STA721 Homework 9

Assume the model

$$\mathbf{Y} = \mathbf{1}\alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{X}$ is $n \times p$ and full column rank for the problems below. Without loss, you may assume that $\mathbf{1}^T\mathbf{X} = \mathbf{0}_p$ for below.

1. What is $E_{\mathbf{Y}|\boldsymbol{\beta},\phi}[\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2]$, the expected MSE for OLS under the full model?

2. Find $\tilde{\boldsymbol{\beta}} = \mathsf{E}_{\boldsymbol{\beta}|\mathbf{Y}}[\boldsymbol{\beta} \mid \mathbf{Y}, g]$, the posterior mean under the Zellner g-prior:

$$p(\alpha, \phi) \propto \phi^{-1}$$
$$\boldsymbol{\beta} \mid \phi, g \sim N(\mathbf{0}_p, \frac{g}{\phi}(\mathbf{X}^T\mathbf{X})^{-1})$$

   as a function of the MLE $\hat{\boldsymbol{\beta}}$.

3. Find the *sampling distribution* of $\tilde{\boldsymbol{\beta}}$. (i.e as a function of $\mathbf{Y}$ given parameters, what is the distribution of $\tilde{\boldsymbol{\beta}}$?)

4. Is the posterior mean $\tilde{\boldsymbol{\beta}}$ unbiased for estimating $\boldsymbol{\beta}$? If not, what is the bias? (again the expectation is with respect to the distribution for $\mathbf{Y} \mid \boldsymbol{\beta}, \phi$

5. Find $\mathsf{E}_{\mathbf{Y}|\boldsymbol{\beta},\phi}[\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2]$ assuming model (1) and express as a function of g, $\|\boldsymbol{\beta}\|^2$ and expected MSE for OLS (if possible). This expectation should be taken with respect to the sampling distribution of $\mathbf{Y}$ not the posterior distribution of $\boldsymbol{\beta}$.

6. The Gauss-Markov Theorem showed that out of the class of unbiased linear estimators, the MLE has the smallest variance. If we use the posterior mean above, can the posterior mean have a smaller loss than the MLE for estimating $\boldsymbol{\beta}$? Can it be much worse? Explain. (Make a plot to illustrate with $g/(1+g)$ on the x-axis and MSE on the y-axis); add curves for 1) the sum of the squared bias terms and 2) the variance term (the part from the trace) as a additional lines using different line types and a legend. (you may need to assume or fix values for some quantities that go into the loss, if so how sensitive are the plots/conclusions to those assumptions?).

7. Can you find a value of g to minimize the Expected MSE with the Bayes estimator? If so, what is it? Add this point to you graph above. Simplify as much as possible. With this g will the Expected MSE with the Bayes estimator always be smaller than that with the MLE/OLS? If this depends on the parameters, describe how you could estimate it.