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

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

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 $\|f^{(r)}\|_\infty \leq A_r(\beta, M) < \infty$.

Solution 1 Let K be $C^\infty(\mathbb{R})$ and supported on $[-1,1]$. We repeatedly integrate by parts to obtain, for $s = 1, \ldots, r$,

$$\int_{-\infty}^{\infty} K_h^{(r)}(x-y) f(y)\, dy = \int_{-\infty}^{\infty} K_h^{(r-1)}(x-y) f'(y)\, dy = \int_{-\infty}^{\infty} K_h^{(r-s)}(x-y) f^{(s)}(y)\, dy,$$

where the boundary terms vanish since K is supported on $[-1,1]$. In particular this holds for $s = r$.

We use the usual bias-variance decomposition. First, the bias term:

$$\mathbb{E}\{\widehat{f}_h^{(r)}(x)\} = \tfrac{1}{n} \sum_{i=1}^{n} \mathbb{E}\{K_h^{(r)}(x - X_i)\}.$$

By definition of $C_{den}^{\beta}(M)$, for

$$R(h,z,x) = f^{(r)}(x-hz) - f^{(r)}(x) - (-hz) f^{(r+1)}(x) - \cdots - \tfrac{(-hz)^{l-r-1}}{(l-r-1)!} f^{(l-1)}(x)$$

we have, for some x^* with $|x^* - x| \leq |hz|$,

$$(l-r-1)!|R(h,z,x)| \leq |hz|^{l-r-1} |f^{(l-1)}(x^*) - f^{(l-1)}(x)| \leq |hz|^{l-r-1} M |hz|^{\beta-l+1} = M|hz|^{\beta-r}.$$

Since K is kernel of order $l-r$,

$$\begin{aligned} bias\{\widehat{f}_h^{(r)}(x)\} &= \int K(z) \left[f^{(r)}(x-hz) - f^{(r)}(x) + hz f^{(r+1)}(x) - \cdots - \tfrac{(-hz)^{l-r-1}}{(l-r-1)!} f^{(l-1)}(x) + R(h,z,x)\right] dz \\ &= \int K(z) R(h,z,x) dz. \end{aligned}$$

The last integral is bounded in absolute value by $M' h^{\beta-r} |\mu|_{\beta-r}(K)$, with $M' = M/(l-r-1)!$.

For the variance term,

$$\begin{aligned}\mathrm{Var}\{\widehat{f}_h^{(r)}(x)\} &\le \tfrac{1}{n}\int (K_h^{(r)})^2(x-y)f(y)\,dy = \tfrac{1}{nh}\left\{\int (K_h^{(r)})^2(z)f(x-hz)\,dz\right\}\\ &= \tfrac{1}{nh^{2r+1}}\left\{\int (K^{(r)})^2(z)f(x-hz)\,dz\right\} \le \tfrac{1}{nh^{2r+1}}R(K^{(r)})\|f\|_\infty \le \tfrac{R(K^{(r)})C_0(\beta,M)}{nh^{2r+1}}.\end{aligned}$$

Conclude that

$$MSE\{\widehat{f}^{(r)}(x)\} \le M'^2h^{2\beta-2r}|\mu|^2_{\beta-r}(K) + \tfrac{R(K^{(r)})C_0(\beta,M)}{nh^{2r+1}}$$

and hence

$$h = \left(\tfrac{(2r+1)R(K^{(r)})C_0(\beta,M)}{2(\beta-r)M'^2|\mu|^2_{\beta-r}(K)n}\right)^{1/(2\beta+1)} \qquad \Longrightarrow \qquad MSE\{\widehat{f}^{(r)}(x)\} \le O\left(n^{-\frac{2(\beta-r)}{2\beta+1}}\right),$$

where the constant depend on K, r, β and M but not on the density f.

Lastly, we have

$$|f^{(r)}(x)| \le |f^{(r)}(x) - \mathbb{E}\widehat{f}^{(r)}(x)| + |\mathbb{E}\widehat{f}^{(r)}(x)| \le M'h^{\beta-r}|\mu|_{\beta-r}(K) + \tfrac{1}{h}\|K^{(r)}\|_\infty$$

and we can define $A_r(\beta, M)$ as the infimum of this with respect to h.

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 $o(n^{-4/5})$.

Solution 2 Mimicking the proof from last week, we now have $\mu_2(K) = 0$, so $\mathrm{bias}(\widehat{f}_h(x)) = o(h^2)$ and $\mathrm{Var}(\widehat{f}_h(x)) = \frac{R(K)f(x)}{nh} + o((nh)^{-1})$. Thus if $h = c_\epsilon n^{-1/5}$ then

$$MSE(\widehat{f}_h(x)) = \tfrac{R(K)f(x)}{nh} + o((nh)^{-1}) + o(h^4) = \tfrac{R(K)f(x)}{c_\epsilon} n^{-4/5} + o(n^{-4/5})$$

Thus if we choose $c_\epsilon = 2/R(K)f(x)\epsilon$ we are done.

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

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

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

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

Solution 3 The inequality is obvious if $t \in \{0, 1\}$. Since y^p is convex on $[0, \infty)$ we have

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

Now assume $\mathbb{E}\|X\|^p + \mathbb{E}\|Y\|^p < \infty$ (there is nothing to prove otherwise) and let

$$t = \tfrac{(\mathbb{E}\|g(X)\|^p)^{1/p}}{(\mathbb{E}\|f(X)\|^p)^{1/p} + (\mathbb{E}\|g(X)\|^p)^{1/p}}$$

to obtain for $\|f\|_p = (\mathbb{E}\|f(X)\|^p)^{1/p}$

$$\|f + g\|_p^p \leq \tfrac{\|f\|_p^p(\|f\|_p + \|g\|_p)^{p-1}}{\|f\|_p^{p-1}} + \tfrac{\|g\|_p^p(\|f\|_p + \|g\|_p)^{p-1}}{\|g\|_p^{p-1}} = (\|f\|_p + \|g\|_p)^p.$$

Finally, apply this for $Z = (X, Y)^\top$, $f(Z) = X$ and $g(Z) = Y$. (The inequality is obvious if $\|f\|_p \in \{0, \infty\}$ or $\|g\|_p \in \{0, \infty\}$).

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.

Solution 4 Given $\epsilon > 0$ let $M < \infty$ such that $\mathbb{P}(\|X_1\| > M) < \epsilon$ and $\mathbb{P}(\|Y_1\| > M) < \epsilon$. Then for all $k \geq 1$

$$\mathbb{P}(\|(X_k, Y_k)^\top\|^2 > 2M^2) \leq \mathbb{P}(\|X_k\|^2 > M^2) + \mathbb{P}(\|Y_k\|^2 > M^2) < 2\epsilon$$

since X_k has the same distribution as X_1 and Y_k has the same distribution as Y_1. Therefore, the sequence (X_k, Y_k) is tight and by Prokhorov theorem has a convergent (in distribution) subsequence to some random vector (X, Y). That $X \sim P$ and $Y \sim Q$ follows from the continuous mapping theorems with the functions $g(x, y) = x$ or $g(x, y) = y$.

Now, we have for all $L > 0$ that

$$\liminf_{k\to\infty} \mathbb{E}\|X_{n_k} - Y_{n_K}\|^p \geq \liminf_{k\to\infty} \mathbb{E} f_L(X_{n_k}, Y_{n_k}) = \mathbb{E} f_L(X, Y) = \mathbb{E}\min(L, \|X - Y\|^p).$$

If $\mathbb{E}\|X - Y\|^p < \infty$, then by the dominated convergence theorem, as $L \to \infty$ the last expectation converges to $\mathbb{E}\|X - Y\|^p$. (If $\mathbb{E}\|X - Y\|^p = \infty$, then we need to use monotone convergence instead.)

Now for $W_p^p(P, Q)$, let (X_n, Y_n) be such that $X_n \sim P$, $Y_n \sim Q$, and $W_p^p(P, Q) = \lim \mathbb{E}\|X_n - Y_n\|^p$. Then by the above construction

$$W_p^p(P, Q) \geq \liminf \mathbb{E}\|X_{n_k} - Y_{n_k}\|^p \geq \mathbb{E}\|X - Y\|^p \geq W_p^p(P, Q).$$

Therefore $\mathbb{E}\|X - Y\|^p = W_p^p(P, Q)$.