Kelvin Niu

Homework 8

<u>Setup</u>

$y_i \sim N(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{100} x_{i(100)}, \sigma^2)$
$\sigma^2 = 1$
$\beta_1, \dots, \beta_5 \sim N(0, 2)$
$\beta_6, \dots, \beta_{100} = 0$
$x_i \sim N_{100}(O, I)$
$n = 100$

For this simulation, I set 5 non-zero betas $\beta_1, \dots, \beta_5$ [each drawn from $N(0, 2)$] and 95 betas $\beta_6, \dots, \beta_{100}$ equal to zero. I then simulated 100 x_i's each with its predictors drawn independently from a standard normal $N(0, 1)$. Finally I calculated the values of y_i using the beta and x_i values plus noise distributed $N(0, 1)$.

<u>Implementation of different methods</u>

Fully Bayes ridge: I set the priors for τ and σ^{-2} to be $\tau \sim Ga(0.5, 0.5)$ and $\sigma^{-2} \sim Ga(0.001, 0.001)$. Using the full conditionals (derived in the Appendix), I performed 1000 iterations of Gibbs sampling with the initial value of the betas set to the true values of the betas. After examining the trace plots, no burn-ins were needed, perhaps because my initial values of the betas were the true values.

***t*-prior with $\boldsymbol{\nu = 1}$**: Similarly, I set the priors for τ and σ^{-2} to be $\tau \sim Ga(0.5, 0.5)$ and $\sigma^{-2} \sim Ga(0.001, 0.001)$. Using the full conditionals (derived in the Appendix), I performed 1000 iterations of Gibbs sampling with the initial value of the betas set to the true values of the betas. After examining the trace plots, no burn-ins were needed, perhaps because my initial values of the betas were the true values.

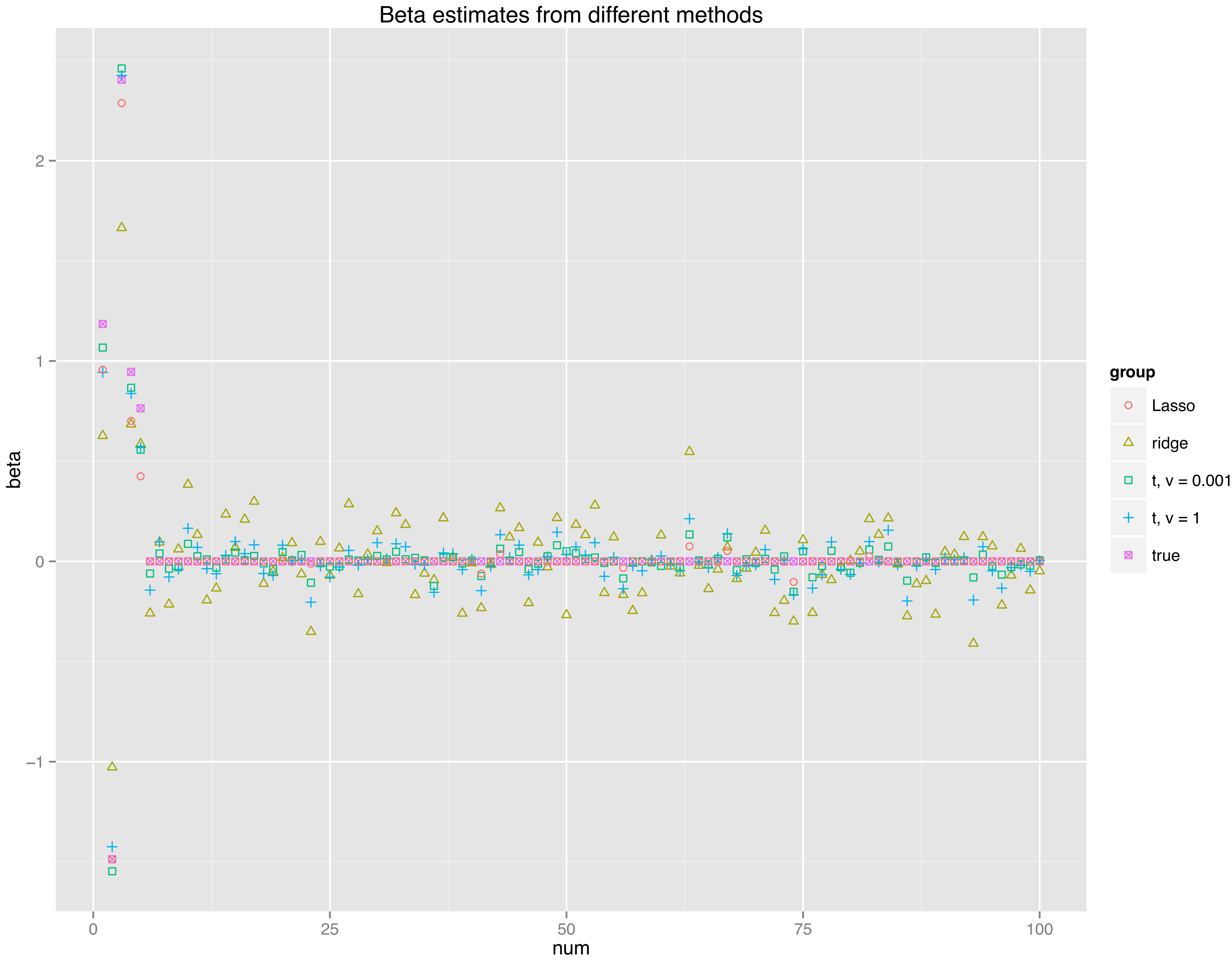
***t*-prior with $\boldsymbol{\nu = 0.001}$**: Similarly, I set the priors for τ and σ^{-2} to be $\tau \sim Ga(0.5, 0.5)$ and $\sigma^{-2} \sim Ga(0.001, 0.001)$. Using the full conditionals (derived in the Appendix), I performed 1000 iterations of Gibbs sampling with the initial value of the betas set to the true values of the betas. After examining the trace plots, no burn-ins were needed, perhaps because my initial values of the betas were the true values.

Lasso: I utilized the glmnet package in R to get Lasso estimates of the betas.

<u>Comparison among different methods</u>

The following is a plot comparing the beta estimates among the different methods.

- The fully Bayes ridge method appears to pull the estimates of the true (non-zero) betas too close to 0 while pulling the estimates of the fake betas too much out from 0
- The t-prior methods appears to estimate the true (non-zero) betas slightly better than Lasso, but Lasso estimates the fake betas better



Beta estimates from different methods
beta
num
2
1
0
−1
0
25
50
75
100
group
Lasso
ridge
t, v = 0.001
t, v = 1
true

For each method, I calculated the MSE on the β's and the predictive MSE on a newly generated test dataset of 100 observations. The results are summarized in the table below.

Table: Comparison of MSE's among different methods for estimating β's

	MSE on β's	Predictive MSE
Fully Bayes ridge	0.0411	4.4731
t-prior with $\nu = 1$	0.0074	1.7688
t-prior with $\nu = 0.001$	0.0027	1.2153
Lasso	0.0026	1.2270

- The fully Bayes ridge regression performs the worst in terms of both estimating β's and predicting new observations
- The t-prior with the smaller ν value performs better than the t-prior with the larger ν value in terms of both MSE on β's and predictive MSE
- The Bayesian method of using a t-prior with $\nu = 0.001$ performs about the same as the frequentist Lasso method in terms of both estimating β's and predicting new observations. Both of these methods do very well compared to the others.

Kelvin Niu

Fully Bayes ridge

$$y \sim N_{100}(X\beta, \sigma^2 I)$$

$$\beta \sim N_{100}(0, \tau^{-1}\sigma^2 I)$$

$$\tau \sim Ga(a, b)$$

$$\sigma^{-2} \sim Ga(c, d)$$

$$\pi(\beta, \tau, \sigma^{-2} | X, y) = L(y | X, \beta, \sigma^{-2})\, \pi(\beta | \tau, \sigma^{-2})\, \pi(\tau)\, \pi(\sigma^{-2})$$

$$\propto |\sigma^2 I|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(y - X\beta)^T(\sigma^2 I)^{-1}(y - X\beta)\right] |\tau^{-1}\sigma^2 I|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\beta^T(\tau^{-1}\sigma^2 I)^{-1}\beta\right] \tau^{a-1} e^{-b\tau} (\sigma^{-2})^{c-1} e^{-d\sigma^{-2}}$$

$$\propto (\sigma^2)^{-50} \exp\left[-\tfrac{1}{2\sigma^2}(y^Ty - y^TX\beta - \beta^TX^Ty + \beta^TX^TX\beta) - \tfrac{\tau}{2\sigma^2}\beta^T\beta\right] (\tau^{-1}\sigma^2)^{-50} \tau^{a-1} e^{-b\tau} (\sigma^{-2})^{c-1} e^{-d\sigma^{-2}}$$

$$\propto (\sigma^{-2})^{c+99} e^{-d\sigma^{-2}} \tau^{a+49} e^{-b\tau} \exp\left[-\tfrac{1}{2\sigma^2}(y^Ty - 2\beta^TX^Ty + \beta^TX^TX\beta + \tau\beta^T\beta)\right]$$

$$\pi(\beta | \tau, \sigma^{-2}, X, y) \propto \exp\left[-\tfrac{1}{2\sigma^2}(-2\beta^TX^Ty + \beta^TX^TX\beta + \tau\beta^T\beta)\right]$$

$$\propto \exp\left[-\tfrac{1}{2\sigma^2}\left[\beta^T(X^TX + \tau I)\beta - 2\beta^T(X^Ty)\right]\right]$$

$$\boxed{\beta | \tau, \sigma^{-2}, X, y \sim N_{100}\left((X^TX + \tau I)^{-1}(X^Ty), (X^TX + \tau I)^{-1}\sigma^2\right)}$$

$$\pi(\tau | \beta, \sigma^{-2}, X, y) \propto \tau^{a+49} \exp\left[-\tau(b + \tfrac{1}{2}\sigma^{-2}\beta^T\beta)\right]$$

$$\boxed{\tau | \beta, \sigma^{-2}, X, y \sim Ga(a + 50, b + \tfrac{1}{2}\sigma^{-2}\beta^T\beta)}$$

$$\pi(\sigma^{-2} | \beta, \tau, X, y) \propto (\sigma^{-2})^{c+99} \exp\left[-\sigma^{-2}(d + \tfrac{1}{2}(y - X\beta)^T(y - X\beta) + \tfrac{1}{2}\tau\beta^T\beta)\right]$$

$$\boxed{\sigma^{-2} | \beta, \tau, X, y \sim Ga(c + 100, d + \tfrac{1}{2}(y - X\beta)^T(y - X\beta) + \tfrac{1}{2}\tau\beta^T\beta)}$$

Kelvin Niu

t-prior

$$y \sim N_{100}(X\beta, \sigma^2 I)$$

$$\beta \sim N_{100}(0, \tau^{-1}L^{-1})$$ where $L = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_{100} \end{bmatrix}$

$$\lambda_j \sim Ga\left(\frac{v}{2}, \frac{v}{2}\right)$$

$$\tau \sim Ga(a, b)$$

$$\sigma^{-2} \sim Ga(c, d)$$

$$\pi(\beta, \lambda, \tau, \sigma^{-2} | x, y) = L(y|x, \beta, \sigma^{-2})\, \pi(\beta|\lambda,\tau)\, \pi(\lambda)\, \pi(\tau)\, \pi(\sigma^{-2})$$

$$\propto |\sigma^2 I|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(y - X\beta)^T(\sigma^2 I)^{-1}(y - X\beta)\right] |\tau^{-1}L^{-1}|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\beta^T(\tau^{-1}L^{-1})^{-1}\beta\right]\left[\prod_{j=1}^{100} \lambda_j^{\frac{v}{2}-1} e^{-\frac{v}{2}\lambda_j}\right] \tau^{a-1}e^{-b\tau}(\sigma^{-2})^{c-1}e^{-d\sigma^{-2}}$$

$$\propto (\sigma^2)^{-50} \exp\left[-\tfrac{1}{2\sigma^2}(y^Ty - y^TX\beta - \beta^TX^Ty + \beta^TX^TX\beta) - \tfrac{\tau}{2}\beta^TL\beta\right] \tau^{50}\left[\prod_{j=1}^{100}\lambda_j^{\frac{1}{2}}\right]\left[\prod_{j=1}^{100}\lambda_j^{\frac{v}{2}-1}\right]\exp\left[-\tfrac{v}{2}\sum_{j=1}^{100}\lambda_j\right]\tau^{a-1}e^{-b\tau}(\sigma^{-2})^{c-1}e^{-d\sigma^{-2}}$$

$$\propto (\sigma^{-2})^{c+49}e^{-d\sigma^{-2}}\tau^{a+49}e^{-b\tau}\left[\prod_{j=1}^{100}\lambda_j^{\frac{v}{2}-\frac{1}{2}}\right]\exp\left[-\tfrac{v}{2}\sum_{j=1}^{100}\lambda_j\right]\exp\left[-\tfrac{1}{2\sigma^2}(y^Ty - 2\beta^TX^Ty + \beta^TX^TX\beta) - \tfrac{\tau}{2}\beta^TL\beta\right]$$

$$\pi(\beta | \lambda, \tau, \sigma^{-2}, x, y) \propto \exp\left[-\tfrac{1}{2}(-2\sigma^{-2}\beta^TX^Ty + \sigma^{-2}\beta^TX^TX\beta + \tau\beta^TL\beta)\right]$$

$$\propto \exp\left[-\tfrac{1}{2}\left[\beta^T(\sigma^{-2}X^TX + \tau L)\beta - 2\beta^T(\sigma^{-2}X^Ty)\right]\right]$$

$$\boxed{\beta | \lambda, \tau, \sigma^{-2}, x, y \sim N_{100}\left((\sigma^{-2}X^TX + \tau L)^{-1}(\sigma^{-2}X^Ty),\ (\sigma^{-2}X^TX + \tau L)^{-1}\right) \quad \text{where } L = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_{100} \end{bmatrix}}$$

$$\pi(\lambda_j | \beta, \lambda_{(-j)}, \tau, \sigma^{-2}, x, y) \propto \lambda_j^{\frac{v}{2}-\frac{1}{2}} \exp\left[-\lambda_j \tfrac{1}{2}(\tau\beta_j^2 + v)\right]$$

$$\boxed{\lambda_j | \beta, \lambda_{(-j)}, \tau, \sigma^{-2}, x, y \sim Ga\left(\frac{v+1}{2}, \tfrac{1}{2}(\tau\beta_j^2 + v)\right)}$$

$$\pi(\tau | \beta, \lambda, \sigma^{-2}, x, y) \propto \tau^{a+49}\exp\left[-\tau(b + \tfrac{1}{2}\beta^TL\beta)\right]$$

$$\boxed{\tau | \beta, \lambda, \sigma^{-2}, x, y \sim Ga(a+50, b + \tfrac{1}{2}\beta^TL\beta)}$$

$$\pi(\sigma^{-2} | \beta, \lambda, \tau, x, y) \propto (\sigma^{-2})^{c+49}\exp\left[-\sigma^{-2}(d + \tfrac{1}{2}(y - X\beta)^T(y - X\beta))\right]$$

$$\boxed{\sigma^{-2} | \beta, \lambda, \tau, x, y \sim Ga\left(c+50, d + \tfrac{1}{2}(y - X\beta)^T(y - X\beta)\right)}$$