STA721

Homework 12

1. Consider the model:

$$\mathbf{Y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$$

where under $H_a : \boldsymbol{\mu} \in C(\mathbf{X}_0, \mathbf{X}_1)$ where the number of parameters is p. (Assume everything is full rank). To test that $\boldsymbol{\beta}_1 = \mathbf{0}_q$, the F statistic

$$F = \frac{\|(\mathbf{P}_{\mathbf{X}_0,\mathbf{X}_1} - \mathbf{P}_{\mathbf{X}_0})\mathbf{Y}\|^2/q)}{\hat{\sigma}^2} \sim F(q, n-p)$$

under the null hypothesis $H_o : \boldsymbol{\mu} \in C(\mathbf{X}_0)$ and $\hat{\sigma}^2$ is the MSE from the full model H_a.

   (a) Show that $\mathbf{P}_{\mathbf{X}_0,\mathbf{X}_1} - \mathbf{P}_{\mathbf{X}_0}$ is a projection and describe the space onto which it projects.

   (b) Show that the numerator and denominator are independent and find the respective distributions of the two quadratic forms.

   (c) Find the expected value of the numerator under H_0 and H_a (simplify as much as possible). Do you expect the numerator to be larger than σ^2 if the alternative hypothesis is true?

   (d) Find the expected value of $\hat{\sigma}^2$ under H_o and H_a.

   (e) If we instead use the estimate of σ^2 from the null model, is that an unbiased estimate of σ^2 under H_0? under H_a? Explain why it is preferable to use the estimate from the largest model.

2. Apply lasso and horseshoe to the Climate Change data (you will have to pick a basis/design matrix). What can they tell you about the questions of interest? If you code the dummy variables or polynomials in latitude differently will this result in different conclusions? (Explore). Do they provide measures of uncertainty about the degree of temperature changes? Write a brief summary of your analysis and recommendation.