Math 6070-1: Spring 2013 Problem set 2

Due date: February 20, 2013

In this assignment, you are asked to experiment with some of the initial practical aspects of density estimation. For this you need to generate 1,000 samples from a N(0, 1) distribution.

1. Create a histogram of your data, with $x_0 = 0$ using several different bin sizes. Since you know the exact sampling distribution, experiment with the bin sizes until you have a feeling for which bin size[s] produce more accurate representations of your data. Explain your findings carefully.

Remember that you have to write your own code for this [though you are encouraged to run a canned program to check your work for yourself].

- 2. Produce a kernel density estimate of the N(0, 1) pdf using your sample, using the following different kernels:
 - (a) $K(x) := (2\pi)^{-1/2} \exp\{-x^2/2\}$ [the standard normal kernel];
 - (b) $K(x) := \frac{1}{2}I\{|x| < 1\}$ [the Unif(-1,1) kernel];
 - (c) $K(x) := I\{0 < x < 1\}$ [the Unif(0, 1) kernel];
 - (d) $K(x) := \exp(-x)I\{x > 0\}$ [the one-sided exponential kernel];
 - (e) $K(x) := \frac{1}{2} \exp(-|x|)$ [the double-exponential kernel];
 - (f) $K(x) := \pi^{-1} \{1 + x^2\}^{-1}$ [the Cauchy kernel].

Do this for an array of different bandwidths in each case. Experiment until you find a satisfactory choice of K and h. Explain your findings carefully.

3. Generate 1,000 samples from a Unif(0, 1) distribution next, and redo all the steps of problem #2. Do you notice a difference in your findings?