STAT 526 TERM 2 2002/3 Homework due Mon Jan 20

Kernel Density Estimation

Let X_i, $i = 1, \cdots, n$, be iid with density f. Then the kernel density estimate of $f(x)$ using kernel K and bandwidth h_n is

$$\hat{f}_h(x) = \frac{1}{nh} \sum K\left(\frac{|X_i - x|}{h_n}\right).$$

(1.) Suppose that K is the normal density with mean 0 and variance σ^2. What value of h minimizes the asymptotic integrated mean squared error (with $w(x) \equiv 1$)? Call this h_{opt}. You will use h_{opt} in problem 3 below.

(2.) When Splus's *ksmooth* uses a normal kernel, what is the value of σ^2? What arguments would you use in Splus's *ksmooth* to calculate $\hat{f}$ with K a standard normal density and the bandwidth $h = 3$?

(3.) You will run a simulation study: Generate 100 X_i's from a density of your choosing (continuous random variables!, best to have f continuous on the real line, with continuous first and second derivatives). You can use *ksmooth* or write your own function. If you use *ksmooth* don't forget to use the info from (2.)!

(a.) For your density, calculate h_{opt} and estimate f using your h_{opt}. You might need to use numerical integration - can use something simple like $\int g(x)\, dx = \sum g(x_i)(x_i - x_{i-1})$ using a fine grid of ordered x_i's.

(b.) Calculate lots of estimates of f, using various values of h. Relate what you see to the formula for asymptotic bias. What h do you like for your data set? Compare the h you choose with h_{opt}. (For plots: just hand in a few estimates of f, including one for h_{opt} and the h that you choose.)

(c.) *Bump-hunting.* When using a normal kernel, the number of bumps in $\hat{f}_h$ is a non-increasing function of h (Silverman, JRSS B 1981). Silverman used this fact to come up with a test statistic for testing that a density is unimodal: h_{crit} = the smallest value of h so that $\hat{f}_h$ is unimodal. Reject the null hypothesis of unimodality if h_{crit} is too large. For your generated data, convince yourself that Silverman's result holds. Find h_{crit} by calculating $\hat{f}_h$ for a lot of h values. Don't worry about getting h_{crit} to a large number of decimal points - just do enough to get the idea. Hand in the plot with $\hat{f}_{h_{crit}}$ and a few others indicating that Silverman's result holds.

(4.) *Some asymptotics.* We can estimate f' by taking the derivative of $\hat{f}_h$. What is $\hat{f}'_h(x)$ for K the standard normal kernel? For K the standard normal kernel, show that, under conditions, $E(\hat{f}'_h(x)) = f'(x) + Ch^2+$ smaller order terms and that $var(\hat{f}'_h(x)) = V/(nh^3)+$ smaller order terms.