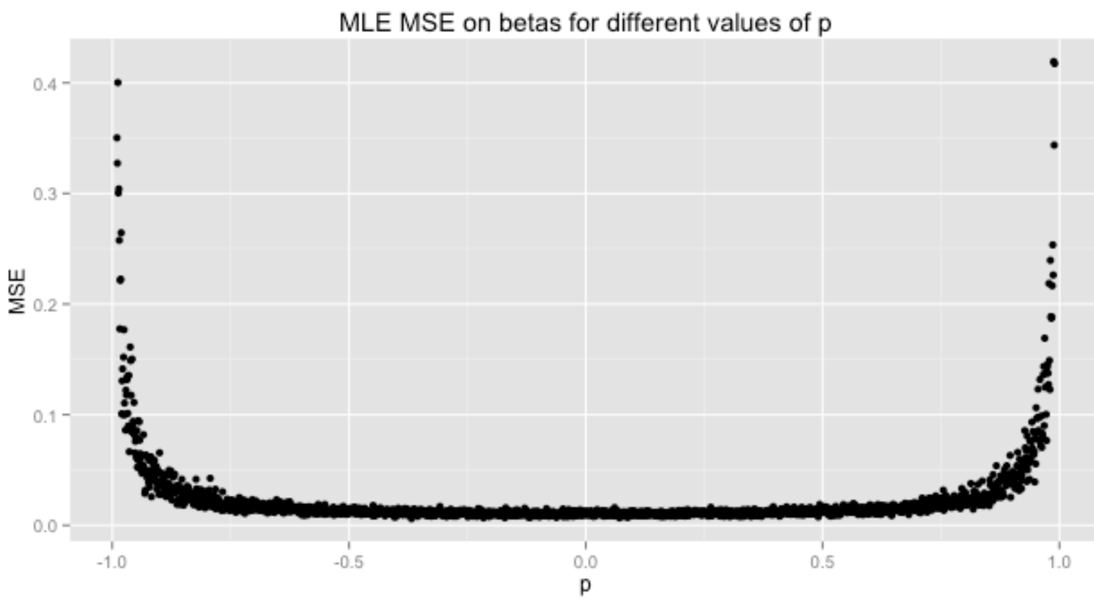
Kelvin Niu

Homework 7

$\begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix} \sim N_2\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$
$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$
$e_i \sim N(0, \sigma^2)$
$n = 100$

<u>MLE estimations of β's</u>
I first selected various values of ρ from -1 to 1. For each value of ρ, I simulated 30 datasets (x, y), calculated the MSE on the β's, and averaged the MSE across the datasets. Below is a plot comparing the MSE for different values of ρ.



As the covariance between X_{i1} and X_{i2} increases in absolute value, the MSE on the betas appears to increase, thus the estimates of the β's get worse.

For the rest of the questions, I chose to simulate datasets with low covariance $\rho = 0.1$.

<u>Bayes ridge regression estimations of β's</u>
$\beta_0 \sim N(0, \tau^{-1}\sigma^2)$
$\sigma^{-2} \sim Ga(0.001, 0.001)$ [weak prior]

Hyper-prior on penalty
If we put a prior on $\tau \sim Ga(0.5, 0.5)$ and proceed to estimate the marginal posterior distribution of τ using Gibbs sampling with 10,000 iterations on the full conditionals (calculated in the Appendix), the below distribution is obtained.

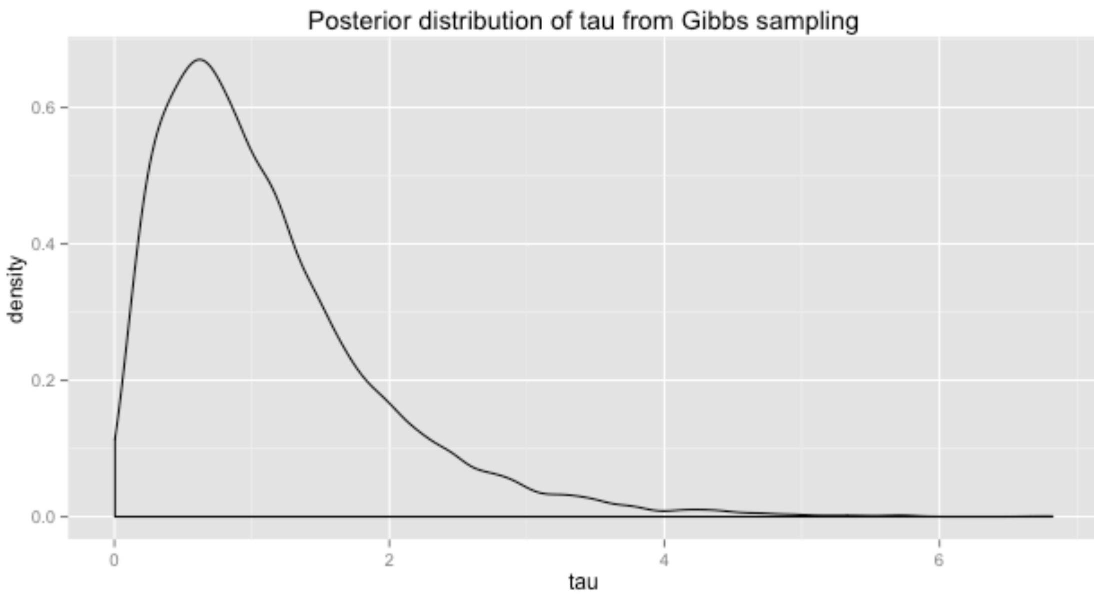


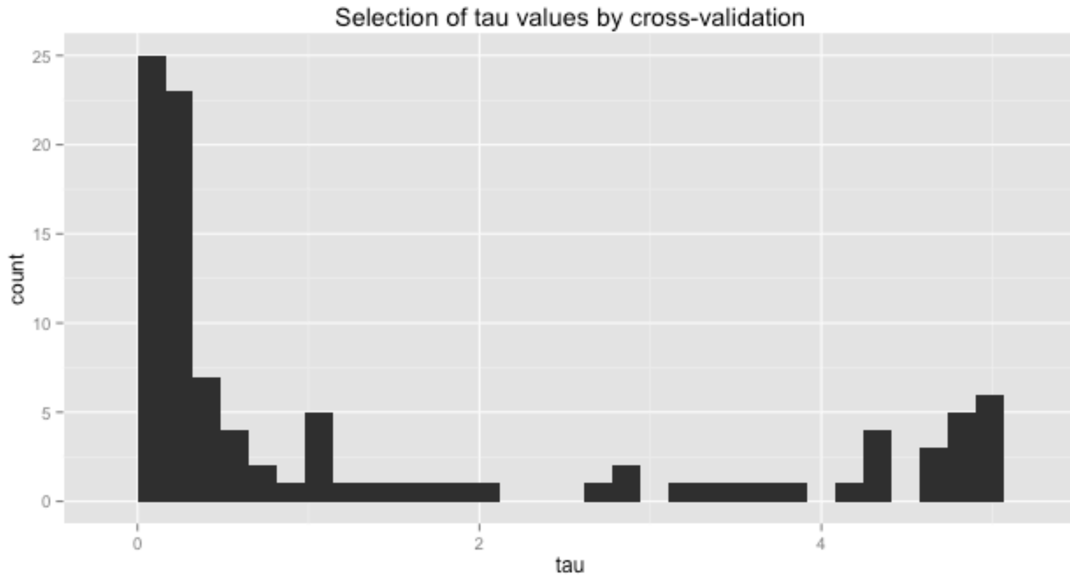
Table: Summary of posterior distribution of τ

	Mean	95% CI
τ	1.1024	[0.1310, 3.2168]

Choosing penalty by cross-validation

If instead we choose τ from a grid of values and use cross-validation to pick the τ that minimizes MSE on a test set (using Gibbs sampling with 1,000 iterations and the marginal posterior means of the β's as point estimates), the selected τ value is not consistent. I did a grid search of τ values from 0.1 to 5 (range based on distribution of τ obtained in previous part), using 90 samples as the training set and 10 samples as the test set. Running this grid search on 100 different datasets, the below distribution of selected τ values is obtained.

Next, I tried the same grid search with a different split using a bigger test set: 50 samples as the training set and 50 samples as the test set. The distribution of selected τ values is below.



The distribution of selected τ values starts to look more like the posterior distribution of τ values from the previous part. Perhaps with more data (greater n), the distribution of selected τ values will resemble the posterior distribution from the previous part.

Comparing estimations of β's using the different methods

To compare the three different methods above, I simulated a dataset, applied each method to the dataset, and recorded the MSE on the β's and the predictive MSE on a new test set of size 10.

Table: Comparison of MSE for the different methods

	MSE on β's	Predictive MSE
MLE	0.0020	0.9856
Bayes ridge regression: hyper-prior on penalty	0.0016	0.9808
Bayes ridge regression: penalty from CV	0.0020	0.8051

It appears that the MLE does the worst in terms of both MSE on β's and predictive MSE. Bayes ridge regression with the hyper-prior estimates the β's the best (in terms of MSE) but has a worse predictive MSE than Bayes ridge regression with cross-validation. Bayes ridge regression with cross-validation has the best predictive MSE, but estimates the β's worse than Bayes ridge regression with the hyper-prior (in terms of MSE). Thus, there appears to be a tradeoff between estimating the β's and predictive accuracy. However, because the MSE's were close and calculated on one dataset, the results may vary.

Kelvin Niu

Bayes ridge regression with hyperparameter on prior

$$y_i \sim N(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \sigma^2)$$

$$\beta_i \sim N(0, \tau^{-1}\sigma^2)$$

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

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

The joint posterior distribution of $\beta_0, \beta_1, \beta_2, \tau, \sigma^{-2}$ is

$$\pi(\beta_0, \beta_1, \beta_2, \tau, \sigma^{-2} \mid x, y) \propto L(y \mid x, \beta, \sigma^{-2})\, \pi(\beta_0 \mid \tau, \sigma^{-2})\, \pi(\beta_1 \mid \tau, \sigma^{-2})\, \pi(\beta_2 \mid \tau, \sigma^{2})\, \pi(\tau)\, \pi(\sigma^{-2})$$

$$\propto \left[\prod_{i=1}^{n} \sigma^{-1} \exp\left[-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}))^2\right]\right] (\tau^{-1}\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{\tau}{2\sigma^2}\beta_0^2\right) (\tau^{-1}\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{\tau}{2\sigma^2}\beta_1^2\right) (\tau^{-1}\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{\tau}{2\sigma^2}\beta_2^2\right) \tau^{a-1} e^{-b\tau} (\sigma^{-2})^{c-1} e^{-d\sigma^{-2}}$$

$$\propto \sigma^{-n-2c-1} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}))^2\right] \exp\left[-\frac{\tau}{2\sigma^2}(\beta_0^2 + \beta_1^2 + \beta_2^2)\right] \tau^{a+\frac{1}{2}} e^{-b\tau} e^{-d\sigma^{-2}}$$

We can find the full conditionals of $\beta_0, \beta_1, \beta_2, \tau, \sigma^{-2}$ by looking at the relevant terms in the joint posterior

$$\pi(\beta_0 \mid x, y, \beta_1, \beta_2, \tau, \sigma^{-2}) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2\beta_0 y_i + \beta_0^2 + 2\beta_0\beta_1 x_{i1} + 2\beta_0\beta_2 x_{i2})\right] \exp\left[-\frac{\tau}{2\sigma^2}\beta_0^2\right]$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\left(-2\beta_0\sum_{i=1}^{n} y_i + n\beta_0^2 + 2\beta_0\beta_1\sum_{i=1}^{n} x_{i1} + 2\beta_0\beta_2\sum_{i=1}^{n} x_{i2} + \tau\beta_0^2\right)\right]$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\left[(n+\tau)\beta_0^2 - 2\left(\sum_{i=1}^{n} y_i - \beta_1\sum_{i=1}^{n} x_{i1} - \beta_2\sum_{i=1}^{n} x_{i2}\right)\beta_0\right]\right]$$

$$\propto \exp\left[-\frac{n+\tau}{2\sigma^2}\left[\beta_0^2 - 2\frac{1}{n+\tau}\left(\sum_{i=1}^{n} y_i - \beta_1\sum_{i=1}^{n} x_{i1} - \beta_2\sum_{i=1}^{n} x_{i2}\right)\beta_0\right]\right]$$

$$\boxed{\beta_0 \mid x, y, \beta_1, \beta_2, \tau, \sigma^{-2} \sim N\left(\frac{1}{n+\tau}\left(\sum_{i=1}^{n} y_i - \beta_1\sum_{i=1}^{n} x_{i1} - \beta_2\sum_{i=1}^{n} x_{i2}\right), \frac{\sigma^2}{n+\tau}\right)}$$

$$\pi(\beta_1 \mid x, y, \beta_0, \beta_2, \tau, \sigma^{-2}) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2\beta_1 x_{i1} y_i + \beta_1^2 x_{i1}^2 + 2\beta_1\beta_0 x_{i1} + 2\beta_1\beta_2 x_{i1} x_{i2})\right] \exp\left[-\frac{\tau}{2\sigma^2}\beta_1^2\right]$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\left(-2\beta_1\sum_{i=1}^{n} x_{i1} y_i + \beta_1^2\sum_{i=1}^{n} x_{i1}^2 + 2\beta_1\beta_0\sum_{i=1}^{n} x_{i1} + 2\beta_1\beta_2\sum_{i=1}^{n} x_{i1} x_{i2} + \tau\beta_1^2\right)\right]$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\left[\left(\sum_{i=1}^{n} x_{i1}^2 + \tau\right)\beta_1^2 - 2\left(\sum_{i=1}^{n} x_{i1} y_i - \beta_0\sum_{i=1}^{n} x_{i1} - \beta_2\sum_{i=1}^{n} x_{i1} x_{i2}\right)\beta_1\right]\right]$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_{i1}^2 + \tau\right)\left[\beta_1^2 - 2\left(\sum_{i=1}^{n} x_{i1}^2 + \tau\right)^{-1}\left(\sum_{i=1}^{n} x_{i1} y_i - \beta_0\sum_{i=1}^{n} x_{i1} - \beta_2\sum_{i=1}^{n} x_{i1} x_{i2}\right)\beta_1\right]\right]$$

$$\boxed{\beta_1 \mid x, y, \beta_0, \beta_2, \tau, \sigma^{-2} \sim N\left(\left(\sum_{i=1}^{n} x_{i1}^2 + \tau\right)^{-1}\left(\sum_{i=1}^{n} x_{i1} y_i - \beta_0\sum_{i=1}^{n} x_{i1} - \beta_2\sum_{i=1}^{n} x_{i1} x_{i2}\right), \left(\sum_{i=1}^{n} x_{i1}^2 + \tau\right)^{-1}\sigma^2\right)}$$

By symmetry with β_1,

$$\boxed{\beta_2 \mid x, y, \beta_0, \beta_1, \tau, \sigma^{-2} \sim N\left(\left(\sum_{i=1}^{n} x_{i2}^2 + \tau\right)^{-1}\left(\sum_{i=1}^{n} x_{i2} y_i - \beta_0\sum_{i=1}^{n} x_{i2} - \beta_1\sum_{i=1}^{n} x_{i1} x_{i2}\right), \left(\sum_{i=1}^{n} x_{i2}^2 + \tau\right)^{-1}\sigma^2\right)}$$

$$\pi(\tau \mid x, y, \beta_0, \beta_1, \beta_2, \sigma^{-2}) \propto \tau^{a+\frac{1}{2}} \exp\left[-\tau\left[\frac{1}{2}(\beta_0^2\sigma^{-2} + \beta_1^2\sigma^{-2} + \beta_2^2\sigma^{-2}) + b\right]\right]$$

$$\boxed{\tau \mid x, y, \beta_0, \beta_1, \beta_2, \sigma^{-2} \sim Ga\left(a + \frac{3}{2}, b + \frac{1}{2}(\beta_0^2\sigma^{-2} + \beta_1^2\sigma^{-2} + \beta_2^2\sigma^{-2})\right)}$$

$$\pi(\sigma^{-2} \mid x, y, \beta_0, \beta_1, \beta_2, \tau) \propto (\sigma^{-2})^{\frac{n}{2}+c+\frac{1}{2}} \quad \exp\left[-\sigma^{-2}\left(\frac{1}{2}\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}))^2+\frac{\tau}{2}(\beta_0^2+\beta_1^2+\beta_2^2)+d\right)\right]$$

$$\boxed{\sigma^{-2} \mid x, y, \beta_0, \beta_1, \beta_2, \tau \sim Ga\left(c+\frac{n}{2}+\frac{3}{2},\ d+\frac{1}{2}\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}))^2+\frac{\tau}{2}(\beta_0^2+\beta_1^2+\beta_2^2)\right)}$$