

**MA 751**  
**M. Kon**

**Problem Set 10 (Data Assignment 3)**  
**Due Thurs. 4/21/22**

**Lectures 19, 20**

**Reading: 7.11, 8.1-8.3, 8.7, 8.8**

Problem 8.6. You should do this problem using smoothing splines (section 5.4). If you use a software package to do parts of this problem, please explain clearly what the package is doing and the rationale behind it, including all parts of the package procedure used.

Using smoothing splines:

- (a) You can use (8.21) to get an estimate of  $\hat{\sigma}$ . But note that this single number will *not* represent the typical error bars you will expect around your curve.
- (b) The Bayesian method will require you to use a large matrix  $\mathbf{H}$  (corresponding to the large number of smoothing spline basis elements).
- (c) You can bootstrap 100 samples and repeat the procedure in (a) on each one to get your estimates  $\hat{f}^*$  for each bootstrap. How can you use these to obtain a confidence interval?

Also try to (d) use cross-validation to estimate the error of the smoothing spline estimate in part (a), and also estimate this error using a bagging version of this estimator.

Please show your source code/commands as well as results.

Please also see suggestions under PS 10 related to this problem.