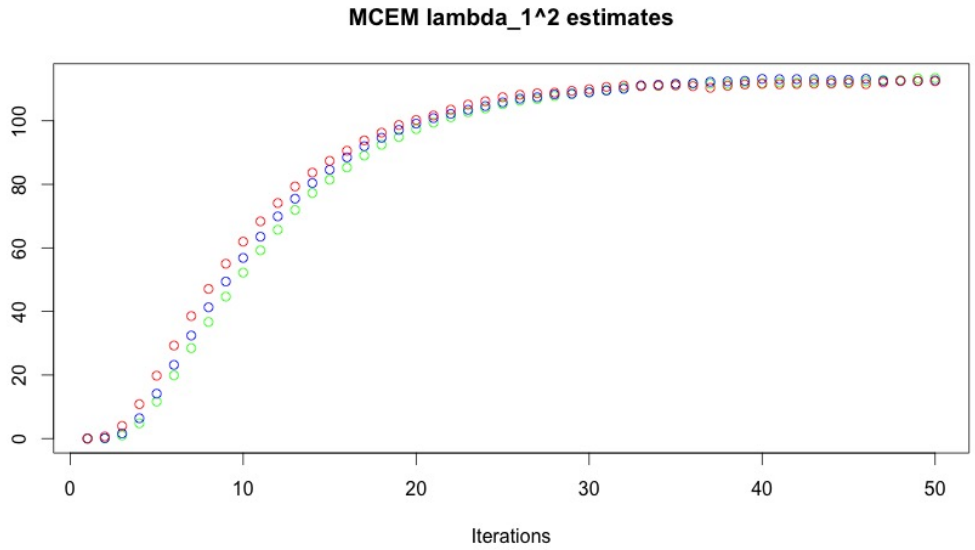
Simulation ($d = 200, n = 500$): Gibbs Sampling Results



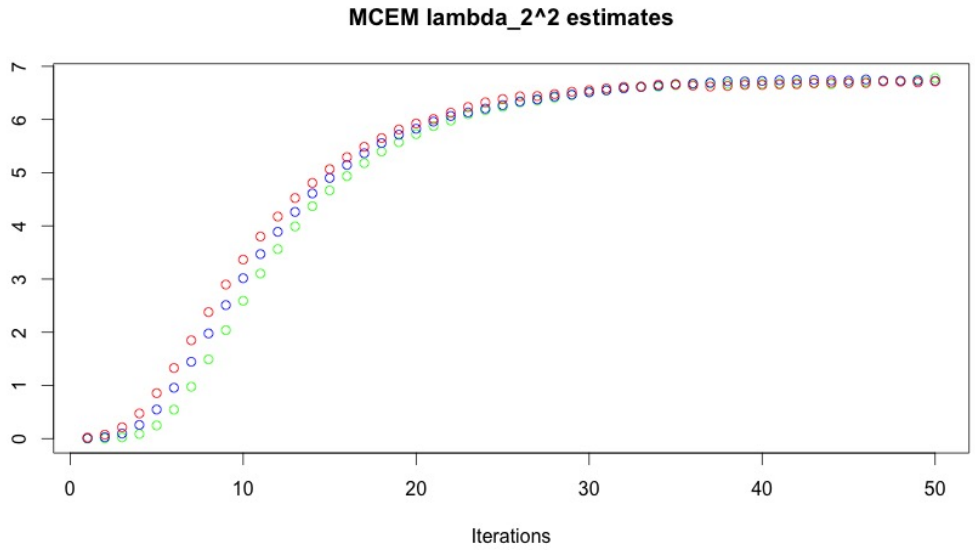




Simulation ($d = 200, n = 500$): Monte Carlo EM Results



λ_1^2 estimates converge to ≈ 111.3.



λ_2^2 estimates converge to ≈ 6.6.

Model Fitting: Issues with λ_1^2 and λ_2^2 Estimation

We begin investigating the behaviour of our MCMC algorithm by simulating data from the model, where the underlying true $\mathbf{W}$ is known.

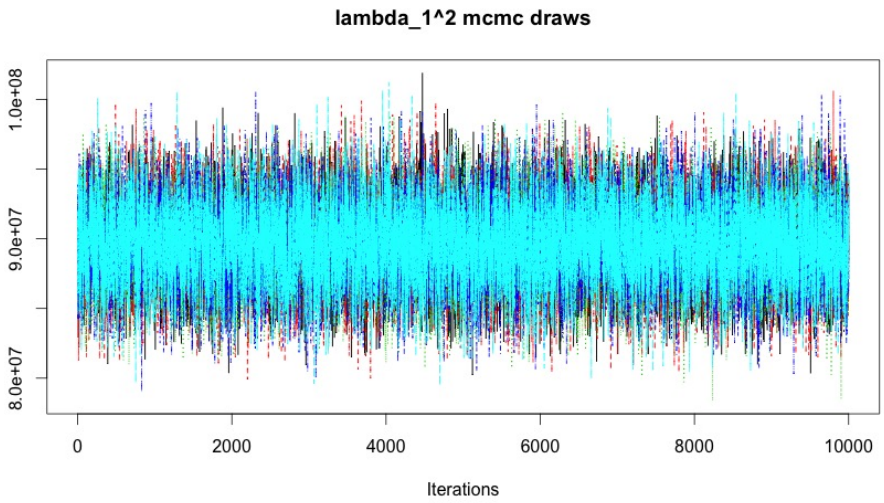
The behaviour changes drastically in two different settings.

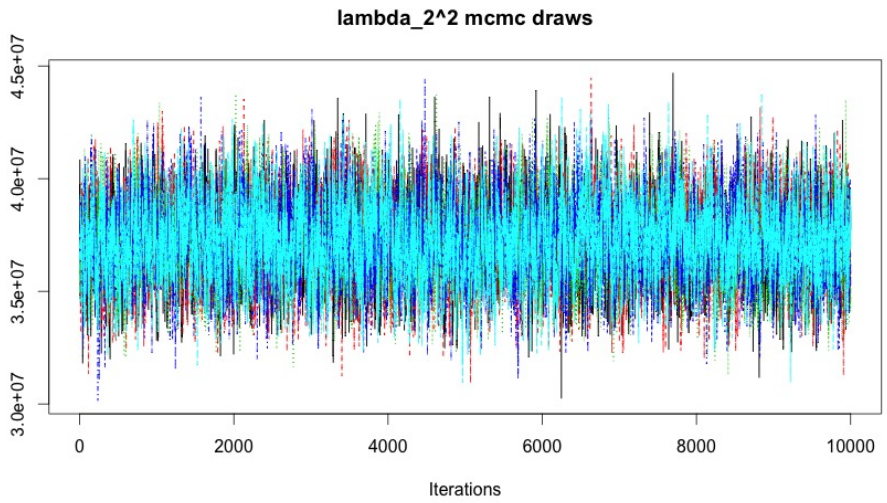
Case 2

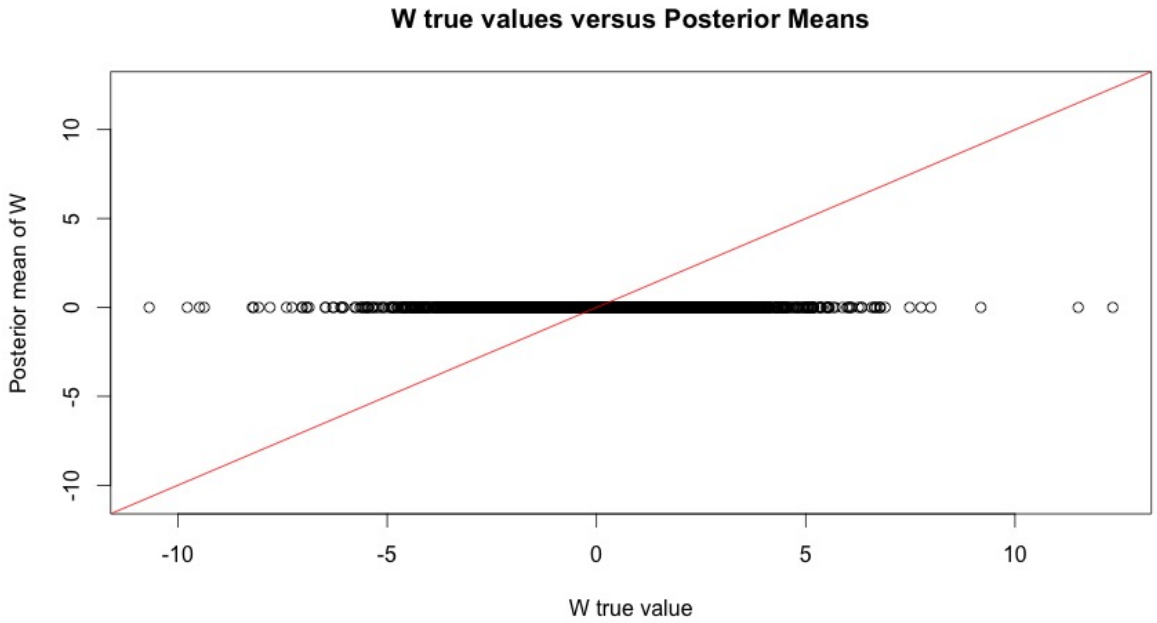
number of SNPs (d) $\approx$ or $\geq$ number of simulated observations (n)

Behaviour:
Gibbs sampling λ_1^2 and λ_2^2 estimates converge to very large values.
MCEM diverges.
Resulting posterior means of $\mathbf{W}$ are overshrunk; they are poor estimates.

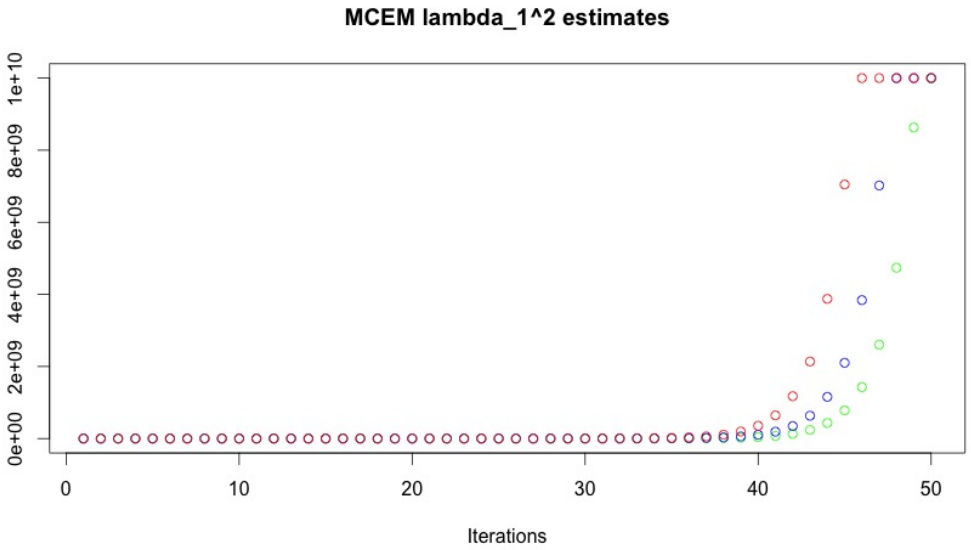
Simulation ($d = 510, n = 500$): Gibbs Sampling Results



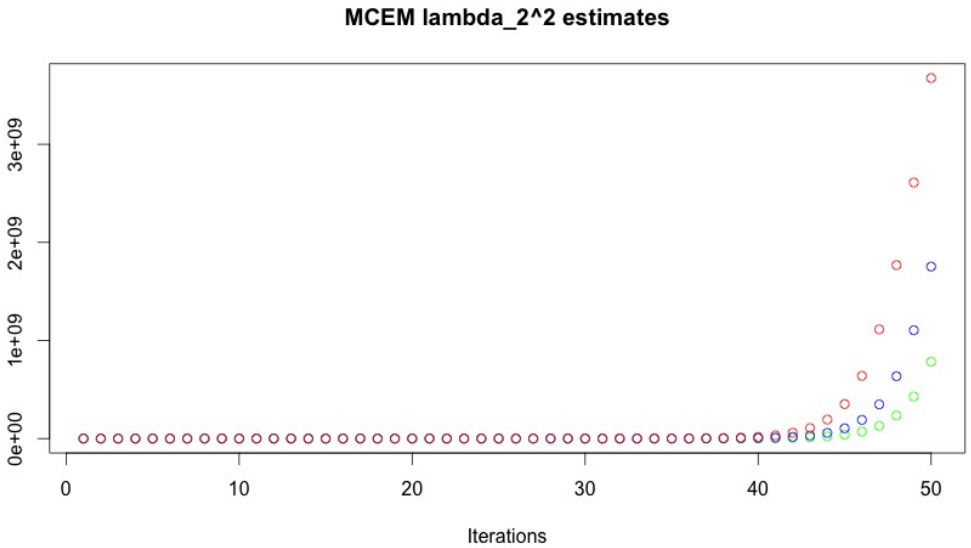




Simulation ($d = 510, n = 500$): Monte Carlo EM Results



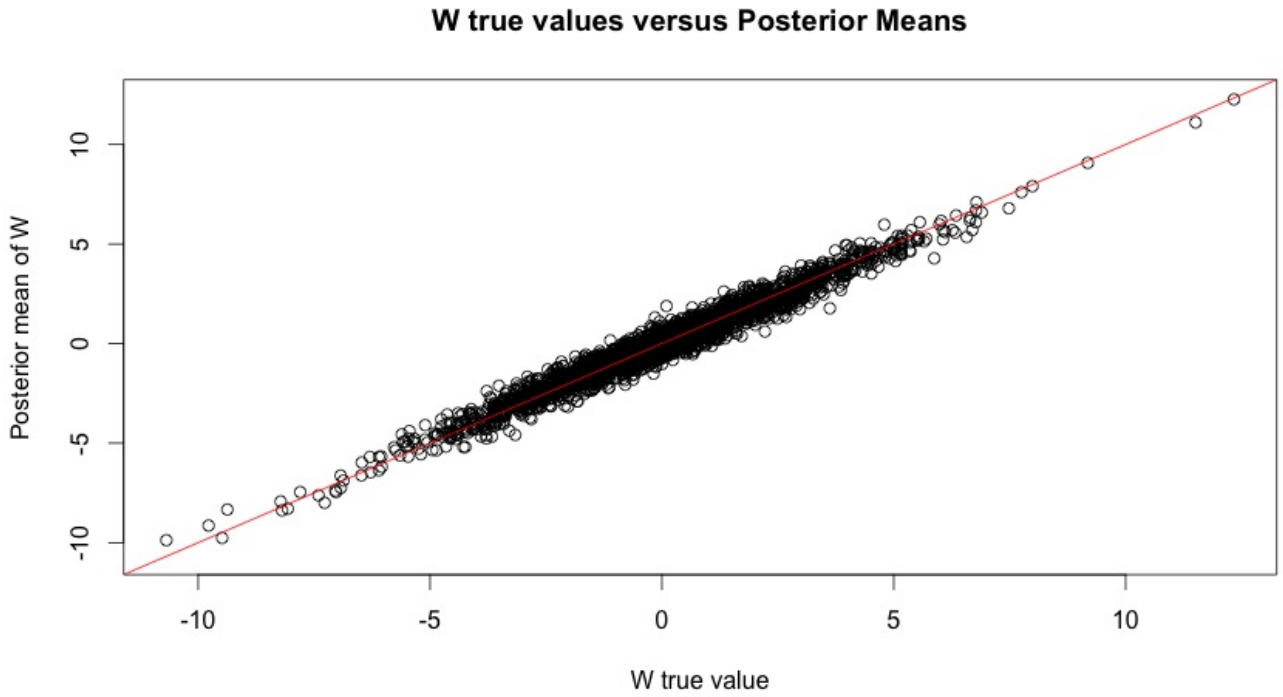
λ_1^2 estimates diverge to infinity.



λ_2^2 estimates diverge to infinity.

Simulation ($d = 510, n = 500$): Fixed λ_1^2 and λ_2^2 Results

Aside from the full Gibbs model and MCEM for estimation of the λ^2's, we note that when λ_1^2 and λ_2^2 are fixed at their true values, the mcmc algorithm performs well in cases where $d \geq n$.



Model Fitting: λ_1^2, λ_2^2 Estimation Discussion

Problem with choosing tuning parameters:

- With a large number of SNPs, and in particular with weak effects, choosing the tuning parameters based on the likelihood/posterior leads to over shrinkage.
- Study the shape of the marginal likelihood, $p(\mathbf{Y}|\,\lambda_1^2, \lambda_2^2)$.

Model Fitting: Studying the Marginal Likelihood

Marginal likelihood:

$$p(\mathbf{Y}|\,\lambda_1^2, \lambda_2^2) = \int p\left(\mathbf{Y}, \mathbf{W}, \sigma^2, \underset{\sim}{\tau^2}, \underset{\sim}{\omega^2} \,|\, \lambda_1^2, \lambda_2^2\right) d\,\mathbf{W}\, d\,\sigma^2\, d\,\underset{\sim}{\tau^2}\, d\,\underset{\sim}{\omega^2}$$

W is marginalized out of the expression by using the basic properties of the Gaussian distribution.

$$\mathbf{Y}\,|\,\underset{\sim}{\tau^2}, \underset{\sim}{\omega^2}, \sigma^2 \;\sim\; MVN(0,\; (I_c \otimes \mathbf{X})\Sigma_w(I_c \otimes \mathbf{X}^T) + \sigma^2\, I_{cn}\,)$$

$$where\;\; \Sigma_w = \sigma^2\, I_c \otimes Diag\left\{ \left(\frac{1}{\omega_i^2} + \frac{1}{\tau_{k(i)}^2} \right)^{-1},\; i = 1, \ldots, d \right\}$$

Model Fitting: Studying the Marginal Likelihood

$$p(\mathbf{Y}|\,\lambda_1^2,\lambda_2^2) = \int \left[\int_0^\infty p(\mathbf{Y},|\,\sigma^2,\underset{\sim}{\tau^2},\underset{\sim}{\omega^2})p(\sigma^2)d\,\sigma^2\right] p(\underset{\sim}{\tau^2}\,|\,\lambda_1^2)p(\underset{\sim}{\omega^2}\,|\,\lambda_2^2)\;d\,\underset{\sim}{\tau^2}\;d\,\underset{\sim}{\omega^2}$$

- Using properties of the Inv-Gamma distribution, σ^2 is analytically integrated out of the expression.
- The remaining integration is analytically intractable. We use a plug-in approximation.

$$p(\mathbf{Y}|\,\lambda_1^2,\;\lambda_2^2) \;=\; E_{\underset{\sim}{\tau^2},\,\underset{\sim}{\omega^2}}\left[p(\mathbf{Y}|\,\underset{\sim}{\tau^2},\underset{\sim}{\omega^2})\right] \;\approx p(\mathbf{Y}|\,E[\underset{\sim}{\tau^2}]\,,E[\underset{\sim}{\omega^2}]\,)$$

$$E[\tau_k^2] = \frac{m_k c+1}{\lambda_1^2} \;\;;\;\; E[\omega_i^2] = \frac{c+1}{\lambda_2^2}$$

Model Fitting: Studying the Marginal Likelihood

Marginal Likelihood Approximation

$$p(\mathbf{Y}|\,\lambda_1^2,\ \lambda_2^2) \approx$$

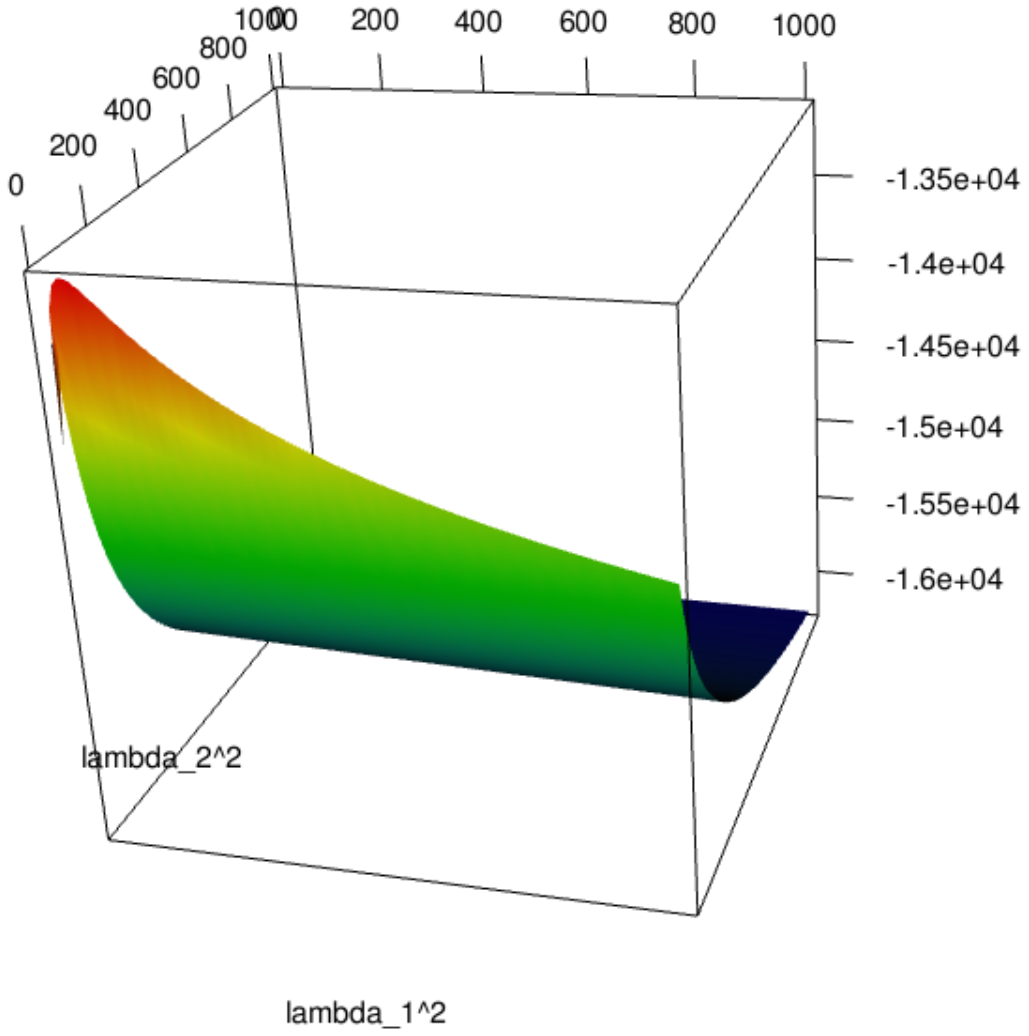
$$(2\pi)^{-\frac{nc}{2}}\, a_\sigma^{b\sigma}\, \frac{\Gamma(\frac{nc}{2}+a_\sigma)}{\Gamma(a_\sigma)} \times \left|(\mathrm{I_c}\otimes\mathbf{X})\left(\mathrm{I_c}\otimes \mathrm{Diag}\left\{\left(\frac{\lambda_2^2}{\mathrm{c}+1}+\frac{\lambda_1^2}{\mathrm{m_{k(i)}c}+1}\right)^{-1}\right\}\right)(\mathrm{I_c}\otimes\mathbf{X}^{\mathrm{T}})+\mathrm{I_{cn}}\ )\right|^{-\frac{1}{2}} \times$$

$$\left(b_\sigma+\frac{1}{2}\mathbf{Y}^T\left[(\mathrm{I_c}\otimes\mathbf{X})\left(\mathrm{I_c}\otimes \mathrm{Diag}\left\{\left(\frac{\lambda_2^2}{\mathrm{c}+1}+\frac{\lambda_1^2}{\mathrm{m_{k(i)}c}+1}\right)^{-1}\right\}\right)(\mathrm{I_c}\otimes\mathbf{X}^{\mathrm{T}})+\mathrm{I_{cn}}\ )\right]^{-1}\mathbf{Y}\right)^{-(\frac{nc}{2}+a_\sigma)}$$

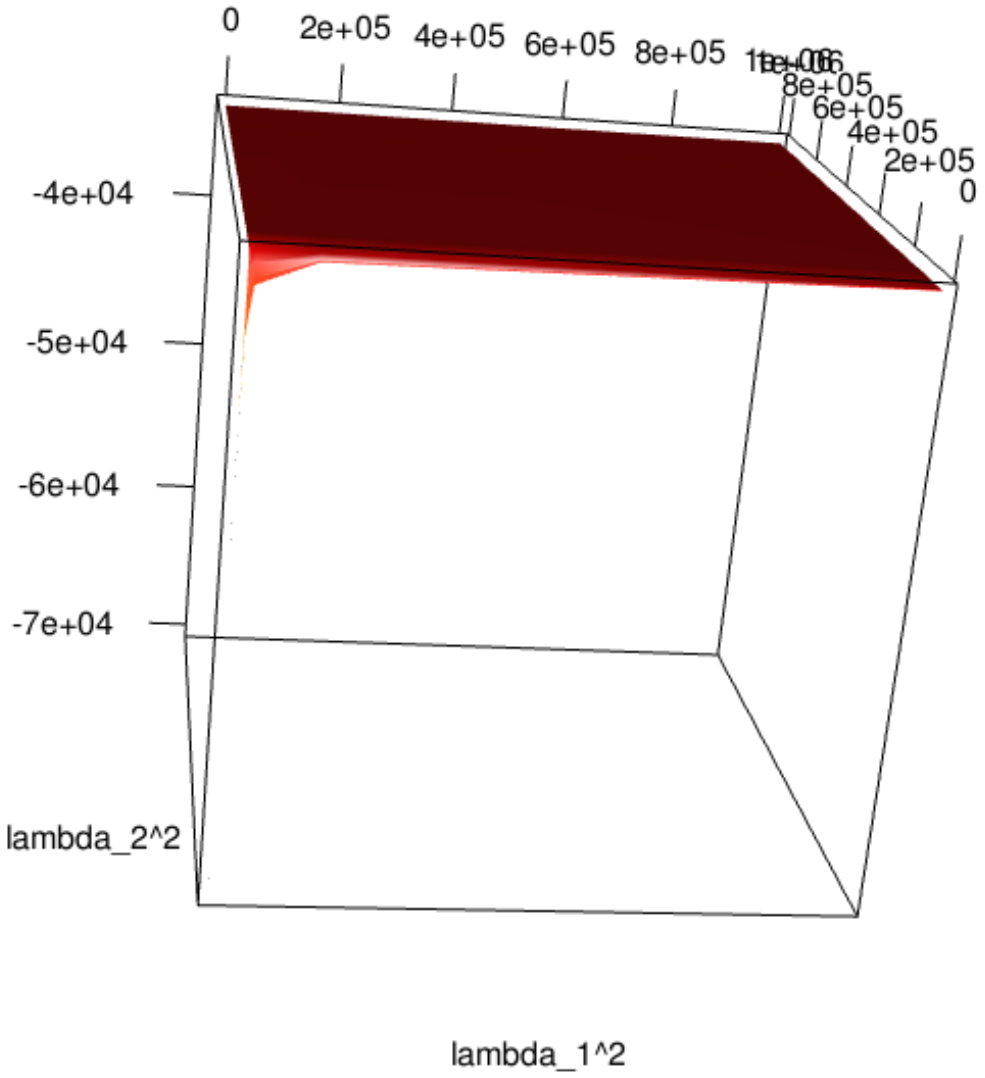
- The approximation is evaluated over a grid of $(\lambda_1^2, \lambda_2^2)$ values for different sets of simulated data.

'Nicely Behaved' Marginal Likelihood Approximaiton

- simulated data:
  $d = 200; c = 5; n = 500$
- maximum point at:
  $\lambda_1^2 = 30.4; \lambda_2^2 = 0.1$
- Gibbs Sampler performs well with fixed $\hat{\lambda}_1^2$ and $\hat{\lambda}_2^2$.



'Poorly Behaved' Marginal Likelihood Approximation



- simulated data:
  $d = 486; c = 12; n = 632$
- data simulated with weak signals
- maximum point at:
  $\lambda_1^2 = 10^5; \lambda_2^2 = 10^4$
- Gibbs Sampler leads to heavy overshrinking with fixed $\hat{\lambda}_1^2$ and $\hat{\lambda}_2^2$.

Model Fitting: Cross-Validation for λ_1^2, λ_2^2?

- For Bayesian lasso and related hierarchical models Park and Casella (2008) and Kyung et al. (2010) found that '*putting* λ *into the Gibbs sampler seems as effective as choosing it by cross-validation*'.
- For the model we have developed, under certain settings (number of SNPs large, weak effects), we find empirically that cross-validation avoids some of the observed problems with FB and MML choice of the tuning parameters.
- Combining Gibbs sampling with CV over a 2-D grid of tuning parameters is computationally intensive.

Model Fitting: WAIC

- We use WAIC (Watanabee, 2010) which does not require any data splitting for its computation and can be viewed as an approximation to leave-one-out cross-validation (Gelman, Hwang and Vehtari, 2013).

$$WAIC = -2\sum_{l=1}^{n} \log E_{\mathbf{W},\sigma^2}[p(\mathbf{y}_\ell \mid \mathbf{W}, \sigma^2)|\mathbf{y}_1, \ldots, \mathbf{y}_n]$$

$$+2\sum_{l=1}^{n} V_{\mathbf{W},\sigma^2}[\log p(\mathbf{y}_\ell \mid \mathbf{W}, \sigma^2)|\mathbf{y}_1, \ldots, \mathbf{y}_n]$$

- We run Gibbs samplers in parallel over a 2D grid for λ_1^2, λ_2^2 and choose the tuning parameters minimizing WAIC.

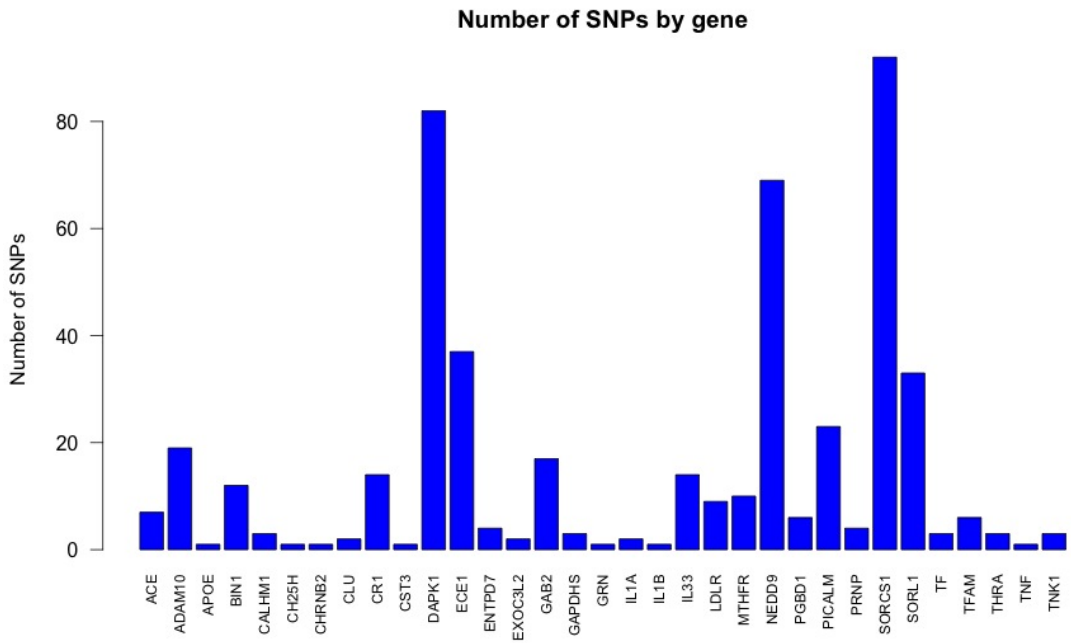
Outline

Simulation Study: The Data

Genetic Data

The SNP covariates used for data simulation come from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

We include genetic data on 632 subjects over 486 SNPs belonging to 33 different genes.



Simulation Study: The Data

True **W** Structure

A **W** matrix is simulated from its prior distribution with the following settings.

- number of SNPs $(d) = 486$
- SNPs are partitioned into 33 (K) genes
- number of phenotypes $(c) = 12$
- $\sigma^2 = \lambda_1^2 = \lambda_2^2 = 2$

Sparsity is introduced to **W** by setting all but 50 rows to zero.
Only the following rows are left at their simulated values.

- rows corresponding to 5 genes of SNP sizes 14, 10, 6, 4, 1 (35 SNPs)
- rows corresponding to 15 other SNPs

Simulation Study: Methodology

The genetic data and sparse **W** matrix are used to simulate 100 sets of response variables. We apply the Wang et al. method and our Gibbs-WAIC Bayesian method to each of the 100 datasets.

Wang et al. model fitting

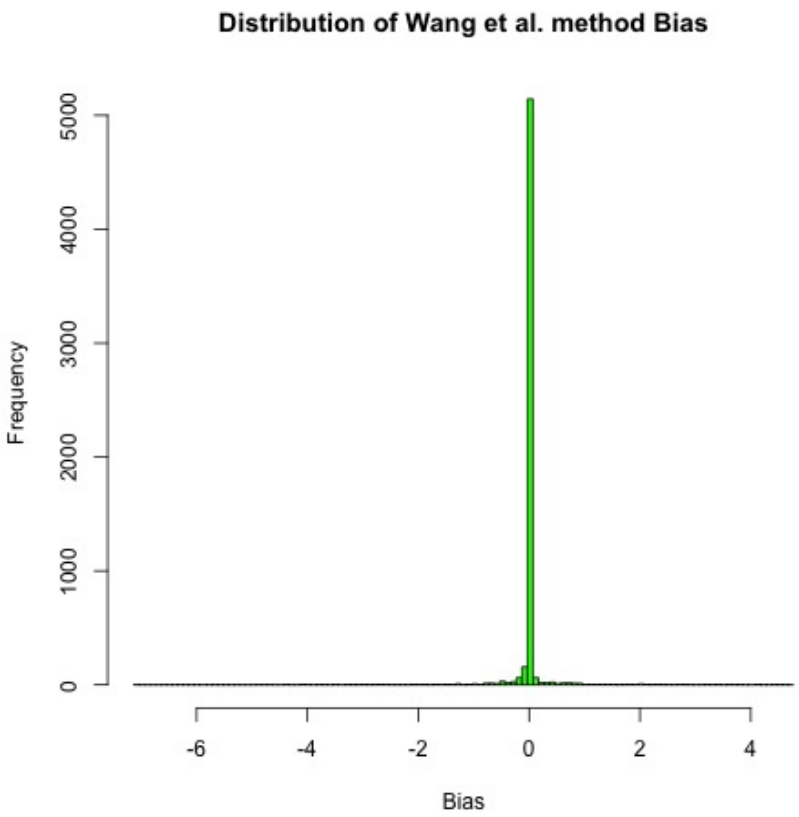
Tuning parameters, γ_1 and γ_2, are chosen via 5-fold cross-validation in the range of $(10^{-5}, 10^{-4}, \ldots, 10^4, 10^5)$.

Bayesian model fitting

The model is fit with fixed λ_1^2, λ_2^2 values in the range of $(0.01, 0.1, 1, 10, 100)$ for a total of 25 mcmc runs in each dataset. The model with the minimum WAIC is selected.

Simulation Study: Estimator Bias

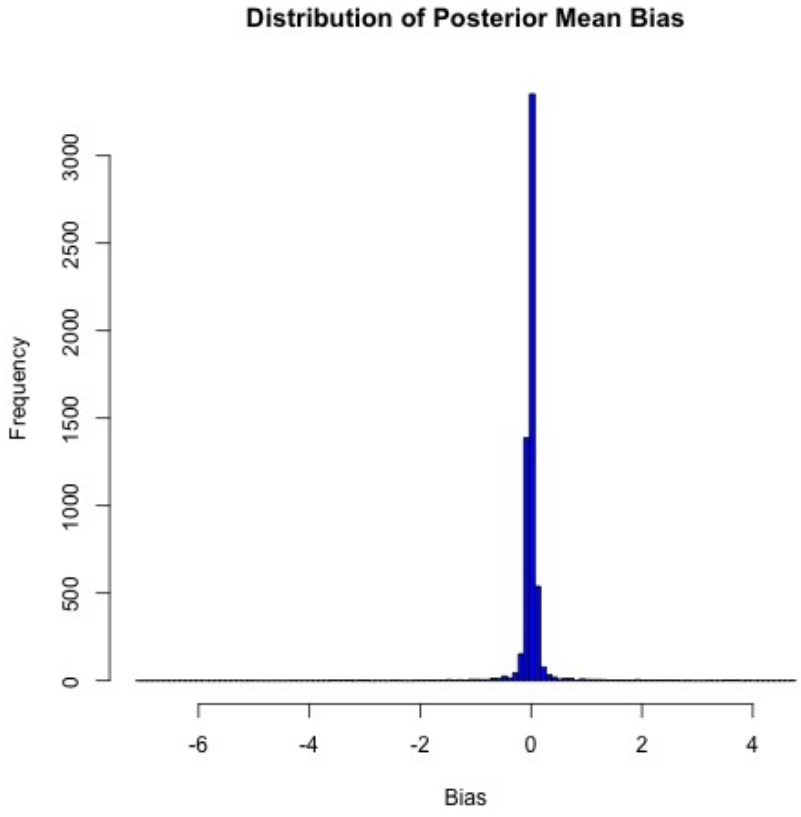
Wang et al. Estimator Bias



mean absolute bias $= 0.0905$

bias range $= [-7.13, 4.63]$

Posterior Mean Bias

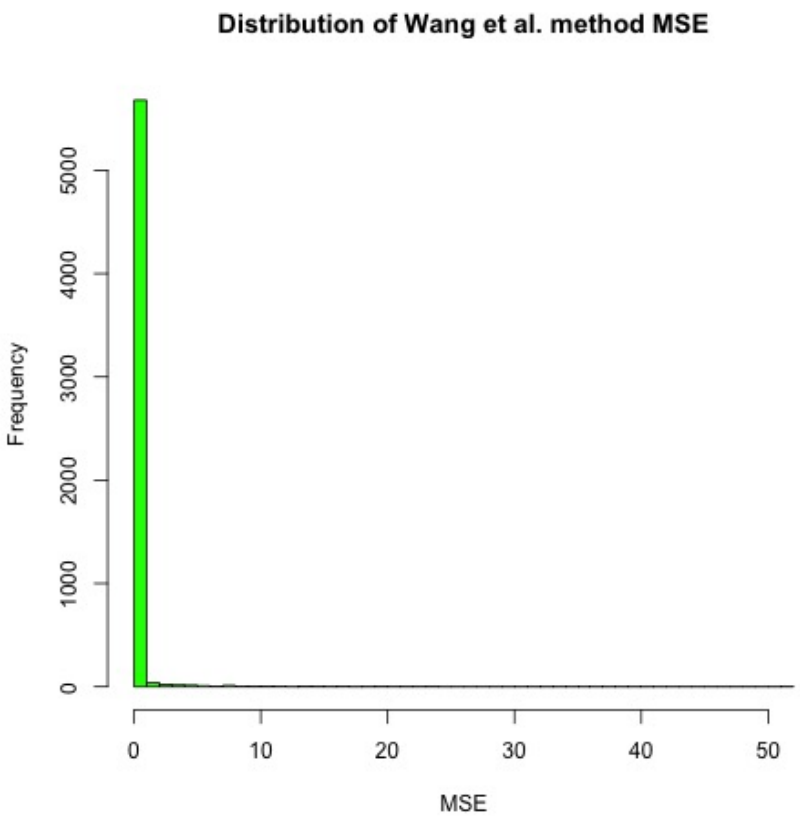


mean absolute bias $= 0.0992$

bias range $= [-5.07, 4.77]$

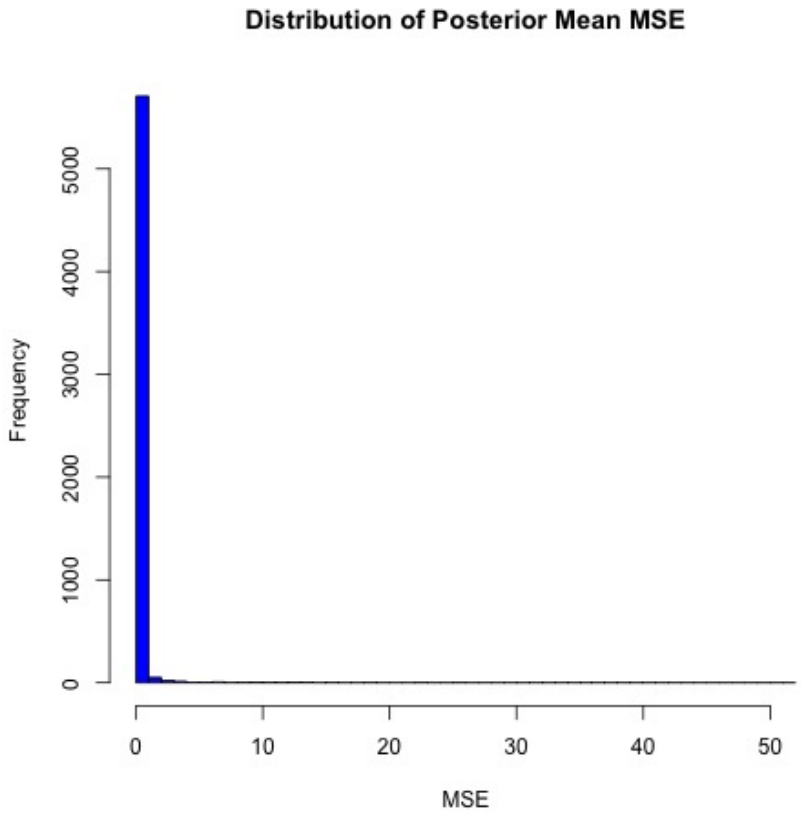
Simulation Study: Estimator MSE

Wang et al. Estimator MSE



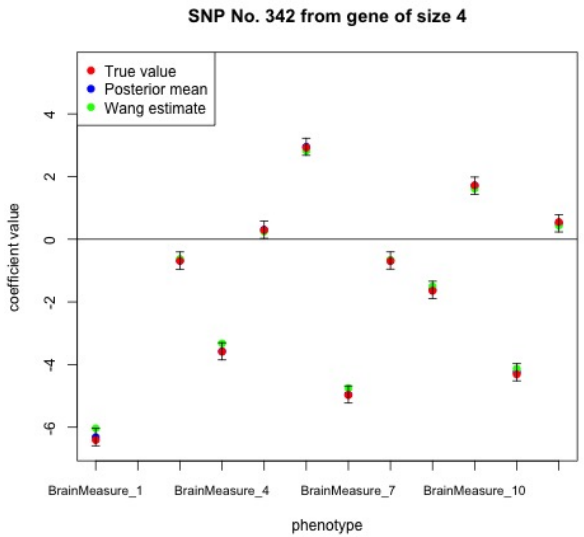
mean MSE = 0.1796

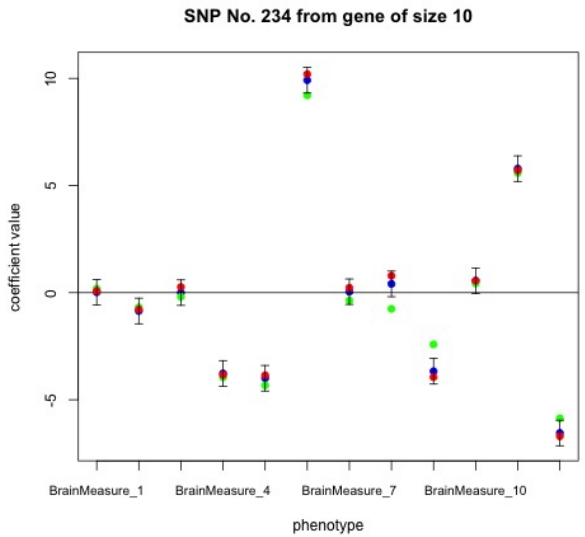
Posterior Mean MSE

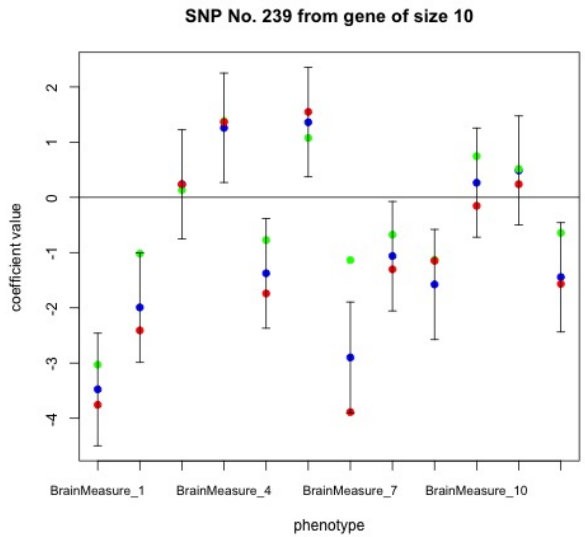


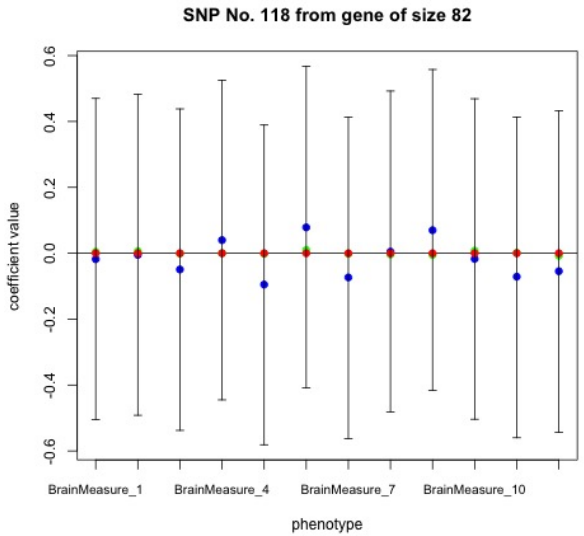
mean MSE = 0.1978

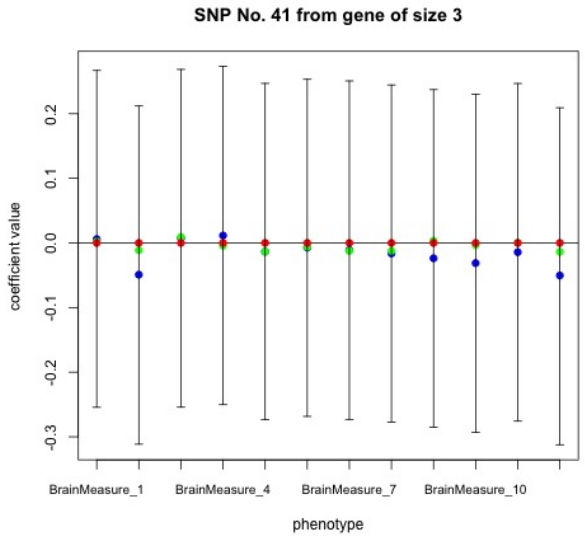
Simulation Study: Sample 95% CI Coverage Probabilities

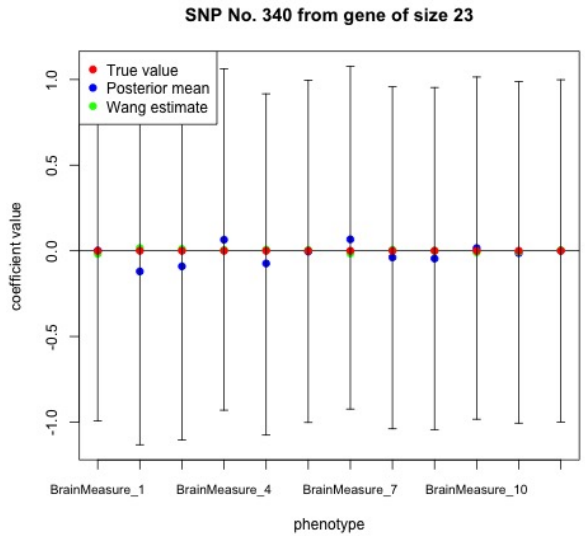






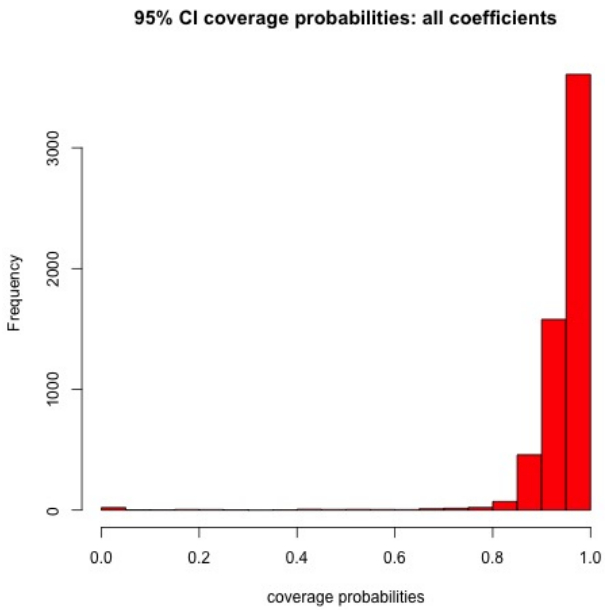






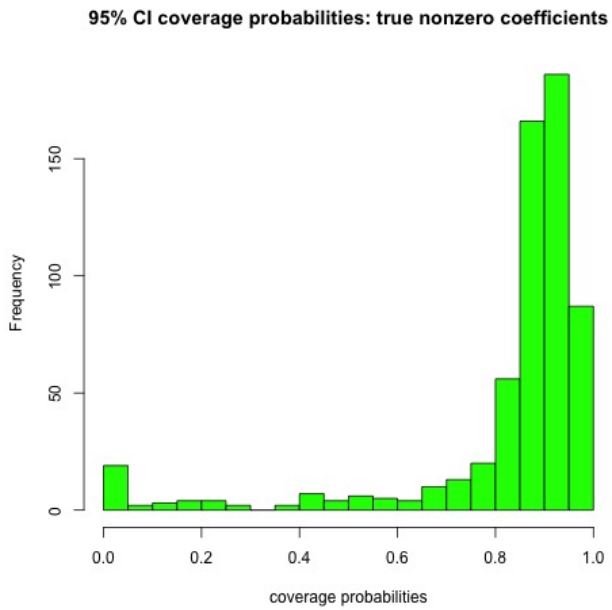
Simulation Study: 95% CI Coverage Probability Summaries

All Coefficients



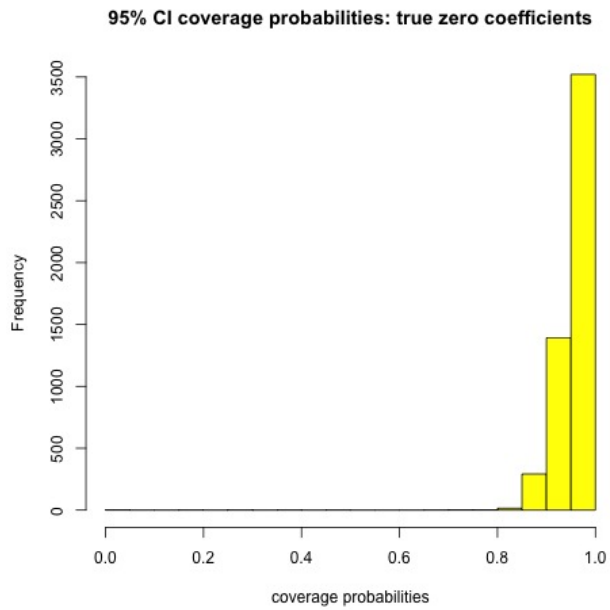
mean coverage probability = 95.18%

True Non-Zero Coefficients



mean coverage probability = 83.23%

True Zero Coefficients



mean coverage probability = 96.54%

The Bayesian intervals seem to have reasonably adequate frequentist coverage.

ADNI Data Application: The Data

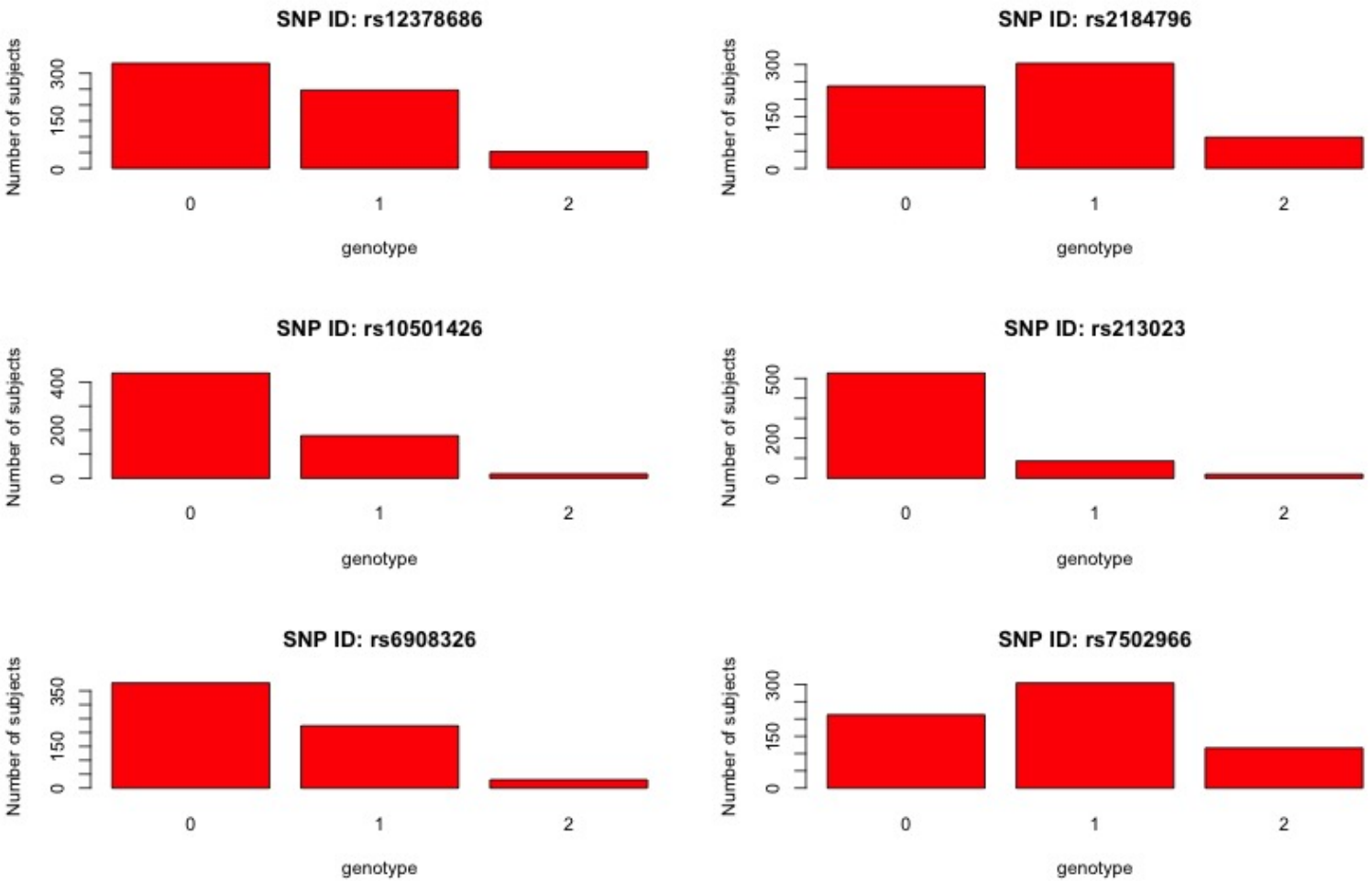
- Both genetic and structural MRI data used in this project were obtained from the Alzheimer's Disease Neuroimaging Initiative 1 (ADNI-1) database.
- Data has been collected and processed to be similar to the data presented and analysed by Wang et al. [2012].
- We include genetic and brain measurement data on 632 subjects.

ADNI Data Application: Genetic Data

- Among all SNPs, Wang et al. [2012] only include SNPs belonging to the top 40 Alzheimer's Disease (AD) candidate genes listed on the AlzGene database as of June 10, 2010.
- Data presented here are queried from the most recent genome build as of December 2014, from ADNI-1 genomic data.
- After quality control and imputation steps, the genetic data used in this study includes 486 SNPs from 33 genes.

ADNI Data Application: Genetic Data

Figure : Example of SNP counts included in the dataset



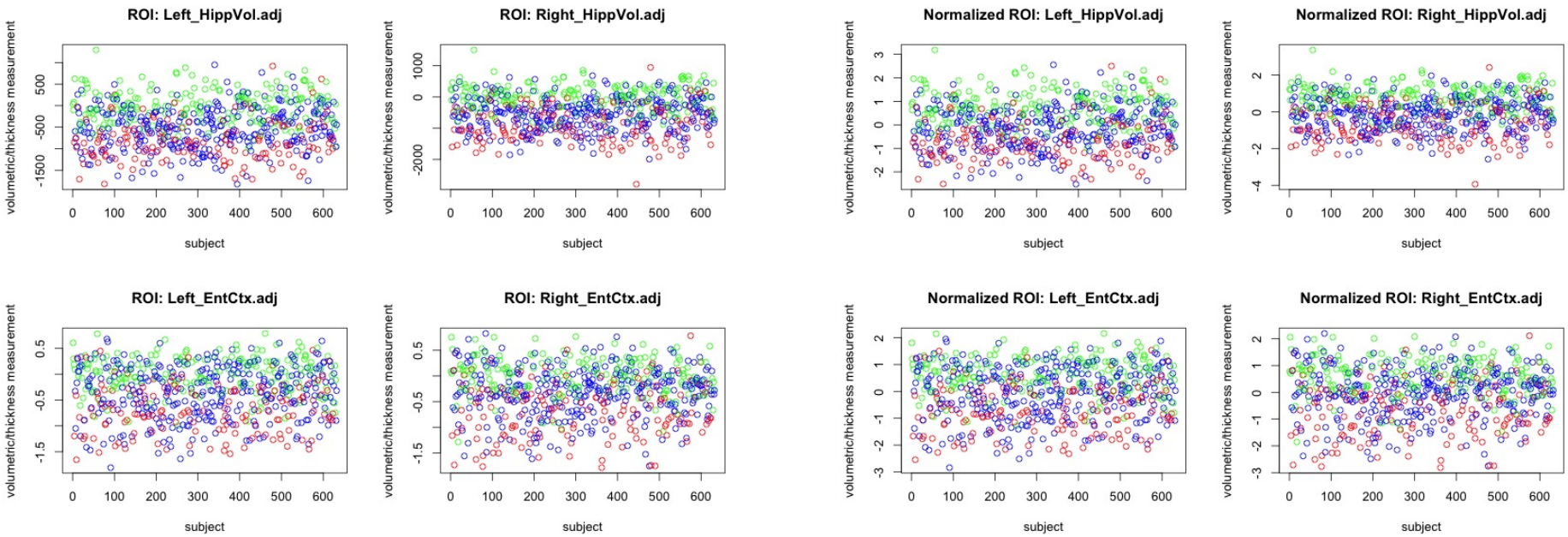
ADNI Data Application: MRI Data

FreeSurfer measurements define volumetric and cortical thickness values. A subset of FreeSurfer measures from 12 regions of interest are selected to be included for identifying significant SNPs.

ID	Region of Interest (ROI)
Left_HippVol Right_HippVol	volume of hippocampus
Left_EntCtx Left_Parahipp Right_EntCtx Right_Parahipp	thickness of entorhinal cortex and thickness of parahippocampal gyrus
Left_Precuneus Right_Precuneus	thickness of precuneus
Left_MeanFront Right_MeanFront	mean thickness of caudal midfrontal, rostral midfrontal, superior frontal, lateral orbitofrontal, and medial orbitofrontal gyri and frontal pole
Left_MeanLatTemp Right_MeanLatTemp	Mean thickness of inferior temporal, middle temporal, and superior temporal gyri

Wang et al. [2012] include these ROIs in their study based on knowledge that they are related to Alzheimer's Disease. MRI measures are adjusted for age, gender, education, handedness, and baseline total intracranial volume (ICV) based on regression weights from healthy controls.

ADNI Data Application: MRI Data



- FreeSurfer measures are scaled and centered prior to fitting the models.
- The figure on the left depicts adjusted measurements from 4 regions of interest prior to being scaled and centered; the figure on the right afterwards.
- Colours represent the disease status of subjects. (Green = CN ; Blue = LMCI ; Red = AD)

ADNI Data Application: Methodology

We apply the Wang et al. method and our Gibbs-WAIC Bayesian method to the data.

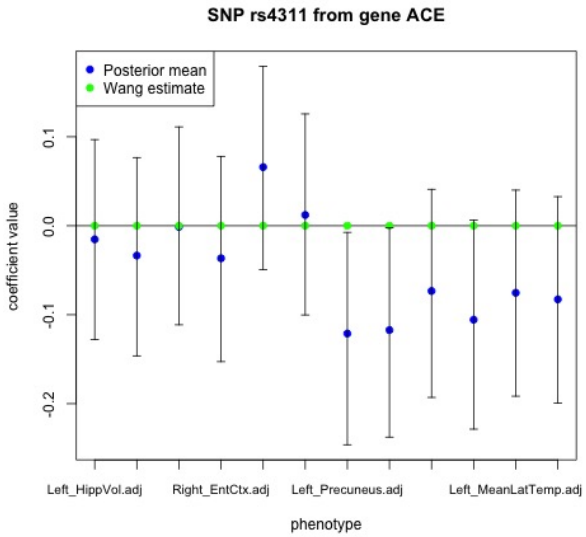
Wang et al. model fitting and SNP selection

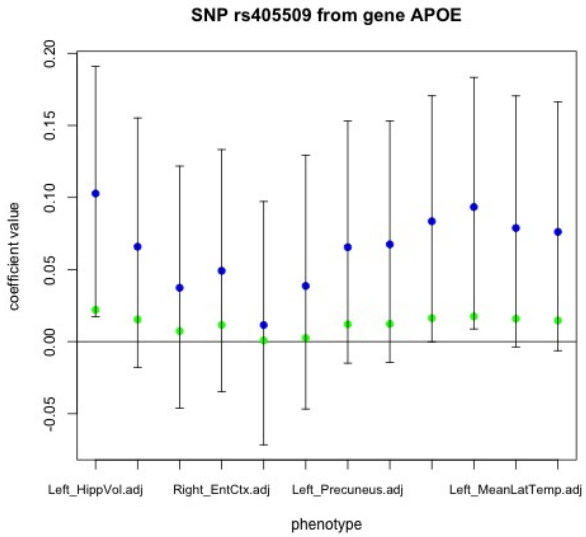
- Tuning parameters, γ_1 and γ_2, are chosen via 5-fold cross-validation in the range of $(10^{-5}, 10^{-4}, \ldots, 10^{4}, 10^{5})$.
- Wang et al. assign weights to each SNP by summing the absolute values of the estimated coefficients of a single SNP over all phenotypes.
- SNPs are ranked based on their weights.

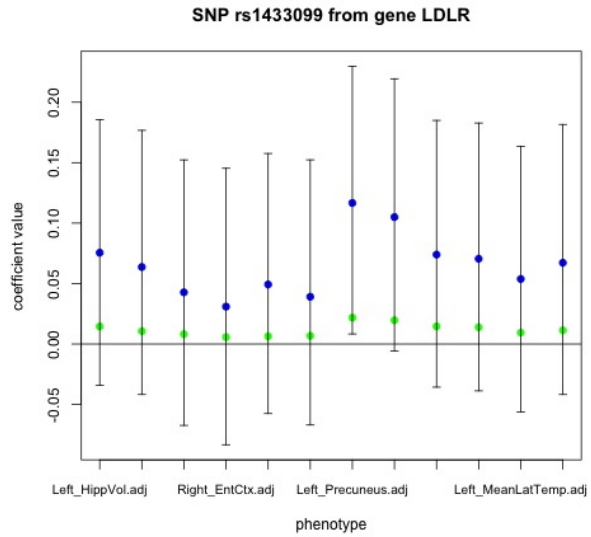
Bayesian model fitting and SNP selection

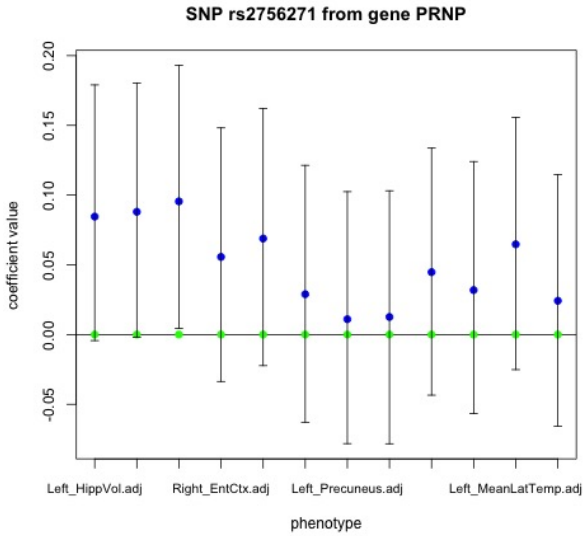
- The model is fit with fixed λ_1^2, λ_2^2 values in the range of $(10^{-3}, 10^{-2}, \ldots, 10^{2}, 10^{3})$ for a total of 49 mcmc runs. The model with the minimum WAIC is selected.
- There are a total of 5 SNPs that have 95% CI's that do not contain zero.

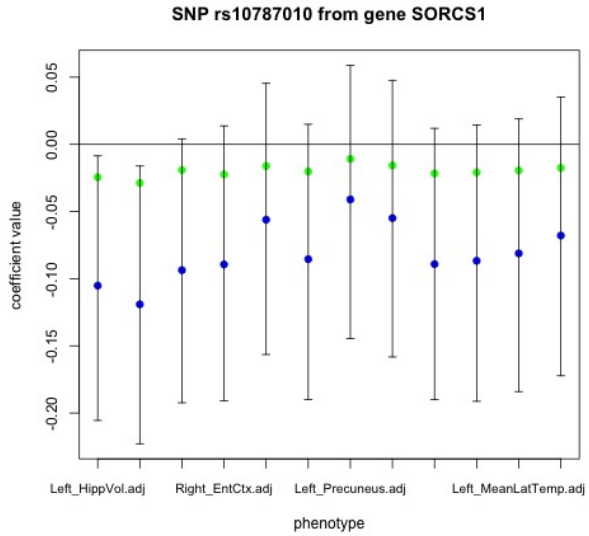
ADNI Data Application: Bayesian Model Selected SNPs



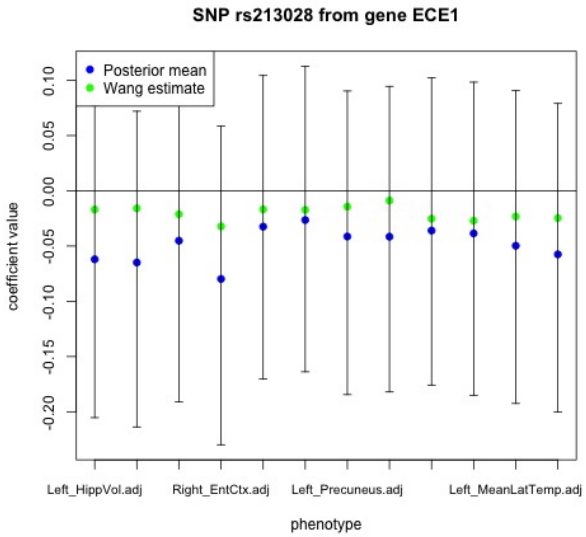


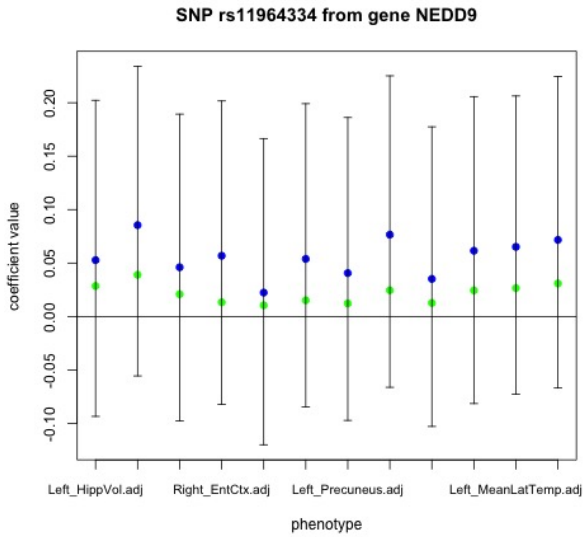


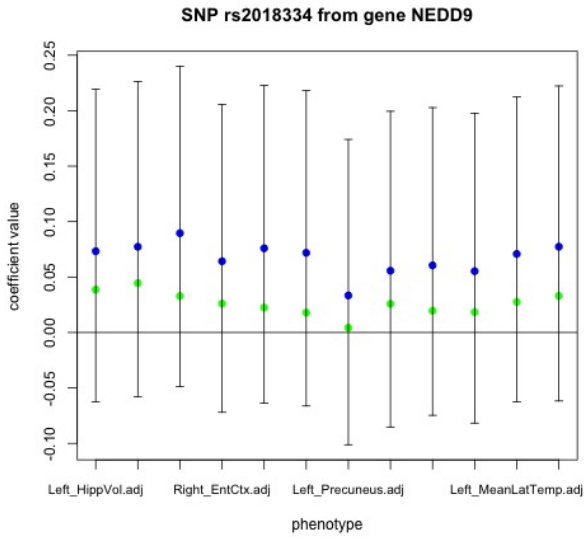


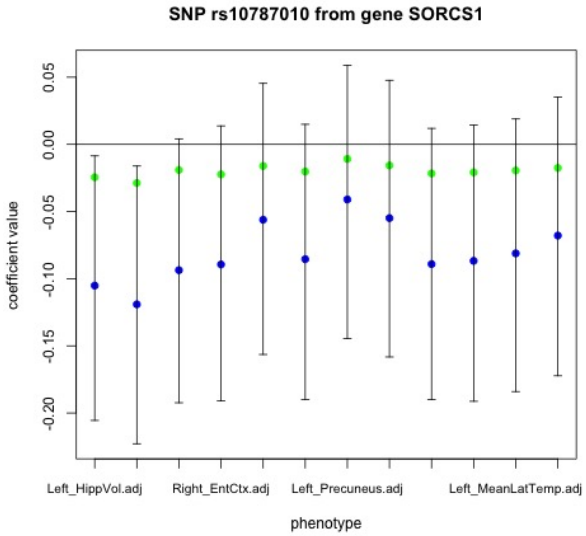


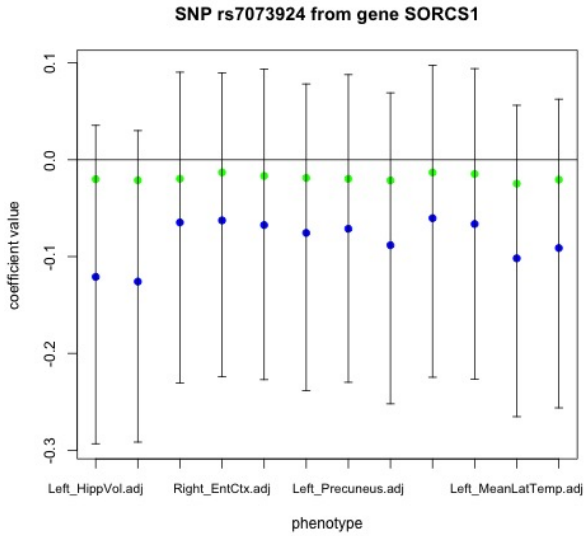
ADNI Data Application: Top 5 Wang et al. ranked SNPs











Outline

Discussion

- We use a hierarchical Bayes representation of the estimator proposed by Wang et al. [2012] and develop a Gibbs sampling approach for obtaining interval estimates.
- Interval estimates seem to have reasonable coverage probabilities for the settings considered.
- Extending numerical studies to compare with (i) non-parametric bootstrap and (ii) Bayesian approach using spike-and-slab priors.
- There are some obvious model improvements to be considered.
- Tuning parameters: comparison of hierarchical Bayes, empirical Bayes, and cross-validation yields unexpected results.

References

Minjung Kyung, Jeff Gill, Malay Ghosh, and George Casella. Penalized regression, standard errors, and bayesian lassos. *Bayesian Analysis*, 5(2):369–411, 2010.

Trevor Park and George Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

Hua Wang, Feiping Nie, Heng Huang, Sungeun Kim, Kwangsik Nho, Shannon L Risacher, Andrew J Saykin, Li Shen, et al. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the adni cohort. *Bioinformatics*, 28(2):229–237, 2012.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.