Long Problem Set 7

For each problem below, complete the requested calculations and answer the accompanying questions. Your responses are evaluated on the appropriateness of your approach and the insightfulness of your analysis. Your answers are due before you pick up the second exam.

1. The goal when smoothing data is to improve the signal-to-noise ratio without distorting the underlying signal. The data in the file “StepFunction.RData” consists of four objects: the vector x, which contains 200 values for plotting on the x-axis; the vector y, which contains 200 values for a step-function that satisfies the following criteria

$$y = 0 \text{ for } x \leq 75 \text{ and for } x \geq 126$$

$$y = 1 \text{ for } 75 < x < 126$$

   the vector n, which contains 200 values drawn from random normal distribution with a mean of 0 and standard deviation of 0.1, and the vector s, which is the sum of y and n. In essence, y is the pure signal, n is the noise, and s is a noisy signal. Using this data, complete the following tasks: (a) Determine the mean signal, the standard deviation of the noise, and the signal-to-noise ratio for the noisy signal using just the data in the object s. (b) Explore the effect of applying to the noisy signal, one pass each of moving average filters of widths 5, 7, 9, 11, 13, 15, and 17. For each moving average filter, determine the mean signal, the standard deviation of the noise, and the signal-to-noise ratio. Organize these measurements using a table and comment on your results. Prepare a single plot that displays the original noisy signal and the smoothed signals using widths of 5, 9, 13, and 17, off-setting each so that all five signals are displayed. Comment on your results. (c) Repeat (b) using Savitzky-Golay quadratic/cubic smoothing filters of widths 5, 7, 9, 11, 13, 15, and 17; see Table I of the original paper (link on the course schedule) for each filter’s coefficients. (d) Considering your results for (b) and for (c), what filter and what width provides the greatest improvement in the signal-to-noise ratio with the least distortion of the original signal’s step-function? Be sure to justify your choice.

2. The file “FFT.RData” contains four time-domain signals (a, b, c, and d), each consisting of 32 values and each consisting of a single peak. Examine the data and describe how the time-domain signals differ from each other in terms of peak position, peak width at the baseline, peak area, and peak height. Complete a FFT of each time-domain signal and examine the real part of each in the frequency-domain. Describe how the frequency-domain signals differ from each other in terms of initial intensity, the distance between peak maxima (or between peak minima), and the rate at which the peak maxima (or peak minima) decrease in value. How do the features in the time-domain signals relate to the features in the frequency-domain signals?

3. The file “VeryNoisySignal.RData” consists of a single object, sn, that contains 1024 points of noisy data with a hint of a signal. Prove that there is a signal in this file by using any one moving average or Savitzky-Golay smoothing filter of your choice and using a Fourier filter. Present your results in a single figure that shows the original signal, the signal after smoothing, and the signal after Fourier filtering. Comment on your results.

4. Use a Savitzky-Golay nine-point cubic second-derivative filter to remove the background from the data in file “BkgdRemoval.RData.” Build a calibration model using these results, and report the calibration equation and a plot of the calibration curve.