3rd sheet of exercises
due: 29 November 2007

Exercise 9. We consider the kernel density estimator $\hat{f}_X$ where the kernel function K is a compactly supported and bounded density and the bandwidth h is positive.

(a) Show that, for any density f_X, we have

$$\|\hat{f}_X - f_X\|_1 \stackrel{n\to\infty}{\longrightarrow} 0 \text{ i.p.} \quad \Longrightarrow \quad E\|\hat{f}_X - f_X\|_1 \stackrel{n\to\infty}{\longrightarrow} 0\,,$$

and that $E\|\hat{f}_X - f_X\|_1 \geq \|E\hat{f}_X - f_X\|_1$.

(b) Prove that $\|E\hat{f}_X - f_X\|_1 \stackrel{n\to\infty}{\longrightarrow} 0$ for some density f_X implies $h \to 0$. Conclude that weak $L_1(\mathbb{R}^d)$-consistency of the kernel estimator for at least one f_X implies $h \to 0$.
Tip: Consider the proof of Lemma 3.1 again.

Exercise 10. Without a proof, Lebesgue's density theorem may be used:

Let f be a d-variate density; μ denotes the d-variate Lebesgue measure; and $S_r(x)$ denotes the closed ball around $x \in \mathbb{R}^d$ with the radius $r > 0$. Then we have

$$\frac{1}{\mu(S_r(x))} \int_{S_r(x)} |f(x) - f(y)| dy \stackrel{r\to 0}{\longrightarrow} 0\,,$$

for almost all $x \in \mathbb{R}^d$ in Lebesgue sense.

Now assume the setting of Exercise 9 and the existence of a density f_X so that $\|\hat{f}_X - f_X\| \stackrel{n\to\infty}{\longrightarrow} 0$ in probability.

(a) Conclude from Exercise 9(a) that $E\|\hat{f}_X - E\hat{f}_X\|_1 \stackrel{n\to\infty}{\longrightarrow} 0$.

(b) Let $S_r(x)$ be a ball in $\mathbb{R}^d$ containing the support of K. Show that

$$E\|\hat{f}_X - E\hat{f}_X\|_1 \geq \int [K_h * f_X](x) \cdot \Big(1 - \int_{S_{rh}(x)} f_X(y) dy\Big)^n dx\,.$$

(c) Assume that $nh^d \to c$ with $c \in [0, \infty)$. Determine the pointwise limit of $\Big(1 - \int_{S_{rh}(x)} f_X(y) dy\Big)^n$ as $n \to \infty$ for almost all $x \in \mathbb{R}^d$.

(d) Study the integral in (b), applying the result of part (c) and Fatou's lemma. Finally, conclude that $nh^d \to \infty$.

Exercise 11. Show that the kernel regression estimator with a kernel K and a bandwidth $h > 0$ satisfies

$$\hat{g}(x) = \arg\min_{a \in \mathbb{R}} \sum_{j=1}^{n} |Y_{j,n} - a|^2 K\big((x - x_{j,n})/h\big)\,,$$

in the fixed design case.

Exercise 12. Consider the standard regression model with fixed design under the additional condition that the i.i.d. regression errors $\varepsilon_{j,n}$ are normal with mean zero. How can one estimate the variance of the $\varepsilon_{j,n}$ without knowing the regression function g in advance?