HW9

Zach White

10/13/2016

$$Y = 1\alpha + X\beta + \epsilon$$

With X as a full column rank matrix and assume $1^T X = 0_p$

Exercise 1

$$\begin{aligned}
E_{Y|\beta\phi}[||\hat{\beta} - \beta||] &= E_{Y|\beta\phi}[(\hat{\beta} - \beta)^T(\hat{\beta} - \beta)] \\
&= \sigma^2 tr[(X^T X)^{-1}] \text{ Note } (X^T X) = U\Lambda U^T \\
&= \sigma^2 tr[(U\Lambda U^T)^{-1}] \\
&= \sigma^2 tr[(U^T U\Lambda)^{-1}] \\
&= \sigma^2 tr[(\Lambda)^{-1}] \\
&= \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1}
\end{aligned}$$

Exercise 2

We are given that $p(\alpha, \phi) \propto \phi^{-1}$ and $\beta|g, \phi \sim N(0_p, g\phi^{-1}(X^T X)^{-1})$.

$\beta|Y, \phi \sim N(\frac{g}{1+g}\hat{\beta} + \frac{1}{1+g}0_p, \frac{g}{1+g}\phi^{-1}(X^T X)^{-1}) = N(\frac{g}{1+g}\hat{\beta}, \frac{g}{1+g}\phi^{-1}(X^T X)^{-1})$. Thus, $\tilde{\beta} = E_{\beta|Y,g}[\beta|Y, g] = \frac{g}{1+g}\hat{\beta}$

$\tilde{\beta} = E_{\beta|Y,g}[\beta|Y, g] = \frac{g}{1+g}\hat{\beta}$

Exercise 3

We know that $\tilde{\beta} = \frac{g}{1+g}\hat{\beta}$. We know that the following will follow a normal distribution, and so we find the expectation and variance of the given values.

$E[\tilde{\beta}|Y, g] = E[\frac{g}{1+g}\hat{\beta}] = \frac{g}{1+g}E[\hat{\beta}] = \frac{g}{1+g}\beta$

$\text{Var}[\tilde{\beta}|Y, g] = \text{Var}[\frac{g}{1+g}\hat{\beta}] = \frac{g^2}{(1+g)^2}\text{Var}[\hat{\beta}] = \frac{g^2}{(1+g)^2}\phi^{-1}(X^T X)^{-1}$

Thus, the sampling distribution of $\tilde{\beta}|Y, g, \phi \sim N(\frac{g}{1+g}\beta, \frac{g^2}{(1+g)^2}\phi^{-1}(X^T X)^{-1})$

Exercise 4

This is clearly biased. The definition of bias is $E[\hat{\theta}|\theta_T] - \theta_T = 0$ or in other words $E[\hat{\theta}|\theta_T] = \theta_T$. According to our model, $E[\tilde{\beta}|Y, g] - \beta = \frac{g}{1+g}\beta - \beta$, which is clearly not zero. Thus this values is the bias.

Exercise 5

$E_{Y|\beta\phi}[||\tilde{\beta}-\beta||] = E_{Y|\beta\phi}[(\tilde{\beta}-\beta)^T(\tilde{\beta}-\beta)]$. Note that this is a quadratic form. The expecation of a quadratic form is as follows E$[y^TAy] = tr[A\Sigma] + \mu^TA\mu$ where $y \sim N(\mu,\Sigma)$. Now in our case, $y = \tilde{\beta}-\beta \sim (\frac{g}{1+g}\beta, \frac{g^2}{(1+g)^2}\phi^{-1}(X^TX)^{-1})$.

Thus for our case,

$$
\begin{aligned}
E_{Y|\beta\phi}[||\tilde{\beta}-\beta||] = E_{Y|\beta\phi}[(\tilde{\beta}-\beta)^T(\tilde{\beta}-\beta)] &= tr[A\Sigma] + \mu^TA\mu \\
&= tr[\frac{g^2}{(1+g)^2}(X^TX)^{-1}\phi^{-1}] + [\frac{g}{1+g}\beta-\beta]^T[\frac{g}{1+g}\beta-\beta] \\
&= \frac{g^2}{(1+g)^2}\sigma^2 tr[(X^TX)^{-1}] + \frac{1}{(1+g)^2}\beta^T\beta \\
&= \frac{g^2}{(1+g)^2}E_{Y|\beta\phi}[||\hat{\beta}-\beta||] + \frac{1}{(1+g)^2}||\beta||^2 \\
&= \frac{1}{(1+g)^2}(g^2E_{Y|\beta\phi}[||\hat{\beta}-\beta||] + ||\beta||^2)
\end{aligned}
$$

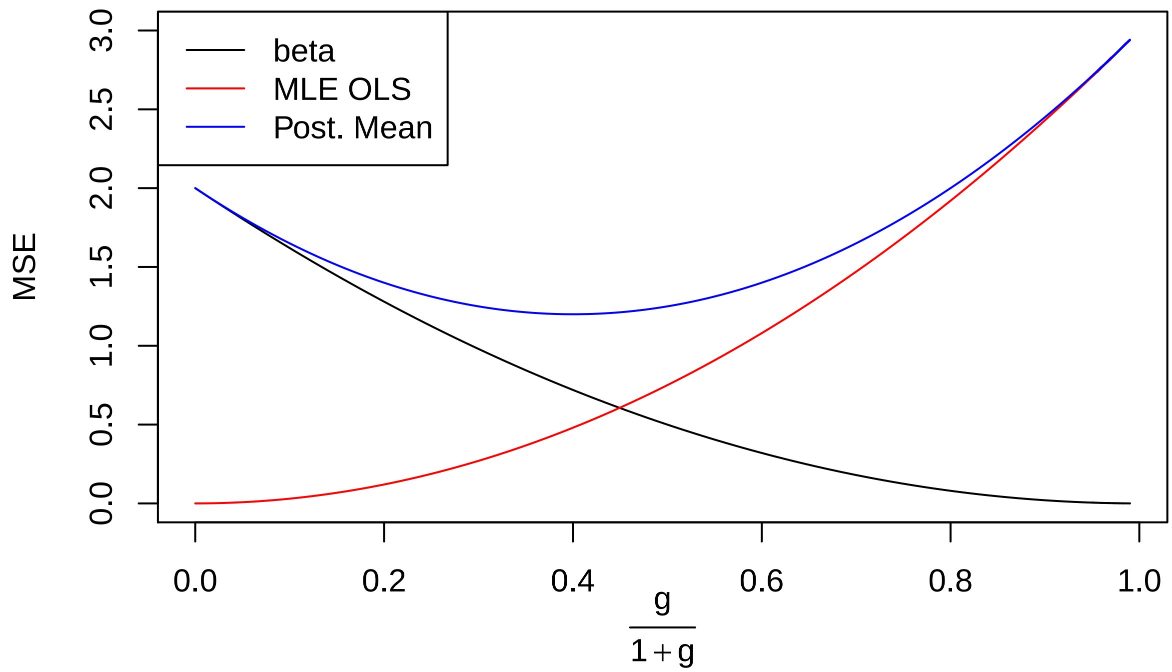
This is clearly a function with the desired arguments.

$$
\begin{aligned}
(\frac{g}{1+g}\beta-\beta)^T(\frac{g}{1+g}\beta-\beta) &= \frac{g^2}{(1+g)^2}\beta^T\beta - 2(\frac{g}{1+g})\beta^T\beta + \beta^T\beta \\
&= (\frac{g^2}{(1+g)^2} - 2(\frac{g(1+g)}{(1+g)^2}) + \frac{(1+g)^2}{(1+g)^2})\beta^T\beta \\
&= \frac{1}{(1+g)^2}\beta^T\beta \\
&= \frac{1}{(1+g)^2}||\beta||^2
\end{aligned}
$$

Exercise 6

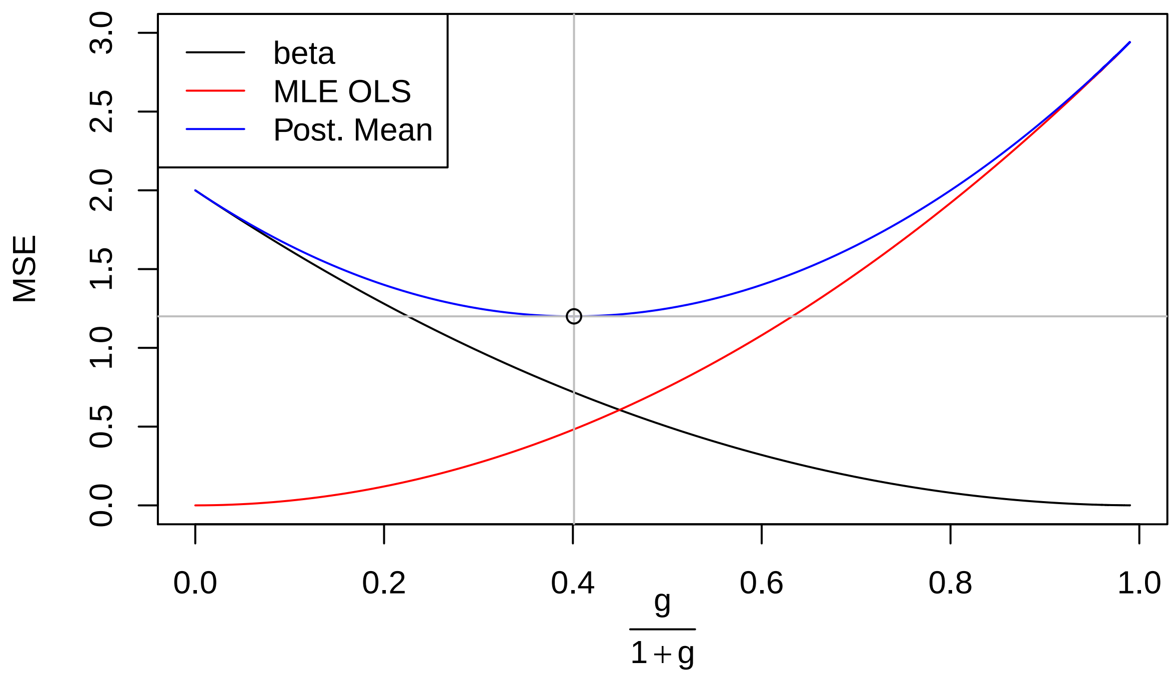
The posterior mean can have a smaller loss than the MLE when estimating β. Note that in our case, $_{Y|\beta\phi}[||\tilde{\beta}-\beta||] = \frac{g^2}{(1+g)^2}E_{Y|\beta\phi}[||\hat{\beta}-\beta||] + \frac{1}{(1+g)^2}||\beta||^2$, and it is important to note that $(1-\frac{g}{1+g})^2 = (\frac{1+g}{1+g} - \frac{g}{1+g})^2 = \frac{1}{(1+g)^2}$. We now designate two arbitrary values for the MSE of the OLS and for β. Now, in this case, our Bayes estimator is uniformly better than our preset MSE of the OLS, which we designated as 3. If we look at the plot, our posterior MSE has a lower value on all values of $\frac{g}{1+g}$.

```
mse.ols = 3
beta.2 = 2

g = seq(0,100, by = .01)

x.axis = g/(1+g)

beta.weight = (1/(1+g)^2) * beta.2
mse.ols = (g^2 / (1+g)^2) *mse.ols
total.weight = beta.weight + mse.ols

plot(x.axis,beta.weight, type = "l", ylim = c(0,3), ylab = "MSE", xlab = expression(frac(g,1+g)))
lines(x.axis,mse.ols, type = "l",col = "red")
lines(x.axis,total.weight, type = "l" , col = "blue")
legend("topleft", c("beta","MLE OLS", "Post. Mean"), col = c("black", "red", "blue"), lty = c(1,1,1))
```



Exercise 7

```
min.mse = min(total.weight)
g.min = g[which.min(total.weight)]
x.place = g.min / (1+g.min)


plot(x.axis,beta.weight, type = "l", ylim = c(0,3), ylab = "MSE", xlab = expression(frac(g,1+g)))
lines(x.axis,mse.ols, type = "l",col = "red")
lines(x.axis,total.weight, type = "l" , col = "blue")
abline(v = x.place, col = "grey")
abline(h = min.mse, col = "grey")
points(x.place,min.mse)
legend("topleft", c("beta","MLE OLS", "Post. Mean"), col = c("black", "red", "blue"), lty = c(1,1,1))
```



In our case, the value of g that minimizes our Bayes estimator is 0.67. Analytically, we find that the value for g that will minimize this function with respect to $\frac{g}{1+g}$ is $g = \frac{||\beta||^2}{MSE_{OLS}} = \frac{||\beta||^2}{E_{Y|\beta\phi}[||\hat{\beta}-\beta||]}$. With this g value, our MSE_{Bayes} should always be lower than the MSE_{OLS}.

Let $x = \frac{g}{1+g}$, $b = ||\beta||^2$, and finally $c = MSE_{OLS}$.

$$\frac{g^2}{(1+g)^2}E_{Y|\beta\phi}[||\hat{\beta}-\beta||] + \frac{1}{(1+g)^2}||\beta||^2 < MLE_{OLS}$$
$$x^2c + (1-x)^2 b < c$$
$$x^2c + b - 2bx + bx^2 - c < 0$$
$$(b+c)x^2 - 2bx + (b-c) < 0$$

Now this equation has the roots, $\frac{b \pm c}{b+c}$, which means that $x \in (\frac{||\beta||^2 - MSE_{OLS}}{||\beta||^2 + MSE_{OLS}}, 1)$, which means that $MSE_{bayes} < MSE_{OLS}$

Also, note to find the maximal g.

$$^2c + (1-x)^2 b = 0$$
$$x^2c + b - 2bx + bx^2 = 0$$
$$(b+c)x^2 - 2bx + b = 0$$
$$x = \frac{b}{b+c}$$
$$\frac{g}{1+g} = \frac{||\beta||^2}{||\beta||^2 + MSE_{OLS}}$$

We can see that the optimal $g = \frac{||\beta||^2}{MSE_{OLS}}$