Homework 4

STAT 547, Fall 2019

You are encouraged to discuss the homework questions with classmates or the instructor, but you must write and submit your individual copy. Please write down the name of the persons with whom you discussed the homework, and submit your homework in a single pdf file through Canvas. Due on Oct 29.

1. Consider the univariate nonparametric regression setting where we have a sample (X_i, Y_i), $i = 1, \ldots, n$, which satisfies $Y_i = \mu(X_i) + \epsilon_i$, and the error variance $\mathrm{var}(\epsilon_i) \equiv \sigma^2 > 0$ is a constant. Assume the density of X_i is positive, continuous, and supported on $[0, 1]$. The kernel $K(\cdot)$ is a symmetric continous density function supported on $[-1, 1]$ with $\int_{-1}^{1} K^2(x)dx < \infty$. The regression function μ is assumed to be twice differentiable with a bounded second derivative. Derive the asymptotic bias and variance for the Nadaraya–Watson estimator at a left boundary point $x_0 = ch$, where $c \in [0, 1)$, as $h \to 0$ and $nh \to \infty$.

   [For example, $c = 0$ implies $x_0 = 0$, so only design points falling within $[0, h]$ willl be utilized.]

2. Investigate the scallop abundance data.

   ```
   library(SemiPar)
   data(scallop)
   ```

   Perform bivariate smoothing with the total catch as the response using local polynomial. Vary the bandwidths and degrees and visualize the results.

3. Consider again the yeast gene expression data used in Homework 2. The gene expression profiles may be slightly noisy, so before performing an FPCA we may want to presmooth the individual curves.

   (a) Perform smoothing for each gene expression profile using appropriate smoothing parameters (bandwidth, roughness penalty, etc), and then apply FPCA on the presmoothed curved. Visualize and compare the functional data input, mean functions, and eigenfunctions obtained with and without presmoothing.

   (b) Estimate the derivative curve of each raw gene expression profile. To use local polynomials, one can use the `deriv=1` argument of `locfit`. Then apply FPCA to analyze the estimated derivative curves.