Applied Statistics 2013

Homework 5 - Due 29th of April

Exercise 1. *Let $Y_1, \ldots, Y_n$ be i.i.d. random variables from a continuous distribution with unknown density f. We want to estimate $f(0)$, the value of this density at the origin. Let h_n be a small positive number and consider the following estimator*

$$\hat{f}_n(0) = \frac{1}{nh_n} \sum_{i=1}^{n} \mathbf{1}\{Y_i \in [-h_n/2, h_n/2]\} \ .$$

Assume f is at least two times continuously differentiable. Therefore the Taylor expansion of f around zero is simply

$$f(t) = f(0) + f'(0)\ t + f''(0)\frac{t^2}{2} + o(t^2) \ , \qquad \text{as } t \to 0 \ .$$

a) Show that the bias of $\hat{f}_n(0)$ for estimating $f(0)$ is equal to

$$f''(0)\frac{h_n^2}{24} + o(h_n^2) \ ,$$

as $h_n \to 0$.

b) Show that the variance of $\hat{f}_n(0)$ is equal to

$$\frac{f(0)}{nh_n} + o\left(\frac{1}{nh_n}\right) \ ,$$

as $nh_n \to \infty$.

c) Given the above results, what is the appropriate choice of h_n in order to minimize $\mathbb{E}[(\hat{f}_n(0) - f(0))^2]$?

Exercise 2. *Do exercise 2 of*
http://www.win.tue.nl/~rmcastro/AppStat2013/files/simple_regression.pdf

Exercise 3. *Do exercise 3 of*
http://www.win.tue.nl/~rmcastro/AppStat2013/files/simple_regression.pdf

Exercise 4. *Look up the definition of $\hat{r}_{(-i)}(x_i)$ in section 5.3 of the book (All of nonparametric statistics). Show that the Nadaraya-Watson estimator based on all data except for pair (x_i, Y_i), evaluated at x_i, coincides with this definition.*

Exercise 5. *Prove theorem 5.34 (Wasserman, exercise 2 (and 6) of chapter 5).*