NONPARAMETRIC SMOOTHING APTS

Lab 2 R. J. Samworth/Sep 2011

Comments and corrections to r.samworth@statslab.cam.ac.uk

1. Plot the Lloyds bank share price data. Now, in different colours and without using built-in `R` functions designed for the purpose, add the local polynomial estimators of order $p = 0, 1, 2$ and 3 with a Gaussian kernel. Use the `dpill` function in the `KernSmooth` package in each case for comparison (even though this is really only appropriate for local linear regression). You should choose the scale of the y-axis to give yourself some room for the final part of the question below. In the cases $p = 0, 1$, check that you get the same results using both the weighted least squares and explicit forms of the estimators.

At $x = 40$, add to your plot the effective kernels of each of the estimators above using the corresponding colours. Experiment with the effect of adding random noise (say with standard deviation 10) to the y-values in the data.

2. In the random design nonparametric regression model for independent and identically distributed pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, observe that the regression function m may be expressed as

$$m(x) = \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x, y)}{f_X(x)} \, dy,$$

where $f_{X,Y}$ is the joint density of (X_1, Y_1) and f_X is the marginal density of X_1. Find the estimator of $m(x)$ that results from estimating f_X and $f_{X,Y}$ using kernel density estimators with symmetric kernel K (and the corresponding product kernel $K_P(x, y) = K(x)K(y)$ in the latter case) and a common bandwidth.

3. Consider the random design homoscedastic regression model. We write

$$\hat{s}_{r,h}(x) = \frac{1}{n} \sum_{i=1}^{n} (X_i - x)^r K_h(X_i - x).$$

Using the fact that

$$\hat{s}_{r,h}(x) = \begin{cases} h^r \mu_r(K) f(x) + o_P(h^r) & \text{if } r \text{ is even} \\ h^{r+1} \mu_{r+1}(K) f'(x) + o_P(h^{r+1}) & \text{if } r \text{ is odd,} \end{cases}$$

compute an asymptotic expression for the conditional bias $\mathbb{E}\{\hat{m}_h(x; 0) | X_1, \ldots, X_n\} - m(x)$ of the local constant estimator at $x \in (0, 1)$ when using a second order kernel with $\int_{-\infty}^{\infty} x K(x) \, dx = 0$ and $\mu_2(K) < \infty$.