# Computational Statistics Midterm

Liam Dillingham

April 12, 2019

## 1 Problem 1

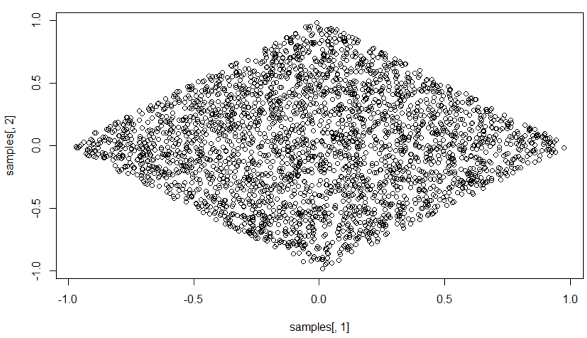
Derive any one scheme for drawing samples from the diamond shaped area:

---

The R code for this is shown below:

```
arfunction = function() {
  while(TRUE) {
    x = runif(1,min = -1,max = 1)
    y = runif(1,min = -1, max = 1)
    if (abs(x) + abs(y) <= 1){
      rv = c(x, y)
      return(rv)
    }
  }
}
samples = matrix(nrow = 3000, ncol = 2)
for(i in 1:3000) {
  x = arfunction()
  samples[i,1] = x[1]
  samples[i,2] = x[2]
}
plot(samples[,1],samples[,2])
```

And the corresponding plot:



## 2 Problem 2

It is well-known that ridge regression tends to give similar coefficient values to correlated variables, whereas the lasso may give quite different coefficient values to correlated vari-ables. We will now explore this property in a very simple setting.
Suppose that $n = 2, p = 2, x_{11} = x_{12}, x_{21} = x_{22}$. Furthermore, suppose that $y_1 + y_2 = 0$, $x_{12} + x_{22} = 0$, and $x_{11} + x_{21} = 0$, so that the estimate for the intercept in a least squares, ridge regression, or lasso model is zero: $\hat{\beta}_0 = 0$.

### 2.1 Write out the ridge regression optimization problem in this setting

I'm trying to visualize the problem at hand here. We have a system that looks like this:

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

So we want to find the values of the $\hat{\beta}_i$'s in the equation $x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 = \hat{y}_i$ such that the computed value, $(\hat{y}_i - y_i)^2$ is minimized for all $i$. In our $p = 2$ case, we want to minimize $(\hat{y_1} - y_1)^2 + (\hat{y_2} - y_2)^2$. However, since we are performing ridge regression, we alter this equation such that we are minimizing the equation:

$$(y_1 - (m_i x_{1i} + b_i)^2 + (y_2 - (m_1 x_{2i} + b_i))^2 + \lambda m_i^2$$

For each $i \in p = 2$. By doing some algebra, we rearrange the equation to:

$$(y_1^2 + y_2^2) - 2m_i(x_{1i}y_1 + x_{2i}y_2) - 2b_i(y_1 + y_2) + m_i^2(x_{1i}^2 + x_{2i}^2) + 2m_i b_i(x_{1i} + x_{2i}) + 2b_i^2 + \lambda m_1^2$$

Since $\overline{x_i} = \frac{x_1 + x_2 + \ldots + x_n}{n}$, and in this case, $n = p = 2$, then $\overline{x_i} = \frac{x_1 + x_2}{2}$, or $x_1 + x_2 = 2\overline{x_i}$. So we can replace a few of these terms:

$$2\overline{y^2} - 4m_i\overline{x_i y} - 4b_i\overline{y} + 2m_i^2\overline{x_i^2} + 4m_i b_i\overline{x_i} + 2b_i^2 + \lambda m_i^2$$

To find the line of best fit, we want to find the minimum point on the error surface produced by $m$ and $b$. We can find those values by calculating the partial derivates for $m$ and $b$, and setting them equal to 0.

$$\frac{\partial SSE}{\partial m_i} = -4\overline{x_i y} + 4m_i\overline{x_i^2} + 4b_i\overline{x_i} + 2\lambda m_i = 0$$

$$\frac{\partial SSE}{\partial b_i} = -4\overline{y} + 4m_i\overline{x_i} + 4b_i = 0$$

We want to coerce these equations into a $y = mx + b$ form, so we do some more algebraic manipulation:

$$\frac{\partial SSE}{\partial m_i} = m_i(\frac{4\overline{x_i^2} + 2\lambda}{4\overline{x_i}}) + b_i = \frac{\overline{x_i y}}{\overline{x_i}} = 0$$

$$\frac{\partial SSE}{\partial b_i} = m_i\overline{x_i} + b_i = \overline{y} = 0$$

And then we have a system of equations, and we can solve for $m$ and $b$. we multiply the partial derivative with respect to $m$ by -1, and add it to the other equation:

$$m_i(\overline{x_i} - \frac{4\overline{x_i^2} + 2\lambda}{4\overline{x_i}}) = \overline{y} - \frac{\overline{x_i y}}{\overline{x_i}}$$

Then, to isolate $m_i$, we divide both sides by its coefficient and simplify:

$$m_i = \frac{4\overline{y}\ \overline{x_i} - \overline{x_i y}}{4(\overline{x_i})^2 - 4\overline{x_i^2} + 2\lambda}$$

And to solve for $b_i$ we substitute back into the equation and solve: $m_i\overline{x_i} + b_i = \overline{y} \Rightarrow b_i = \overline{y} - m_i\overline{x_i}$.

$$b_i = \overline{y} - \frac{4\overline{y}\ \overline{x_i} - \overline{x_i y}}{4(\overline{x_i})^2 - 4\overline{x_i^2} + 2\lambda}\ \overline{x_i} = \overline{y} - \frac{4\overline{y}\ (\overline{x_i})^2 - \overline{x_i y}\ \overline{x_i}}{4(\overline{x_i})^2 - 4\overline{x_i^2} + 2\lambda}$$

Now, it is worth noting a few key facts about this problem. Since $y_1 + y_2 = 0$, $x_{12} + x_{22} = 0$, and $x_{11} + x_{21} = 0$, then $x_{1i} + x_{2i} = 0$, for all $i$. Thus $\overline{x_i} = 0$. So we can eliminate a few terms.

$$m_i = \frac{4\overline{y}\ 0 - \overline{x_i y}}{4(0)^2 - 4\overline{x_i^2} + 2\lambda} = \frac{-\overline{x_i y}}{-4\overline{x_i^2} + 2\lambda}$$

$$b_i = 0 - \frac{0 - \overline{x_i y}\ 0}{4(0)^2 - 4\overline{x_i^2} + 2\lambda} = 0$$

## 2.2 Argue that in this setting, the ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$.

note that $x_{11} = x_{12}, x_{21} = x_{22}$. Since $b_i = 0$ for all $i$, then we don't need to calcualte $b$. However, let's calculate $m_1$.
$-\overline{x_1 y} = \frac{1}{2}(x_{11}y_1 + x_{21}y_2)$, and $\overline{x_1^2} = \frac{1}{2}(x_{11}^2 + x_{21}^2)$. So,

$$m_1 = \frac{-\frac{1}{2}(x_{11}y_1 + x_{21}y_2)}{-2(x_{11}^2 + x_{21}^2) + 2\lambda} \text{ and likewise,}$$

$$m_2 = \frac{-\frac{1}{2}(x_{12}y_1 + x_{22}y_2)}{-2(x_{12}^2 + x_{22}^2) + 2\lambda}$$

using the identities $x_{11} = x_{12}, x_{21} = x_{22}$ in the equation for $m_2$,

$$m_2 = \frac{-\frac{1}{2}(x_{11}y_1 + x_{21}y_2)}{-2(x_{11}^2 + x_{21}^2) + 2\lambda} = m_1$$

Using the notation $\hat{\beta}_1 = m_1$ and $\hat{\beta}_2 = m_2$, we have satisfied $\hat{\beta}_1 = \hat{\beta}_2$.

## 2.3 Write out the lasso optimzation problem in this setting.

Note that for lasso regression, instead of penalizing the SSE by $\lambda m_i^2$, we penalize it by $\lambda|m_i|$. We can actually skip a few of the initial steps of the previous problem, since the differences begin at derivation. Note that instead of differentiating $m_i^2$, we differentiate $|m_i|$, where $\frac{\partial SSE}{\partial m_i} = \frac{m_i}{|m_i|}$. Starting with our original equation adjusted for lasso regression:

$$2\overline{y^2} - 4m_i\overline{x_i y} - 4b_i\overline{y} + 2m_i^2\overline{x_i^2} + 4m_i b_i\overline{x_i} + 2b_i^2 + \lambda|m_i|$$

We then take the partial derivatives:

$$\frac{\partial SSE}{\partial m_i} = -4\overline{x_i y} + 4m_i\overline{x_i^2} + 4b_i\overline{x_i} + \lambda\frac{m_i}{|m_i|} = 0$$

$$\frac{\partial SSE}{\partial b_i} = -4\overline{y} + 4m_i\overline{x_i} + 4b_i = 0$$

However, due to the absolute value of $m_i$, we will be breaking this equation down into two cases, $m_i = m_i$ when $m_i \geq 0$, and $m_i = -m_i$ when $m_i < 0$. And again, coercing the equations into our $mx + b$ form:

*Case 1:* $m_i < 0$

$$\frac{\partial SSE}{\partial m_i} = -4\overline{x_i y} + 4m_i\overline{x_i^2} + 4b_i\overline{x_i} + \lambda\frac{m_i}{-m_i} = 0$$

$$\frac{\partial SSE}{\partial m_i} = m_i\frac{\overline{x_i^2}}{\overline{x_i}} + b_i = \frac{\overline{x_i y}}{\overline{x_i}} + \frac{1}{4}\lambda \text{ And for } b_i,$$

$$\frac{\partial SSE}{\partial b_i} = m_i\overline{x_i} + b_i = \overline{y} = 0$$

We multiply the equation for $m_i$ by -1, add the two equations together, and simplify

$$m_i = \frac{\overline{x_i}\ \overline{y} - \overline{x_i y} - \frac{\lambda}{4}}{(\overline{x_i})^2 - \overline{x_i^2}}$$

Let's first simplify this equation further with the previously known fact that $\overline{x_i} = 0$.

$$m_i = \frac{-\overline{x_i y} - \frac{\lambda}{4}}{-\overline{x_i^2}} \quad \text{then } b_i \text{ equals:}$$

$$b_i = \frac{-\overline{x_i y} - \frac{\lambda}{4}}{-\overline{x_i^2}}\ 0 + b_i = 0 \text{ , Since we already know } \overline{x_i} \text{ and } \overline{y_i} = 0$$

now for case 2.

*Case 2:* $m_i \geq 0$

$$\frac{\partial SSE}{\partial m_i} = -4\overline{x_i y} + 4m_i\overline{x_i^2} + 4b_i\overline{x_i} + \lambda\frac{m_i}{m_i} = 0$$

$$\frac{\partial SSE}{\partial m_i} = m_i\frac{\overline{x_i^2}}{\overline{x_i}} + b_i = \frac{\overline{x_i y}}{\overline{x_i}} - \frac{1}{4}\lambda \text{ And for } b_i,$$

$$\frac{\partial SSE}{\partial b_i} = m_i\overline{x_i} + b_i = \overline{y} = 0$$

The situation is very similar. We can already infer that $b_i = 0$, due to the multiplication from the fact that $\overline{x_i}$ and $\overline{y_i} = 0$. Thus,

$$m_i = \frac{-\overline{x_i y} + \frac{\lambda}{4}}{-\overline{x_i^2}}$$

### 2.4 Argue that in this setting, the lasso coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are not unique - in other words, there are many possible solutions to the optimization problem in (3). Describe these solutions.

Because of the absolute value function that makes up lasso, the error surface can be diamond or v-shaped. So if that error surface is stretched in some direction, rather than stretching the bottom of a bowl to make some sort of trough, the bottom of the pointed bowl is stretched, such that the minimum solution becomes a line rather than a single point.

# 3 Problem 3

A study on the metabolism in 15-year-old females yielded the following data:
$x = (91, 504, 557, 609, 693, 727, 764, 803, 857, 929, 970, 1043, 1089, 1195, 1384, 1713)$. Their energy intake in megajoules, follows a joint density.

### 3.1 Prove the full conditional distributions of $\theta$ and $\sigma^2$

Using the provided full condition, we derive the necessary conditional distributions:

$$f(\sigma^2 \mid x, \theta) = \left[\frac{1}{(\sigma^2)^{n/2}} exp\left(-\Sigma_i(x_i - \theta)^2/(2\sigma^2)\right)\right] \times \left[\frac{1}{\tau} exp\left(-(\theta - \theta_0)^2/(2\tau^2)\right)\right] \times \left[\frac{1}{(\sigma^2)^{a+1}} exp\left(\frac{-b}{\sigma^2}\right)\right]$$

Starting with deriving $\sigma^2 \mid x, \theta$:

$$f(\sigma^2 \mid x, \theta) = \left[\frac{1}{(\sigma^2)^{n/2}} exp\left(-\Sigma_i(x_i - \theta)^2/(2\sigma^2)\right)\right] \times \left[\frac{1}{(\sigma^2)^{a+1}} exp\left(\frac{-b}{\sigma^2}\right)\right] \quad \text{and rearrange:}$$

$$f(\sigma^2 \mid x, \theta) = \left[\frac{1}{(\sigma^2)^{n/2}} \frac{1}{(\sigma^2)^{a+1}} exp\left(-\Sigma_i(x_i - \theta)^2/(2\sigma^2)\ \frac{-b}{\sigma^2}\right)\right]$$

$$f(\sigma^2 \mid x, \theta) = \left[\frac{1}{(\sigma^2)^{n/2+a+1}} exp\left(-\Sigma_i(x_i - \theta)^2/(2\sigma^2)\ \frac{-b}{\sigma^2}\right)\right]$$

$$f(\sigma^2 \mid x, \theta) = \left[\frac{1}{(\sigma^2)^{n/2+a+1}} exp\left(-\left(\frac{1}{2}\frac{\Sigma_i(x_i - \theta)^2}{\sigma^2} + \frac{b}{\sigma^2}\right)\right)\right]$$

Thus the parameters of $\sigma^2 \mid x, \theta$ have the form $a = n/2 + a$ and $b = \frac{1}{2}\Sigma_i(x_i - \theta)^2 + b$. Next, we prove the conditional for $\theta \mid x, \sigma^2$:

$$f(\theta \mid x, \sigma^2) = \left[\frac{1}{(\sigma^2)^{n/2}} exp\left(-\Sigma_i(x_i - \theta)^2/(2\sigma^2)\right)\right] \times \left[\frac{1}{\tau} exp\left(-(\theta - \theta_0)^2/(2\tau^2)\right)\right] \times \left[\frac{1}{(\sigma^2)^{a+1}} exp\left(\frac{-b}{\sigma^2}\right)\right]$$

$$f(\theta \mid x, \sigma^2) = \frac{1}{(\sigma^2)^{n/2+a+1}} exp\left(\frac{-\Sigma_i(x_i - \theta)^2}{2\sigma^2} + \frac{-(\theta - \theta_0)^2}{2\tau^2} + \frac{-b}{\sigma^2}\right)$$

By expanding the summation, we get: $(x_1^2 + x_2^2 + ... + x_n^2) + (-2\theta x_1 - 2\theta x_2 - ... - 2\theta x_n) + n\theta^2$. Using the identity $n\overline{x} = x_1 + x_2 + ... + x_n$, we can remove the summation term from the equation:

$$= \frac{1}{(\sigma^2)^{n/2+a+1}} exp\left(\frac{-2\tau^2 n(\overline{x^2} - 2\theta\overline{x} + \theta^2) + (2\sigma^2)(-(\theta - \theta_0^2))}{4\sigma^2\tau^2} + \frac{-b}{\sigma^2}\right)$$

$$= \frac{1}{(\sigma^2)^{n/2+a+1}} exp\left( \frac{-2\tau^2 n(\overline{x^2} - 2\theta\overline{x} + \theta^2) + (-2\sigma^2)(\theta^2 - 2\theta\theta_0 + \theta_0^2)}{4\sigma^2\tau^2} + \frac{-b}{\sigma^2} \right)$$

$$= \frac{e^{-b/\sigma^2}}{(\sigma^2)^{n/2+a+1}} exp\left( \frac{-2\tau^2 n\overline{x^2} - 4\tau^2\theta\overline{x} + 2\tau^2\theta^2 - 2\sigma^2\theta^2 + 4\sigma^2\theta_0\theta - 2\sigma^2\theta_0^2}{4\sigma^2\tau^2} \right)$$

$$= \frac{e^{-b/\sigma^2}}{(\sigma^2)^{n/2+a+1}} exp\left( \frac{\theta^2(\tau^2 - \sigma^2)}{2\sigma^2\tau^2} + \frac{\theta(\sigma^2\theta_0 - \tau^2\overline{x})}{\sigma^2\tau^2} - \frac{\tau^2 n\overline{x^2} - \sigma^2\theta_0^2}{2\sigma^2\tau^2} \right)$$

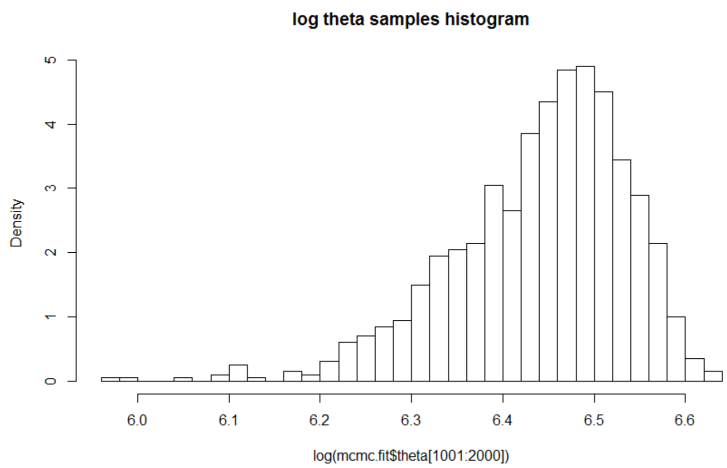
note that in a binomial expansion, $(a+b)^2 = a^2 + 2ab + b^2$. Since $\theta$ is our variable $a$, then our constant term $b^2$ is equivalent to: $\tau^2 n\overline{x^2} - \sigma^2\theta_0^2$. This explains where the numerator comes from regarding the mean of our joint distribution. But the denominator is still mysterious.

### 3.2 Using the normal model above, implement the gibbs sampler

```
library(invgamma)
x = c(91, 504, 557, 609, 693, 727, 764, 803, 857,
      929, 970, 1043, 1089, 1195, 1384, 1713)
n = length(x)
a = 3
b = 3
tau2 = 10
theta0 = 5

theta_mu = function(sig2) {
  mu1 = (sig2 * theta0) / (sig2 + n * tau2)
  mu2 = (n * tau2 * mean(x)) / (sig2 + n * tau2)
  return(mu1 + mu2)
}

theta_sig = function(sig2) {
  sig = (sig2 * tau2) / (sig2 + n * tau2)
  return(sig)
}

gibbs <- function(burn = 1000, nmc = 2000){

  theta <- rep(0, nmc+burn)
  sigma2 <- rep(0, nmc+burn)

  sigma2[1] = rinvgamma(1, (n/2) + a, 0.5 * sum((x - theta0)^(2)) + b)
  theta[1] = theta0;

  for (i in 2:(burn+nmc)) {

    sigma2[i] <- rinvgamma(1, (n/2) + a, theta[i-1])
    theta[i] <- rnorm(1, theta_mu(sigma2[i]), theta_sig(sigma2[i]))
  }
  return(list(sigma2=sigma2,theta=theta))
```

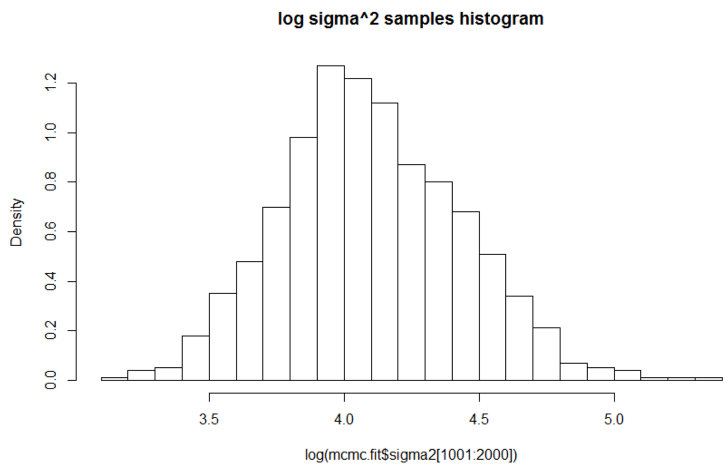
```
}

mcmc.fit <- gibbs()
log.theta = log(mcmc.fit$theta[1001:2000])
log.sigma2 = log(mcmc.fit$sigma2[1001:2000])
hist(log.sigma2, breaks = 30, main="log sigma^2 samples histogram", freq = F)
hist(log.theta, breaks = 30, main="log theta samples histogram", freq = F)

quant.theta = c(quantile(log.theta, 0.05), quantile(log.theta, 0.95))
quant.sigma2 = c(quantile(log.sigma2, 0.05), quantile(log.sigma2, 0.95))

interval.theta = sum((log.theta > quant.theta[1]) & (log.theta < quant.theta[2]))
interval.sigma2 = sum((log.sigma2 > quant.sigma2[1]) & (log.sigma2 < quant.sigma2[2])
```

## 3.3 Plot histograms of $\log(\theta)$ and $\log(\sigma^2)$ and report 90% posterior probability intervals for both




90% posterior probability intervals for $\log(\theta)$ and $\log(\sigma^2)$ note that ($n = 1000$):

```
> interval.theta
[1] 900
> interval.sigma2
[1] 900
```



log sigma^2 samples histogram
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
3.5
4.0
4.5
5.0
log(mcmc.fit$sigma2[1001:2000])