STA721

Homework 11

1. Derive the full conditionals in Casella and Park (2008) see website for link to paper.

2. As a variation on the simulation study in Nott & Kohn (Biometrika 2005) (nott-kohn.R), we will explore shrinkage estimators in the normal linear model

$$\mathbf{Y} \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\sigma^2) \tag{1}$$

where $\mathbf{X}$ has been generated to have a given correlation structure (see the R code nott-kohn.R in the Shrinkage section). Two of the variables have a correlation of near 0.99, with the others more modest. Of the 20 variables, only 8 are related to $\mathbf{Y}$.

   (a) Calculate the $E[(\hat{\beta}-\beta)^T(\hat{\beta}-\beta)]$, the expected MSE for OLS under the full model.

   (b) For each simulation, the OLS coefficients are found and an observed MSE $= (\hat{\beta}^{(s)} - \beta)^T(\hat{\beta}^{(s)} - \beta)$ is computed for each of the s simulated datasets. Does the average of the vector of observed MSEs provide a good estimate of the average of $E[(\hat{\beta} - \beta)^T(\hat{\beta} - \beta)]$? What does the distribution of MSEs look like? Do you think you need to use a larger number of simulations?

   (c) Modify the R-code to use lasso (lars), ridge regression (lm.ridge from MASS or other), and the horseshoe (bhs from monomvn package on CRAN) to estimate β (be careful about which methods standardize variables)). In terms of MSE, which method appears to be best (look at average MSE and side-by-side boxplots)? Which method has the least bias? (most variance?) How do they compare to OLS? Because the methods are compard on the same simulated data, we can use "blocking" to eliminate some of the MC variation. For each simulated data set, take the MSE for all the methods and divide by the smallest MSE for that simulation (hint: use `apply` and `sweep`) and then look at side-by-side boxplots of the relative MSE – those closest to 1 are best.

   (d) (Optional)- repeat the above, but consider predictive MSE for predicting new $\mathbf{Y}^*$'s at new $\mathbf{X}^*$ values with the same correlation structure. Are the methods that are best for estimating $\boldsymbol{\beta}$ also best for estimating $\mathbf{Y}^*$