

# Homework 5: STA 721 Fall19

Your Names

October 30, 2019

You may work in groups of two for this assignment. Both team members should make commits on github and contribute equally to the project. Please look at before class in case there are questions or clarifications needed (or post on Piazza). This complete assignment should be completed using knitr/Sweave.

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

$$Y \sim N(X\beta, I_n\sigma^2) \quad (1)$$

where  $X$  has been generated to have a given correlation structure.

```
n = 50
sigma = 2.5
betatrue = c(4,2,0,0,0,-1,0,1.5, 0,0,0,1,0,.5,0,0,0,0,-1,1,4)
#           int/      X1                      / X2      /X3
# set the random seed so that we can replicate results. Update your seed to be your team number.

set.seed(42)

# number of simulated data sets
nsim = 100

fname = rep("df", nsim)

# create 100 datasets
for (i in 1:nsim) {
  Z = matrix(rnorm(n*10, 0, 1), ncol=10, nrow=n)
  X1 = cbind(Z, (Z[,1:5] %*% c(.3, .5, .7, .9, 1.1) %*% t(rep(1,5)) +
              matrix(rnorm(n*5, 0, 1), ncol=5, nrow=n)))
  X2 <- matrix(rnorm(n*4,0,1), ncol=4, nrow=n)
  X3 <- X2[,4]+rnorm(n,0,sd=0.1)
  X <- cbind(X1,X2,X3)
  X = sweep(X, 2, apply(X,2, mean)) # subtract off the column means
  Y = betatrue[1] + X %*% betatrue[-1] + rnorm(n,0,sigma) # does not have a column of ones for the
  df = data.frame(Y, X)
  fname[i] = paste("df", as.character(i), sep="")
  save(df, file=fname[i])
}
```

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  $Y$ .

- (a) Calculate the  $E[(\hat{\beta} - \beta)^T(\hat{\beta} - \beta)]$ , the expected MSE for OLS under the full model. (should be review!)
- (b) For each simulation, the OLS coefficients are found and an observed  $\text{MSE} = (\hat{\beta}^{(s)} - \beta)^T(\hat{\beta}^{(s)} - \beta)$  is computed for each of the  $s$  simulated datasets.

```
MSE.OLS = rep(NA,nsim)

for( i in 1:nsim) {
  rm(df)
  load(fname[i])
  nk.ols = lm(Y ~ ., data=df)
  coef.ols = coef(nk.ols)
  MSE.OLS[i] = sum((betatrue - coef.ols)^2)
  # print(c(i, MSE.OLS[i])) for the truly bored student
}
```

Does the average of observed MSEs, 49.57, provide a good estimate of the average of  $E[(\hat{\beta} - \beta)^T(\hat{\beta} - \beta)]$ ? What does the distribution of observed MSE minus expected MSEs look like? Do you think you should increase the number of simulated data sets?

- (c) For each of the data sets run lasso, ridge regression (lm.ridge from MASS or other), and the horseshoe (bhs from monomvn package on CRAN or other) to provide estimates of  $\beta$  MSE, variance and squared bias for each simulated dataset; be careful about which methods standardize variables.
  - (d) Derive the full conditional distributions for the Generalized double Pareto model of Armagan, Dunson, and Lee (2011)<sup>1</sup> (<https://arxiv.org/pdf/1104.0861v4.pdf>) (see Section 3) but with the Independent Jeffreys prior on  $\beta_0$  and  $\phi \equiv 1/\sigma^2$ .
  - (e) Implement the Generalized Double Pareto and compare to the above methods, using the default  $\alpha = \eta = 1$ . See Michael Lindon's implementation in R using Rcpp <https://michaellindon.github.io/lindonslog/mathematics/statistics/generalized-double-pareto-shrinkage-priors-sparse-index.html> or implement in JAGS. Be careful about standardizing the predictors and change of variables for coefficients when computing the MSE!
  - (f) In terms of MSE, which method(s) appears to be best (look at average MSE and side-by-side boxplots)? Which method has the least bias? Which method has the most variability?
  - (g) (Optional)- repeat the above simulations, but now 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  $\beta$  also best for estimating  $\mathbf{Y}^*$ ? Which seems to be harder?
2. For the Gaussian ridge prior, the lasso (Double exponential), independent Cauchy, and the Generalized Double pareto, show whether the prior satisfy's any/all of the 3 conditions in Fan and Li (2011) <https://www.tandfonline.com/doi/abs/10.1198/016214501753382273>.

---

<sup>1</sup>Artin Armagan, David Dunson, & Jaeyong Lee (2011). Generalized double Pareto shrinkage Statistica Sinica 23 (2013), 119-143 arXiv: 1104.0861v4