Exercise sheet 12

Exercise 1 Let $r \geq 0$ be an integer. A natural kernel estimator of the rth derivative, $f^{(r)}(x)$ of a density $f(x)$ is

$$\widehat{f}_h^{(r)}(x) = \frac{1}{nh^{r+1}} \sum_{i=1}^{n} K^{(r)}\left(\frac{x-X_i}{h}\right),$$

where K is an appropriate kernel.

Now let $\beta > r$ be a real number and let l be the unique integer such that $l-1 < \beta \leq l$ and consider the class of functions

$$C_{den}^{\beta}(M) = \{f \text{ density} : f \in C^{l-1}, |f^{(l-1)}(x) - f^{(l-1)}(y)| \leq M|x-y|^{\beta-l+1} \forall x, y \in \mathbb{R}\}.$$

Show that, for an appropriate choice of kernel K,

$$\inf_{h>0} \sup_{f \in C_{den}^{\beta}(M)} MSE\left(\widehat{f}_h^{(r)}(x)\right) \leq C(M, \beta, r, K) n^{-\frac{2(\beta-r)}{2\beta+1}}.$$

Moreover, using the results shown in this exercise, prove that $\|f^{(r)}\|_\infty \leq A_r(\beta, M) < \infty$.

Exercise 2 As in the exercise from last week let f be $C^2(M)$ smooth. Let K be a kernel of order 3 such that $R(K) < \infty$. Show that for any $\epsilon > 0$, there exists a $c_\epsilon > 0$ such that if $h = c_\epsilon n^{-1/5}$ then $MSE(\widehat{f}_h(x)) \leq \epsilon n^{-4/5}$ for n large.

In other words, it does not make much sense to talk about "the optimal" h for a single function. This is why we considered estimators that perform well uniformly on large (infinite-dimensional) classes of functions.

With a bit more work one can find a sequence h_n such that the mean squared error is $o(n^{-4/5})$.

Exercise 3 Let $p \geq 1$. Use convexity to show that for $f, g : \mathbb{R}^k \to \mathbb{R}^d$ and $t \in [0, 1]$,

$$\|f(x) + g(x)\|^p \leq \frac{\|f(x)\|^p}{(1-t)^{p-1}} + \frac{\|g(x)\|^p}{t^{p-1}}$$

Choose t wisely to show **Minkowski's inequality**

$$(\mathbb{E}\|X + Y\|^p)^{1/p} \leq (\mathbb{E}\|X\|^p)^{1/p} + (\mathbb{E}\|Y\|^p)^{1/p}$$

Exercise 4 Let (X_k, Y_k) be a sequence of random vectors such that $X_k \sim P$ for all k and $Y_k \sim Q$ for all k, where P and Q are probability distributions. Using Prokhorov theorem, or otherwise, show that there exists a subsequence (X_{n_k}, Y_{n_k}) that jointly converges in distribution to some random vector (X, Y).

Show that $\liminf_{k\to\infty} \mathbb{E}\|X_{n_k} - Y_{n_k}\|^p \geq \mathbb{E}\|X - Y\|^p$. *Hint: you may wish to consider the bounded continuous function* $f_L(x, y) = \min(L, \|x - y\|^p)$ *and then let* $L \to \infty$.

Deduce that the infimum defining the Wasserstein is always attained.