Nonparametric Statistics
Take-home exam

R.A. Gunawan
gunawan@few.vu.nl
Free University (VU); student no.: 1069926
June 15, 2007

Assignment 1

Let $R(\hat{p}(x), p(x)) = \mathrm{MSE}(\hat{p}(x), p(x))) = E(\hat{p}(x) - p(x))^2$ be the risk at point x. Assuming that f'' is absolutely-continuous and bounded in the vincinity of x. Under these conditions:

$$R(\hat{p}(x), p(x)) = \left(\frac{1}{4}\sigma_K^4 h_n^4 (p''(x))^2 + p(x) \frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n} \right) (1 + o(1)) \tag{1}$$

as $n \to \infty$ and where K is a kernel that is symmetric around the y axis. A kernel is a smooth function $K : \mathbb{R} \to \mathbb{R}$ such that $K(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} K(x)dx = 1$, $\int_{-\infty}^{\infty} xK(x)dx = 0$ and $\sigma_K^2 \equiv \int_{-\infty}^{\infty} x^2 K(x)dx > 0$. $h_n > 0$ is the so-called bandwidth. One of the commonly used kernels is the Epanechnikov kernel:

$$K(x) = \frac{3}{4}\max(1 - x^2, 0)$$

which is symmetric around the y axis. Let us prove equation (1). The risk $R(\hat{p}(x), p(x))$ is the sum of the bias $\mathrm{Bias}(\hat{p}(x), p(x)) = E(\hat{p}(x) - p(x))$ to the 2nd power and the variance $\mathrm{Var}(\hat{p}(x) - p(x))$ =$\mathrm{Var}(\hat{p}(x))$. The bias can then be written as follows:

$$\begin{aligned}
&\mathrm{Bias}(\hat{p}(x), p(x)) = E[\hat{p}(x) - p(x)] = E\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) - p(x)\right] \\
&= \frac{1}{nh_n}\sum_{i=1}^{n} E\left[K\left(\frac{x - X_i}{h_n}\right)\right] - p(x) \overset{(*)}{=} \frac{1}{nh_n}\sum_{i=1}^{n} E\left[K\left(\frac{X_i - x}{h_n}\right)\right] - p(x) \\
&\overset{iid}{=} \frac{1}{nh_n} nE\left[K\left(\frac{X_1 - x}{h_n}\right)\right] - p(x) = \frac{1}{h_n}E\left[K\left(\frac{X_1 - x}{h_n}\right)\right] - p(x) \\
&= \frac{1}{h_n}E\left[K\left(\frac{X_1 - x}{h_n}\right)\right] - p(x) = \frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{u - x}{h_n}\right) p(u)du - p(x) \\
&\overset{(**)}{=} \int_{-\infty}^{\infty} K(z)p(x + zh_n)dz - p(x) = \int_{-\infty}^{\infty} K(z)\left(\sum_{i=0}^{\infty} \frac{p^{(i)}(x)}{i!}((x + zh_n) - x)^i\right) dz - p(x) \\
&= \int_{-\infty}^{\infty} K(z)\left(\sum_{i=0}^{\infty} \frac{p^{(i)}(x)}{i!}(zh_n)^i\right) dz - p(x) \\
&= p(x)\int_{-\infty}^{\infty} K(z)dz + p'(x)h_n\int_{-\infty}^{\infty} zK(z)dz + \frac{1}{2}h_n^2 p''(x)\int_{-\infty}^{\infty} z^2K(z)dz(1 + o(1)) - p(x) \\
&\overset{(***)}{=} \frac{1}{2}h_n^2 p''(x)\int_{-\infty}^{\infty} z^2K(z)dz(1 + o(1)) = \frac{1}{2}\sigma_K^2 h_n^2 p''(x)(1 + o(1))
\end{aligned}$$

and the variance can be written as follows:

$$\begin{aligned}
\mathrm{Var}(\hat{p}(x)) &= \mathrm{Var}\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right] \overset{(*)}{=} \frac{1}{(nh_n)^2}\sum_{i=1}^{n}\mathrm{Var}\left[K\left(\frac{X_i-x}{h_n}\right)\right] \\
&\overset{iid}{=} \frac{1}{(nh_n)^2} n\mathrm{Var}\left[K\left(\frac{X_1-x}{h_n}\right)\right] = \frac{1}{nh_n^2}\mathrm{Var}\left[K\left(\frac{X_1-x}{h_n}\right)\right] \\
&\overset{(*4)}{=} \frac{1}{nh_n^2} h_n p(x)\int_{-\infty}^{\infty} K^2(t)dt(1+o(1)) = p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n}(1+o(1))
\end{aligned}$$

as $n \to \infty$. The equality $(*)$ is due to the fact that the kernel K is symmetric around the y axis and, for the variance, the X_i's are independent. The equality $(**)$ is due to the change of variables $z = (u-x)/h_n \Leftrightarrow u = x + zh_n \Leftrightarrow du = h_n\, dz$ and the equality $(***)$ is, because $\int_{-\infty}^{\infty} K(z)dz = 1$ and $\int_{-\infty}^{\infty} zK(z)dz = 0$. To prove equality $(*4)$, first notice that
$\mathrm{Var}\left[K\left((X_1-x)/h_n\right)\right] = E\left[K^2\left((X_1-x)/h_n\right)\right] - \left(E\left[K\left((X_1-x)/h_n\right)\right]\right)^2$ and:

$$\begin{aligned}
E\left[K\left(\frac{X_1-x}{h_n}\right)\right] &= \int_{-\infty}^{\infty} K\left(\frac{u-x}{h_n}\right)p(u)du \overset{(**)}{=} \int_{-\infty}^{\infty} h_n K(z)p(x+zh_n)dz \\
&= \int_{-\infty}^{\infty} h_n K(z)\left(p(x) + p'(x)zh_n + \frac{1}{2}p''(x)(zh_n)^2\right)(1+o(1))dz = 0\,(1+o(1))
\end{aligned}$$

as $n \to \infty$, since $h_n \to 0$ as $n \to \infty$. So:

$$\begin{aligned}
\mathrm{Var}\left[K\left(\frac{X_1-x}{h_n}\right)\right] &= E\left[K^2\left(\frac{X_1-x}{h_n}\right)\right] = \int_{-\infty}^{\infty} K^2\left(\frac{u-x}{h_n}\right)p(u)du \\
&\overset{(**)}{=} \int_{-\infty}^{\infty} h_n K^2(z)p(x+zh_n)(1+o(1))du = \int_{-\infty}^{\infty} h_n K^2(z)p(x)(1+o(1))du
\end{aligned}$$

as $n \to \infty$, which ends the proof of equality $(*4)$. The last equality is again due to the fact that $h_n \to 0$ as $n \to \infty$.
The risk can thus be written as:

$$\begin{aligned}
R(\hat{p}(x), p(x)) &= (\mathrm{Bias}(\hat{p}(x), p(x)))^2 + \mathrm{Var}(\hat{p}(x)) \\
&= ((\frac{1}{2}\sigma_K^2 h_n^2 p''(x))^2 + p(x)\frac{\int_{-\infty}^{\infty} K^2(x)dx}{nh_n})(1+o(1)) \qquad (2)
\end{aligned}$$

as $n \to \infty$ which completes the proof of equation (1).
The risk is optimal when the bandwidth is, which can be found as follows:

$$\begin{aligned}
R(\hat{p}(x), p(x)) &= \left(\frac{1}{4}\sigma_K^4 h_n^4 (p''(x))^2 + p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n}\right)(1+o(1)) \\
\frac{\partial}{\partial h_n}R(\hat{p}(x), p(x)) &= \left(\sigma_K^4 h_n^3 (p''(x))^2 - p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n^2}\right)(1+o(1))
\end{aligned}$$

as $n \to \infty$. Now:

$$\begin{aligned}
\frac{\partial}{\partial h_n}R(\hat{p}(x), p(x)) = 0 &\Leftrightarrow \sigma_K^4 h_n^3 (p''(x))^2 - p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n^2} = 0 \\
\sigma_K^4 h_n^3 (p''(x))^2 = p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n^2} &\Leftrightarrow \sigma_K^4 n h_n^5 (p''(x))^2 = p(x)\int_{-\infty}^{\infty} K^2(t)dt \\
h_n^* &= \left(\frac{p(x)\int_{-\infty}^{\infty} K^2(t)dt}{n\sigma_K^4 (p''(x))^2}\right)^{1/5} \qquad (3)
\end{aligned}$$

and

$$\frac{\partial^2}{\partial h_n^2} R(\hat{p}(x), p(x)) = \left(3\sigma_K^4 h_n^2 (p''(x))^2 + 2p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n^3}\right)(1+o(1))$$

as $n \to \infty$. This is always positive if we plug in $h_n = h_n^*$, because h_n^* is positive and so is σ_K^4, $(p''(x))^2$, $p(x)$, n and $\int_{-\infty}^{\infty} K^2(t)dt$. The last one must be true, because K is a kernel, so $K > 0$. Thus the extremum h_n^* is indeed a minimum. Now, the optimal risk is:

$$R^*(\hat{p}(x), p(x)) = \left(\frac{1}{4}\sigma_K^4 (h_n^*)^4 (p''(x))^2 + p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n^*}\right)(1+o(1))$$

$$= \left(\frac{1}{4}\sigma_K^4 \frac{(p(x))^{4/5}(\int_{-\infty}^{\infty} K^2(t)dt)^{4/5}}{n^{4/5}\sigma_K^{16/5}(p''(x))^{8/5}}(p''(x))^2\right.$$

$$\left. + p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{n}\frac{n^{1/5}\sigma_K^{4/5}(p''(x))^{2/5}}{(p(x))^{1/5}\left(\int_{-\infty}^{\infty} K^2(t)dt\right)^{1/5}}\right)(1+o(1))$$

$$= \left(\frac{1}{4}\left(\frac{1}{n}\sigma_K p(x)\int_{-\infty}^{\infty} K^2(t)dt (p''(x))^{1/2}\right)^{4/5} + \left(\frac{1}{n}p(x)\int_{-\infty}^{\infty} K^2(t)dt\sigma_K (p''(x))^{1/2}\right)^{4/5}\right)(1+o(1))$$

$$= cn^{-4/5} = O(n^{-4/5})$$

for $c \in \mathbb{R}$ as $n \to \infty$. Recall that in general $f(x) = O(g(x))$ as $x \to \infty$ if $|f(x)| \le c|g(x)|$ for $x > x_0$ for some c and x_0. Since the risk and n are already positive, the risk is indeed $O(n^{-4/5})$.

Assignment 2

Assume that the density $p(x, y) \ge 0$ for all $(x, y) \in \mathbb{R}^2$ and let us define $h_n := h_{1,n}h_{2,n}$. A suitable 2-dimensional kernel $K(\cdot,\cdot)$ would be

$$K((X_i - x)/h_{1,n}, (Y_i - y)/h_{2,n}) = K((X_i - x)/h_{1,n})\, K((Y_i - y)/h_{2,n}) \tag{4}$$

for $K(\cdot)$ a 1-dimensional kernel, like for example the Epanechnikov kernel mentioned in assignment 1. Notice that $p(x, y) - p(x', y') \le |p(x, y) - p(x', y')| \le L(|x - x'| + |y - y'|)$ and so $p(x, y) \le L(|x - x'| + |y - y'|) + p(x', y') =: p_{\max}$ for some $(x', y') \in \mathbb{R}$ such that $p(x', y') < \infty$ and where $L, p_{\max} \in \mathbb{R}$ for all $(x, y) \in \mathbb{R}^2$, because $p(x, y)$ satisfies the Lipschitz condition. Such a (x', y') must exist, otherwise $p(x, y)$ would not satisfy $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)dxdy = 1$. So the variance of $\hat{p}(x, y) - p(x, y)$ can be written as follows:

$$\begin{aligned}
\mathrm{Var}[\hat{p}(x,y)] &= \mathrm{Var}\left[\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{X_i - x}{h_{1,n}}, \frac{Y_i - y}{h_{2,n}}\right)\right] = \frac{1}{(nh_n)^2}\mathrm{Var}\left[\sum_{i=1}^n K\left(\frac{X_i - x}{h_{1,n}}, \frac{Y_i - y}{h_{2,n}}\right)\right] \\
&\overset{\text{iid}}{=} \frac{1}{(nh_n)^2} n\mathrm{Var}\left[K\left(\frac{X_1 - x}{h_{1,n}}, \frac{Y_1 - y}{h_{2,n}}\right)\right] = \frac{1}{nh_n^2}\mathrm{Var}\left[K\left(\frac{X_1 - x}{h_{1,n}}, \frac{Y_1 - y}{h_{2,n}}\right)\right] \\
&\le \frac{1}{nh_n^2} E\left[K^2\left(\frac{X_1 - x}{h_{1,n}}, \frac{Y_1 - y}{h_{2,n}}\right)\right] = \frac{1}{nh_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2\left(\frac{u - x}{h_{1,n}}, \frac{v - y}{h_{2,n}}\right)p(u,v)dudv \\
&\overset{(*5)}{=} \frac{1}{nh_n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2(s,t)p(x + sh_{1,n}, y + th_{2,n})dsdt \le \frac{p_{\max}}{nh_n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2(s,t)dsdt = \frac{c_3}{nh_n}
\end{aligned}$$

for $c_3 \in \mathbb{R}$. The inequality '$\le$' is, because $\mathrm{Var}[X] = E[X^2] - (E[X])^2 \le E[X^2]$ for general X. The equality $(*5)$ is due to the change of variables $s = (u - x)/h_{1,n} \Leftrightarrow u = x + sh_{1,n} \Leftrightarrow du = h_{1,n}\,ds$ and $t = (v - y)/h_{2,n} \Leftrightarrow v = y + th_{2,n} \Leftrightarrow dv = h_{2,n}\,dt$.

and the Bias$(\hat{p}(x,y), p(x,y))$ can be written as follows:

$$\begin{aligned}
&\text{Bias}(\hat{p}(x,y), p(x,y)) = E[(\hat{p}(x,y) - p(x,y))^2] = E(\hat{p}(x) - p(x)) \\
&= \frac{1}{nh_n} E\left[\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_{1,n}}, \frac{y - Y_i}{h_{2,n}}\right)\right] - p(x,y) \overset{(*)}{=} \frac{1}{nh_n} \sum_{i=1}^{n} E\left[K\left(\frac{X_i - x}{h_{1,n}}, \frac{Y_i - y}{h_{2,n}}\right)\right] - p(x,y) \\
&\overset{\text{iid}}{=} \frac{1}{nh_n} nE\left[K\left(\frac{X_1 - x}{h_{1,n}}, \frac{Y_1 - y}{h_{2,n}}\right)\right] - p(x,y) = \frac{1}{h_n} E\left[K\left(\frac{X_1 - x}{h_{1,n}}, \frac{Y_1 - y}{h_{2,n}}\right)\right] - p(x,y)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{h_n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K\left(\frac{u - x}{h_{1,n}}, \frac{v - y}{h_{2,n}}\right) p(u,v) du dv - p(x,y) \\
&\overset{(*5)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(s,t) p(x + sh_{1,n}, y + th_{2,n}) ds dt - p(x,y) \\
&\overset{(*6)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(s,t)(p(x + sh_{1,n}, y + th_{2,n}) - p(x,y)) ds dt
\end{aligned}$$

The equality $(*6)$ is due to the fact that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(s,t) ds dt = 1$. Now:

$$\begin{aligned}
&\text{Bias}(\hat{p}(x,y), p(x,y)) \leq |\text{Bias}(\hat{p}(x,y), p(x,y))| \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |K(s,t)||p(x + sh_{1,n}, y + th_{2,n}) - p(x,y)| ds dt \\
&\overset{(*7)}{\leq} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |K(s,t)| \, L(|sh_{1,n}| + |th_{2,n}|) ds dt = \int_{-\infty}^{\infty} |K(s)| \, L|sh_{1,n}| ds + \int_{-\infty}^{\infty} |K(t)| \, L|th_{2,n}| dt \\
&= c_1 h_{1,n} + c_2 h_{2,n}
\end{aligned}$$

for $c_1, c_2 \in \mathbb{R}$. The inequality $(*7)$ is, because $p(x,y)$ satisfies the Lipschitz condition.
Thus the risk $R(\hat{p}(x,y), p(x,y)$ has the following upper bound:

$$\begin{aligned}
R(\hat{p}(x,y), p(x,y) &= E[(\hat{p}(x,y) - p(x,y))^2] = (\text{Bias}(\hat{p}(x,y), p(x,y)))^2 + \text{Var}(\hat{p}(x)) \\
&\leq (c_1 h_{1,n} + c_2 h_{2,n})^2 + \frac{c_3}{nh_{1,n}h_{2,n}}
\end{aligned}$$

where $c_1, c_2, c_3 \in \mathbb{R}$. Again, the risk is optimal when the bandwidth is. However, we only have an upper bound of the risk, so we can only determine an optimal upper bound of the optimal risk.
Let us now choose the 2 bandwidths to be the same: $h_{1,n} = h_{2,n}$. So:

$$R(\hat{p}(x,y), p(x,y)) \leq (c_1 + c_2)^2 h_{1,n}^2 + \frac{c_3}{nh_{1,n}^2} =: R_{up}$$

Now:

$$\begin{aligned}
\frac{\partial R_{up}}{\partial h_{1,n}} &= 2(c_1 + c_2)^2 h_{1,n} - \frac{2c_3}{nh_{1,n}^3} \\
\frac{\partial R_{up}}{\partial h_{1,n}} &= 0 \Leftrightarrow 2(c_1 + c_2)^2 n h_{1,n}^4 = 2c_3 \\
h_{1,n}^* &= \left(\frac{c_3}{(c_1 + c_2)^2 n}\right)^{1/4}
\end{aligned}$$

and

$$\frac{\partial^2}{\partial h_{1,n}^2} R_{up} = 2(c_1 + c_2)^2 + \frac{6c_3}{nh_{1,n}^4}$$

This is always positive if we plug in $h_{1,n} = h_{1,n}^*$, because $(c_1 + c_2)^2$, n and $h_{1,n}^4$ are positive and consequently c_3 is positive. The last statement is true, because $c_3 = (c_1 + c_2)^2 n (h_{1,n}^*)^4$. So the extremum $h^*_{1,n}$ is indeed a minimum. Now, the optimal risk is bounded as follows:

$$\begin{aligned} R^*(\hat{p}(x,y),p(x,y)) \leq R^*_{up} &= (c_1+c_2)^2\frac{c_3^{1/2}}{(c_1+c_2)n^{1/2}} + \frac{c_3}{n}\frac{(c_1+c_2)n^{1/2}}{c_3^{1/2}} \\ &= \frac{(c_1+c_2)c_3^{1/2}}{n^{1/2}} + \frac{c_3^{1/2}(c_1+c_2)}{n^{1/2}} = dn^{-1/2} \end{aligned}$$

for $d \in \mathbb{R}$ and so $R^*(\hat{p}(x,y),p(x,y)) = O(n^{-1/2})$.
as $n \to \infty$. The order is justified as before, since the risk and n are already positive. Remark that this result does not apply when the Lipschitz condition of $p(x,y)$ only holds for points in the neighborhood of some point (x_0,y_0) instead of all points in $\mathbb{R}$. The reason is that inequality $(*7)$ does not hold anymore.

Assignment 3

The risk or MSE can be rewritten as follows:

$$\begin{aligned} MSE(\hat{p}(x),p(x)) &= E[(\hat{p}(x)-p(x))^2] = \mathrm{Var}[\hat{p}(x)] + (E[\hat{p}(x)-p(x)])^2 \\ &= E[(\hat{p}(x))^2] - (E[\hat{p}(x)])^2 + (E[\hat{p}(x)]-p(x))^2 = E[(\hat{p}(x))^2] - 2p(x)E[\hat{p}(x)] + (p(x))^2 \end{aligned}$$

Let us calculate the exact expression of $E[(\hat{p}(x))^2]$ and $E[\hat{p}(x)]$. Notice first that $p(x+zh_n) = e^{-(x+zh_n)}$ if $x+zh_n \geq 0 \Leftrightarrow z \geq -x/h_n$.
Now, let $m := \max\{-1,-x/h_n\} = -\min\{1,x/h_n\}$, so $m = -x/h_n$ if $x \leq h_n$ and $m=-1$ if $x > h_n$. Moreover, let $a = x$ if $x \leq h_n$ and $a = h_n$ if $x > h_n$, then:

$$\begin{aligned} E[\hat{p}(x)] &= E\left[\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)\right] = \frac{1}{nh_n}\sum_{i=1}^n E\left[K\left(\frac{X_i-x}{h_n}\right)\right] \overset{iid}{=} \frac{1}{h_n}E\left[K\left(\frac{X_1-x}{h_n}\right)\right] \\ &= \frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{u-x}{h_n}\right)p(u)du = \int_{-\infty}^{\infty} K(z)p(x+zh_n)dz = \frac{3}{4}\int_m^1 (1-z^2)e^{-(x+zh_n)}dz \\ &= \frac{3}{4}\int_m^1 e^{-(x+zh_n)}dz - \frac{3}{4}\int_m^1 z^2 e^{-(x+zh_n)}dz = \frac{3}{4}e^{-x}\left(\int_m^1 e^{-zh_n}dz - \int_m^1 z^2e^{-zh_n}dz\right) \end{aligned}$$

and

$$\int_m^1 e^{-zh_n}dz = \left[-\frac{1}{h_n}e^{-zh_n}\right]_{z=m}^1 = \frac{1}{h_n}(e^a - e^{-h_n}) \tag{5}$$

$$\begin{aligned} \int_m^1 ze^{-zh_n}dz &= \left[-\frac{1}{h_n}ze^{-zh_n}\right]_{z=m}^1 + \frac{1}{h_n}\int_m^1 e^{-zh_n} \\ &\overset{(5)}{=} \frac{1}{h_n}\left(me^a - e^{-h_n}\right) + \frac{1}{h_n^2}(e^a - e^{-h_n}) \end{aligned} \tag{6}$$

$$\begin{aligned} \int_m^1 z^2e^{-zh_n}dz &= \left[-\frac{1}{h_n}z^2e^{-zh_n}dz\right]_m^1 + \frac{2}{h_n}\int_m^1 ze^{-zh_n}dz \\ &\overset{(6)}{=} \frac{1}{h_n}\left(m^2e^a - e^{-h_n}\right) + \frac{2}{h_n^2}\left(me^a - e^{-h_n}\right) + \frac{2}{h_n^3}(e^a - e^{-h_n}) \end{aligned} \tag{7}$$

So:

$$E[\hat{p}(x)] = \frac{3}{4}e^{-x}\left(\frac{1}{h_n}\left(e^a - e^{-h_n}\right) - \frac{1}{h_n}\left(m^2e^a - e^{-h_n}\right) - \frac{2}{h_n^2}\left(me^a - e^{-h_n}\right) - \frac{2}{h_n^3}\left(e^a - e^{-h_n}\right)\right) \tag{8}$$

or equivalently, if $x > h_n$

$$E[\hat{p}(x)] = \frac{3}{4}e^{-x}\left(\frac{2}{h_n^2}\left(e^{h_n}+e^{-h_n}\right)-\frac{2}{h_n^3}\left(e^{h_n}-e^{-h_n}\right)\right)$$

and if $x \le h_n$

$$E[\hat{p}(x)] = \frac{3}{4}e^{-x}\left(\frac{1}{h_n}\left(e^{x}-e^{-h_n}\right)-\frac{1}{h_n}\left(\frac{x^2}{h_n^2}e^{x}-e^{-h_n}\right)+\frac{2}{h_n^2}\left(\frac{x}{h_n}e^{x}+e^{-h_n}\right)-\frac{2}{h_n^3}\left(e^{x}-e^{-h_n}\right)\right)$$

Now, let us determine the exact expression of $E[(\hat{p}(x))^2]$:

$$\begin{aligned}
E[(\hat{p}(x))^2] &= \int_{-\infty}^{\infty} K^2(z)p(x+zh_n)dz = \frac{3}{4}\int_m^1 (1-z^2)^2 e^{-(x+zh_n)}dz = \frac{3}{4}e^{-x}\int_m^1 (1-z^2)^2 e^{-zh_n}dz \\
&= \frac{3}{4}e^{-x}\left(\left[-\frac{1}{h_n}(1-z^2)^2e^{-zh_n}\right]_m^1 - \int_m^1 \frac{4}{h_n}z(1-z^2)e^{-zh_n}\right) \\
&= \frac{3}{4}e^{-x}\left(\frac{1}{h_n}(1-m^2)^2e^{-mh_n} - \int_m^1 \frac{4}{h_n}ze^{-zh_n}dz + \int_m^1 \frac{4}{h_n}z^3e^{-zh_n}dz\right)
\end{aligned}$$

and

$$\begin{aligned}
\int_m^1 z^3e^{-zh_n}dz &= \left[-\frac{1}{h_n}z^3e^{-zh_n}\right]_m^1 + \frac{3}{h_n}\int_m^1 z^2e^{-zh_n}dz \overset{(7)}{=} \frac{1}{h_n}\left(m^3e^{a}-e^{-h_n}\right) \\
&+ \frac{3}{h_n}\left(\frac{1}{h_n}\left(m^2e^{a}-e^{-h_n}\right)+\frac{2}{h_n^2}\left(me^{a}-e^{-h_n}\right)+\frac{2}{h_n^3}(e^{a}-e^{-h_n})\right) \qquad (9)
\end{aligned}$$

So:

$$\begin{aligned}
E[(\hat{p}(x))^2] &= \frac{3}{4}e^{-x}\left(\frac{1}{h_n}(1-m^2)^2e^{-mh_n} - \int_m^1 \frac{4}{h_n}ze^{-zh_n}dz + \int_m^1 \frac{4}{h_n}z^3e^{-zh_n}dz\right) \\
&= \frac{3}{4}e^{-x}\left(\frac{1}{h_n}(1-m^2)^2e^{-mh_n} - \frac{4}{h_n}\left(\frac{1}{h_n}(me^{a}-e^{-h_n})+\frac{1}{h_n}(e^{a}-e^{-h_n})\right)\right. \\
&+ \frac{4}{h_n}\left(\frac{1}{h_n}\left(m^3e^{a}-e^{-h_n}\right)\right. \qquad (10) \\
&\left.\left.+\frac{3}{h_n}\left(\frac{1}{h_n}\left(m^2e^{a}-e^{-h_n}\right)+\frac{2}{h_n^2}\left(me^{a}-e^{-h_n}\right)+\frac{2}{h_n^3}(e^{a}-e^{-h_n})\right)\right)\right)
\end{aligned}$$

Notice now that the risk or MSE derived in assignment 1 is applicable to this case, because $p''(x) = e^{-x}$ is continuous and it is bounded in a neighborhood of $x \ge 0$, namely $p''(x) \le 1$ for all $x \ge 0$. So for this case:

$$\begin{aligned}
\sigma_K^2 &= \int_{-1}^1 x^2K(x)dx = \frac{3}{4}\int_{-1}^1 x^2(1-x^2)dx = \frac{3}{4}\left(\int_{-1}^1 x^2dx - \int_{-1}^1 x^4dx\right) \\
&= \frac{3}{4}\left(\left[\frac{1}{3}x^3\right]_{-1}^1 - \left[\frac{1}{5}x^5\right]_{-1}^1\right) = \frac{3}{4}\left(\frac{2}{3}-\frac{2}{5}\right) = \frac{1}{5}
\end{aligned}$$

$$\begin{aligned}
\int_{-\infty}^{\infty} K^2(x)dx &= \int_{-1}^1 (1-x^2)^2dx = \int_{-1}^1 (1-2x^2+x^4)dx = \left[x-\frac{2}{3}x^3+\frac{1}{5}x^5\right]_{-1}^1 \\
&= 2\left(1-\frac{2}{3}+\frac{1}{5}\right) = \frac{16}{15}
\end{aligned}$$

Applying these results to equation (2):

$$
\begin{aligned}
R(\hat{p}(x) - p(x)) &= \left(\frac{1}{4}\sigma_K^4 h_n^4 (p''(x))^2 + p(x)\frac{\int_{-\infty}^{\infty} K^2(t)dt}{nh_n}\right)(1 + o(1)) \\
&= \left(\frac{1}{100} h_n^4 e^{-2x} + e^{-x}\frac{16}{15}\frac{1}{nh_n}\right) 1_{[0,+\infty)}(x)(1 + o(1))
\end{aligned}
$$

and applying these results to equation (3):

$$
\begin{aligned}
h_n^* &= \left(\frac{p(x)\int_{-\infty}^{\infty} K^2(t)dt}{n\sigma_K^4 (p''(x))^2}\right)^{1/5} = \left(\frac{16/15e^{-x}}{1/25ne^{-2x}}\right)^{1/5} = \left(\frac{80}{3}\frac{1}{n}e^{x}\right)^{1/5} \\
&= (80/3)^{1/5} n^{-1/5} e^{x/5} \qquad (11)
\end{aligned}
$$

as $n \to \infty$, which is the best oracle choice of the bandwidth.

Assignment 4A

Suppose the available data is (x_i, y_i) for $i = 1, ..., n$ where $Y_i \in \mathbb{R}$ and $x_i = (x_{i1}, ..., x_{ip})^T \in \mathbb{R}^p$ and consider the linear regression model:

$$
Y_i = r(x_i) + \varepsilon_i = \sum_{i=1}^{p} \beta_j x_{ij} + \varepsilon_i, \qquad i = 1, ..., n,
$$

where $E[\varepsilon_i] = 0$ and $\text{Var}[\varepsilon_i] = \sigma^2$. An estimator $\hat{r}_n$ of r is called a linear smoother, if there exists a vector $\ell(x) = (\ell_1(x), ..., \ell_n(x))^T$ for every x such that

$$
\hat{r}_n(x) = \sum_{i=1}^{n} \ell_i(x) Y_i \Leftrightarrow \mathbf{r} = LY \qquad (12)
$$

where $\mathbf{r} = (\hat{r}_n(x_1), ..., \hat{r}_n(x_n))^T$, $Y = (Y_1, ..., Y_n)^T$ and $L = (\ell(x_1), ..., \ell(x_n))^T = (\ell_j(x_i)))_{i,j=1}^n$. Linear smoothers, $\ell(x)$ to be exact, depend on a so-called bandwidth $h > 0$. If the bandwidth is increased, then so is the level of smoothness. If it is too low, then it becomes undersmoothed. If it is too high, then it becomes oversmoothed. The goal is to find just the right bandwidth. This can be determined by minimizing the leave-one-out cross validation score:

$$
\hat{R}(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \hat{r}_n(x_i)}{1 - L_{ii}}\right)^2 \qquad (13)
$$

where L_{ii} is the i^{th} diagonal element of the smoothing matrix L.
The variance σ^2 can be estimated by

$$
\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{r}(x_i))^2}{n - 2\nu + \tilde{\nu}} \qquad (14)
$$

where $\nu = \text{tr}(L)$ and $\tilde{\nu} = \text{tr}(L^T L) = \sum_{i=1}^{n} \|\ell(x_i)\|^2$, as long as r is sufficiently smooth, $\nu = o(n)$ and $\tilde{\nu} = o(n)$ as $n \to \infty$.

We will use three types of linear smoothers for the glass fragments data: regressogram, Nadaraya-Watson kernel estimator with a standard normal kernel and local linear regression.

Regressogram

Suppose that $a \leq x_i \leq b$ for $i = 1, ..., n$ and divide $[a, b]$ into m bins $B_1, ..., B_m$, where

$$B_i = \begin{cases} [a + (i-1)\frac{b-a}{m}, \quad a + i\frac{b-a}{m}), & \text{if } i < m \\ [a + (m-1)\frac{b-a}{m}, \quad b], & \text{if } i = m \end{cases}$$

A regressogram $\hat{r}_n$ is defined as follows:

$$\hat{r}_n(x) = \frac{1}{k_j} \sum_{i:\, x_i \in B_j} Y_i, \qquad x \in B_j \tag{15}$$

where $k_j = \#(B_j)$. A regressogram is a linear smoother, because equation (15) can be written as equation (12), where:

$$\ell_i(x) = \begin{cases} 1/k_j, & x_i \in B_j \\ 0, & x_i \notin B_j \end{cases}$$

The bandwidth is the binwidth $h = (b - a)/m$.
The lowest Cross-Validation score for the data appears to be 7.014422 when using $m = 4$ bins, so $h = (b - a)/4$. The variance is estimated at 6.911277. Bin sizes higher than 6 gives an infinite score. Figure 1 plots the Cross-Validation score against the bin size and figure 2 shows the regressogram that has the lowest score. The dotted line in figure 1 is the estimated variance for different bin sizes.

Nadaraya-Watson kernel estimator

The Nadaraya-Watson kernel estimator is defined as equation (12) with the following weights $\ell_i(x)$:

$$\ell_i(x) = \frac{K(\frac{x-x_i}{h})}{\sum_{j=1}^{n} K(\frac{x-x_j}{h})} \tag{16}$$

where $h > 0$ is the bandwidth and K is a kernel. I will use the standard-normal kernel, i.e.:

$$K(x) = \frac{1}{\sqrt{2\Pi}} e^{-x^2/2} \tag{17}$$

The lowest Cross-Validation score for the data appears to be 6.954199 when $h = 0.003$. The variance is estimated at 15.02898. Bandwidths lower than $h = 0.003$ gives an infinite score. Figure 3 plots the Cross-Validation score against the bandwidth. Figure 4 shows the Nadaraya-Watson kernel estimator with a standard-normel kernel that has the lowest score. Figure 7 shows the estimated variances for different bandwidths.

Local linear regression

The local linear regression is defined as equation (12) with the following weights $\ell_i(x)$:

$$\ell_i(x) = \frac{b_i(x)}{\sum_{j=1}^{n} b_j(x)}, \qquad b_i(x) = K\left(\frac{x_i - x}{h}\right)(S_{n,2}(x) - (x_i - x)S_{n,1}(x),$$

$$S_{n,j}(x) = \sum_{i-1}^{n} K\left(\frac{x_i - x}{h}\right)(x_i - x)^j, \qquad j = 1, 2$$

where $h > 0$ is the bandwidth and K is a kernel. I again used the standard-normal kernel.

The lowest Cross-Validation score for the data appears to be 8.432450 when $h = 0.012$. The variance is estimated at 12.74935. Bandwidths lower than $h = 0.012$ gives an infinite score. Figure 5 plots the Cross-Validation score against the bandwidth. Figure 6 shows the local linear regression with a standard-normel kernel that has the lowest score. Figure 8 shows the estimated variances for different bandwidths.

Assignment 5(i)

Recall that if some variable X is $N(\mu, \sigma^2)$ distributed, then $(X - \mu)/\sigma$ is $N(0, 1)$ distributed. In particular, if some variable Z is $N(0, 1)$ distruted, then σZ is $N(0, \sigma^2)$ distributed.
In general, if X is a $N(0, \sigma^2)$ random variable, then X/σ is a standard normal variable, i.e. a $N(0, 1)$ variable. Now, $\hat{\theta}_i = \hat{\theta}_i(Y) = \hat{\theta}_i(\theta, \xi)$ and we let ξ, be a standard normal variable, so $\xi = \xi(\sigma^2 = 1)$. For arbitrary σ^2, the risk is $R(\hat{\theta}_i(\theta, \xi(\sigma^2)), S) = R(\hat{\theta}_i(\theta, \sigma\xi, S))$.

Assignment 5(ii)

Recall that if Z is a $N(0, 1)$ variable, then $\sigma Z + \mu$ is a $N(\mu, \sigma)$ variable. In particular, ξ_k is $N(0, 1)$ distributed, so Y_k is $N(\theta_k, \varepsilon^2)$ distributed for all k, so $E[Y_k] = \theta_k$ and $\text{Var}[Y_k] = \varepsilon^2$ for all k. Suppose that $\lambda_k = 1\{k \leq N\}$, i.e. $\lambda_k = 1$ if $k \leq N$ and $\lambda_k = 0$ otherwise. Let $\hat{\theta}_k$ be the projection estimator, i.e. $\hat{\theta}_k = \lambda_k Y_k$ for all k. N_0, the optimal N, can then be found as follows:

$$
\begin{aligned}
R(\hat{\theta}, \Theta) = E_\theta \left\| \hat{\theta} - \theta \right\|^2 &= E_\theta \left[\sum_{k=1}^{\infty} (\hat{\theta}_k - \theta_k)^2 \right] = E_\theta \left[\sum_{k=1}^{\infty} (\lambda_k Y_k - \theta_k)^2 \right] \\
&= \sum_{k=1}^{\infty} E_\theta \left[(\lambda_k Y_k - \theta_k)^2 \right] = \sum_{k=1}^{\infty} \left(\text{Var}\left[\lambda_k Y_k\right] + \left(E_\theta \left[\lambda_k Y_k - \theta_k\right]\right)^2 \right) \\
&= \sum_{k=1}^{\infty} (\lambda_k^2 \varepsilon^2 + (1 - \lambda_k)^2 \theta_k^2) \leq N\varepsilon^2 + \sum_{k=N+1}^{\infty} \theta_k^2 = N\varepsilon^2 + \sum_{k=N+1}^{\infty} \theta_k^2 \frac{k^{2\beta}}{k^{2\beta}} \\
&\leq N\varepsilon^2 + \frac{1}{N^{2\beta}} \sum_{k=N+1}^{\infty} k^{2\beta}\theta_k^2 \leq N\varepsilon^2 + \frac{1}{N^{2\beta}} \sum_{k=1}^{\infty} k^{2\beta}\theta_k^2 \leq N\varepsilon^2 + QN^{-2\beta} =: R_{up}
\end{aligned}
$$

$$
\frac{\partial}{\partial N} R_{up} = \varepsilon^2 - 2\beta Q N^{-(2\beta+1)}
$$

$$
\frac{\partial}{\partial N} R_{up} = 0 \Leftrightarrow 2\beta Q N^{-(2\beta+1)} = \varepsilon^2 \Leftrightarrow N_0 = (2\beta Q)^{1/(2\beta+1)} \varepsilon^{-2/(2\beta+1)} = (2\beta Q n)^{1/(2\beta+1)}
$$

since $\varepsilon = (1/n)^{1/2}$ and

$$
\frac{\partial^2}{\partial N^2} R_{up} = 2\beta(2\beta + 1) Q N^{-(2\beta+2)}
$$

This is always positive if we plug in $N = N_0$. To see this, note that n is positive and the fact that $N_0 = (2\beta Q n)^{1/(2\beta+1)} > 0$. So Q and β must be positive, otherwise $Q^{1/(2\beta+1)}$ and $\beta^{1/(2\beta+1)}$ would not be real values and neither is N_0. We conclude that the extremum N_0 is indeed a minimum.

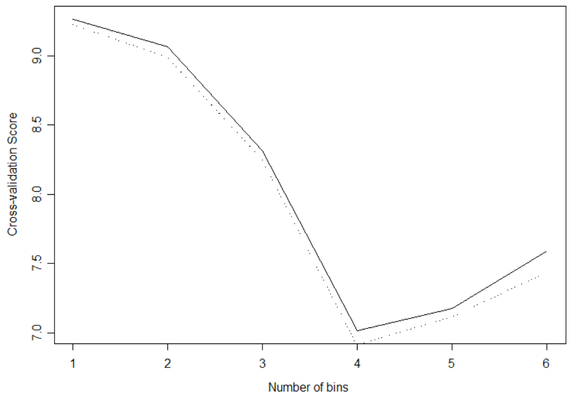

Figure 1: Cross-Validation scores of regressograms for different bin sizes.

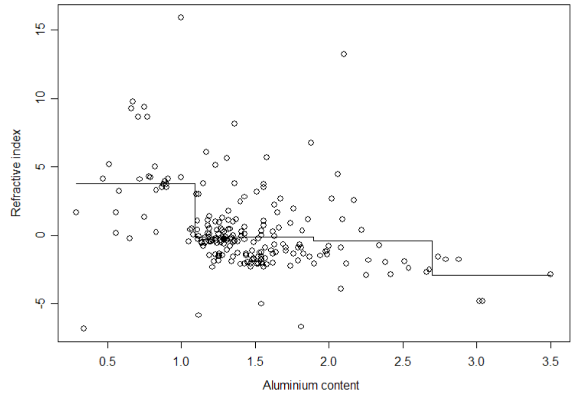

Figure 2: Regressogram with optimal bin size $m = 4$.

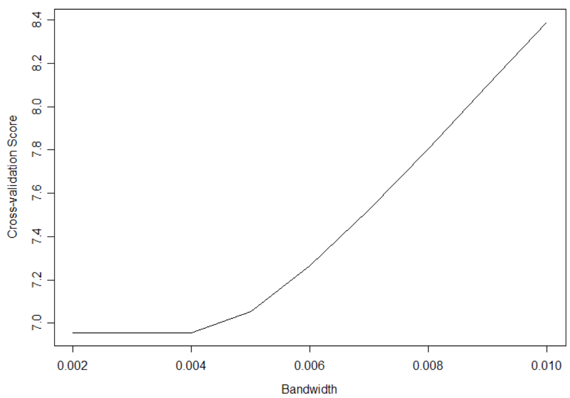

Figure 3: Cross-Validation scores of kernel estimator for different bandwidths.

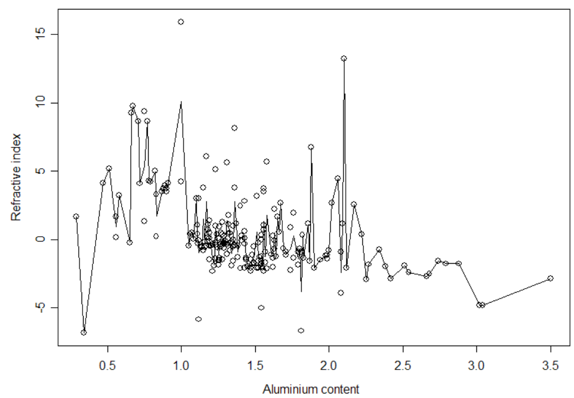

Figure 4: Kernel estimator with optimal bandwidth $h = 0.003$.

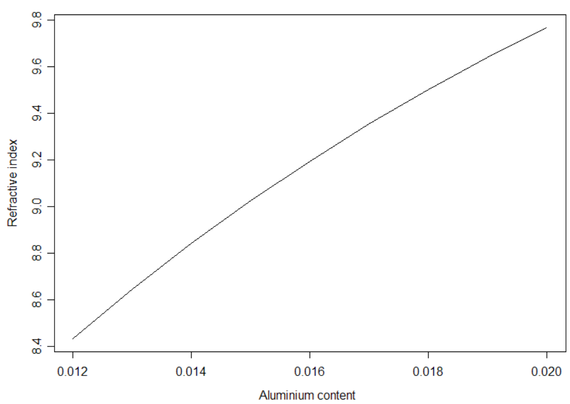

Figure 5: Cross-Validation scores of local linear regression for different bandwidths.

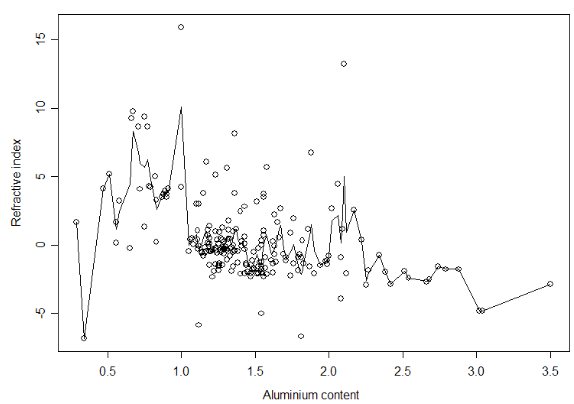

Figure 6: Local linear regression with optimal bandwidth $h = 0.012$.

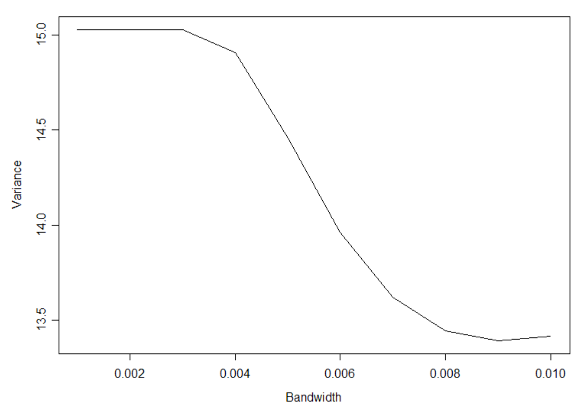

Figure 7: Estimated variances of kernel estimator for different bandwidths.

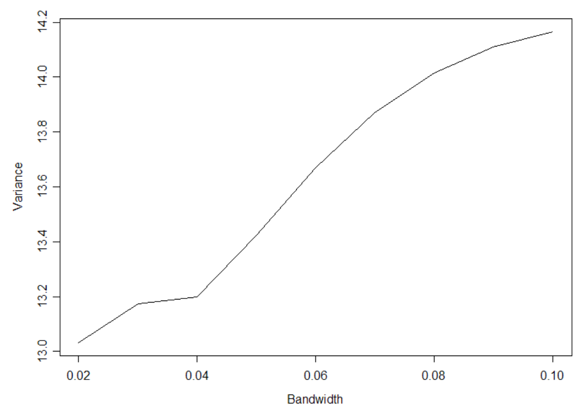

Figure 8: Estimated variances of local linear regression for different bandwidths.

Thus $\hat{\theta}_k(N_0) = Y_k$ if $k \leq N_0 = (2\beta Q n)^{1/(2\beta+1)}$ and $\hat{\theta}_k(N_0) = 0$ otherwise. Now:

$$\begin{aligned} R(\hat{\theta}(N_0), \Theta) &\leq N_0 \varepsilon^2 + Q N_0^{-2\beta} = (2\beta Q)^{1/(2\beta+1)} n^{1/(2\beta+1)} n^{-1} + Q(2\beta Q)^{-2\beta/(2\beta+1)} n^{-2\beta/(2\beta+1)} \\ &= 2\beta(2\beta)^{-2\beta/(2\beta+1)} Q^{1/(2\beta+1)} n^{-2\beta/(2\beta+1)} + Q^{1/(2\beta+1)} (2\beta)^{-2\beta/(2\beta+1)} n^{-2\beta/(2\beta+1)} \\ &= (2\beta+1)(2\beta)^{-2\beta/(2\beta+1)} Q^{1/(2\beta+1)} n^{-2\beta/(2\beta+1)} \end{aligned}$$

Thus $R(\hat{\theta}(N_0), \Theta) \leq \widetilde{C} n^{-2\beta/(2\beta+1)}$ for some $\widetilde{C} \in \mathbb{R}$, so $R(\hat{\theta}(N_0), \Theta) = O(n^{-2\beta/(2\beta+1)})$ as $n \to \infty$, which is indeed the minimax convergence rate. The order is again justified, because the risk and n are already positive. More precisely,

$$\begin{aligned} \widetilde{C} &= (2\beta+1)(2\beta)^{-\frac{2\beta}{2\beta+1}} Q^{\frac{1}{2\beta+1}} = Q^{\frac{1}{2\beta+1}} \left((2\beta)^{\frac{1}{2\beta+1}} + (2\beta)^{-\frac{2\beta}{2\beta+1}} \right) = Q^{\frac{1}{2\beta+1}} \left(\frac{1}{(2\beta)^{2\beta}} + 2\beta \right)^{\frac{1}{2\beta+1}} \\ &= Q^{\frac{1}{2\beta+1}} \left(\frac{1 + (2\beta)^{2\beta+1}}{(2\beta)^{2\beta}} \right)^{\frac{1}{2\beta+1}} \end{aligned}$$

while the minimax constant, or Pinkster's constant, C^* is:

$$C^* = Q^{\frac{1}{2\beta+1}} (2\beta+1)^{\frac{1}{2\beta+1}} \left(\frac{\beta}{\beta+1} \right)^{\frac{2\beta}{2\beta+1}} = Q^{\frac{1}{2\beta+1}} \left(\frac{(2\beta+1)\beta^{2\beta}}{(\beta+1)^{2\beta}} \right)^{\frac{1}{2\beta+1}}$$

They are clearly not equal. So the minimax constant is not attained here.

Assignment 5(iii)

Assuming that γ is positive, the optimal cut-off parameter M_0 here is determined as follows:

$$
\begin{aligned}
R(\hat{\theta}, \Theta) &= E_\theta \left\| \hat{\theta} - \theta \right\|^2 \overset{\text{ass 5(ii)}}{=} \sum_{k=1}^{\infty} (\lambda_k^2 \varepsilon^2 + (1-\lambda_k)^2 \theta_k^2) \leq M\varepsilon^2 + \sum_{k=M+1}^{\infty} \theta_k^2 \\
&= M\varepsilon^2 + \sum_{k=M+1}^{\infty} \theta_k^2 \frac{e^{\gamma k}}{e^{\gamma k}} \leq M\varepsilon^2 + \frac{1}{e^{\gamma M}} \sum_{k=M+1}^{\infty} e^{\gamma k} \theta_k^2 \leq M\varepsilon^2 + \frac{1}{e^{\gamma M}} \sum_{k=1}^{\infty} e^{\gamma k} \theta_k^2 \\
&\leq M\varepsilon^2 + Ce^{-\gamma M} =: R_{up} \\
\frac{\partial}{\partial M} R_{up} &= \varepsilon^2 - \gamma C e^{-\gamma M} \\
\frac{\partial}{\partial M} R_{up} &= 0 \Leftrightarrow \gamma C e^{-\gamma M} = \varepsilon^2 \Leftrightarrow e^{-\gamma M} = \frac{\varepsilon^2}{\gamma C} \Leftrightarrow -\gamma M = \ln\left(\frac{\varepsilon^2}{\gamma C}\right) \\
M_0 &= -\gamma^{-1} \ln\left(\frac{\varepsilon^2}{\gamma C}\right) = -\gamma^{-1} \ln\left(\frac{1}{n\gamma C}\right) = \gamma^{-1} \ln(n\gamma C) = \gamma^{-1}(\ln(n) + \ln(\gamma C))
\end{aligned}
$$

and

$$
\frac{\partial^2}{\partial M^2} R_{up} = \gamma^2 C e^{-\gamma M}
$$

which is always positive, because γ^2, $e^{-\gamma M}$ and C are all positive. The last statement must be true, otherwise M_0 would not be defined, since it would then contain the logarithm of a negative value. So the extremum M_0 is indeed a minimum. Thus $\hat{\theta}_k = Y_k$ if $k \leq M_0 = \gamma^{-1} \ln(n\gamma C)$ and $\hat{\theta}_k = 0$ otherwise. So:

$$
\begin{aligned}
R(\hat{\theta}(M_0), \Theta) &\leq M_0 \varepsilon^2 + Ce^{-\gamma M_0} = \frac{1}{n\gamma} \ln(n\gamma C) + Ce^{-ln(n\gamma C)} = \frac{1}{n\gamma} \ln(n\gamma C) + C(n\gamma C)^{-1} \\
&= \frac{1}{n\gamma} \ln(n\gamma C) + \frac{1}{n\gamma} = \frac{1}{n\gamma} (\ln(n\gamma C) + 1) = \frac{1}{n\gamma} (\ln(n) + \ln(\gamma C) + 1) \approx \frac{1}{n\gamma} \ln(n)
\end{aligned}
$$

if n is large enough. So $R(\theta(\hat{M}_0), \Theta) = O(n^{-1} \ln(n))$. If n is large enough, then $n^{-1} \ln(n) > n^{-2\beta/(\beta+1)}$, meaning that the rate of convergence has become larger now. Thus the optimal projection estimator is performing better here. This makes sense, because the ellipsoid $\mathcal{E}$ contains an exponential power $e^{\gamma k}$, while the ellipsoid Θ has $k^{2\beta}$ and it is known that in general $e^x > x^k$ for all x, regardless of what $k > 0$ is.

Assignment 5(iv)

The optimal cut-off parameter L_0 here is determined as follows:

$$
\begin{aligned}
R(\hat{\theta}, \Theta) = E_\theta \left\| \hat{\theta} - \theta \right\|^2 &\overset{\text{ass 5(ii)}}{=} \sum_{k=1}^{\infty} (\lambda_k^2 \varepsilon^2 + (1-\lambda_k)^2 \theta_k^2) \leq L\varepsilon^2 + \sum_{k=L+1}^{\infty} \theta_k^2 = \left| L\varepsilon^2 + \sum_{k=L+1}^{\infty} \theta_k^2 \right| \\
&\leq \left| L\varepsilon^2 \right| + \sum_{k=L+1}^{\infty} |\theta_k|^2 = L\varepsilon^2 + \sum_{k=L+1}^{\infty} |\theta_k|^2 \leq L\varepsilon^2 + \sum_{k=1}^{\infty} |\theta_k|^2 + |\theta_L|^2 \\
&\leq L\varepsilon^2 + B \sum_{k=1}^{\infty} k^{-2\alpha} + BL^{-2\alpha} =: R_{up}
\end{aligned}
$$

Note that the summation $\sum_{k=1}^{\infty} k^{-2\alpha}$ is a p-series. It only converges if $2\alpha > 1 \Leftrightarrow \alpha > 1/2$, which is indeed the case here. Now:

$$\frac{\partial}{\partial L} R_{up} = \varepsilon^2 - 2\alpha B L^{-(2\alpha+1)}$$
$$\frac{\partial}{\partial L} R_{up} = 0 \overset{\text{ass 5(ii)}}{\Leftrightarrow} L_0 = (2\alpha B n)^{1/(2\alpha+1)}$$

As we already proved in assignment 5(ii) with β and L instead of α and B, this is indeed a minimum. Thus $\hat{\theta}_k = Y_k$ if $k \leq L_0 = (2\alpha B n)^{1/(2\alpha+1)}$ and $\hat{\theta}_k = 0$ otherwise. So:

$$\begin{aligned} R(\hat{\theta}(L_0), \Theta) &\leq L_0 \varepsilon^2 + B \sum_{k=1}^{\infty} k^{-2\alpha} + B L_0^{-2\alpha} \\ &\overset{\text{ass 5(ii)}}{=} (2\alpha+1)(2\alpha)^{-2\alpha/(2\alpha+1)} B^{1/(2\beta+1)} n^{-2\beta/(2\beta+1)} + B \sum_{k=1}^{\infty} k^{-2\alpha} \\ &\approx (2\alpha+1)(2\alpha)^{-2\alpha/(2\alpha+1)} B^{1/(2\beta+1)} n^{-2\beta/(2\beta+1)} \end{aligned}$$

if n is large enough. So $R(\hat{\theta}(L_0), \Theta) = O(n^{-2\beta/(2\beta+1)})$, which is the minimax convergence rate.

Assignment 6(i)

$$\begin{aligned} f_\theta(y) &= f_0(y) + \frac{\theta}{h} k\left(\frac{x-y}{h}\right) \\ |f_\theta(y_1) - f_\theta(y_2)|^{\alpha-r} &\leq \left(|f_0(y_1) - f_0(y_2)| + \frac{|\theta|}{h}\left|k\left(\frac{x-y_1}{h}\right) - k\left(\frac{x-y_2}{h}\right)\right|\right)^{\alpha-r} \\ &\leq |f_0(y_1) - f_0(y_2)|^{\alpha-r} + \frac{|\theta|^{\alpha-r}}{h^{\alpha-r}}\left|k\left(\frac{x-y_1}{h}\right) - k\left(\frac{x-y_2}{h}\right)\right|^{\alpha-r} \\ &\leq |f_0(y_1) - f_0(y_2)|^{\alpha-r} + \delta h^{r+1}\left|k\left(\frac{x-y_1}{h}\right) - k\left(\frac{x-y_2}{h}\right)\right| \\ &\leq L_1|y_1 - y_2|^{\alpha-r} + \delta h^{r+1} L\left|\frac{x-y_1}{h} - \frac{x-y_2}{h}\right|^{\alpha-r} \\ &= L_1|y_1 - y_2|^{\alpha-r} + \delta h^{r+1} L\left|\frac{y_2-y_1}{h}\right|^{\alpha-r} = L_1|y_1 - y_2|^{\alpha-r} + \delta h^r L\,|y_1 - y_2|^{\alpha-r} \\ &= (L_1 + \delta h^r L)\,|y_1 - y_2|^{\alpha-r} \leq L|y_1 - y_2|^{\alpha-r} \end{aligned}$$

The first inequality '$\leq$' is due to the Cauchy-Schwarz inequality and the second inequality is, because $r = \lfloor \alpha \rfloor = \max\{k \in \mathbb{Z} : k < \alpha\}$ and so $0 < \alpha - r \leq 1$. The third is due to the fact that $|\theta| \leq \delta h^{\alpha+1}$ and the fourth is, because $f_0 \in \mathcal{D}(\alpha, L_1)$ and $k(u) \in \mathcal{D}(\alpha, L)$. If h is small enough, then so is $\delta L h^r$. It is known that $L_1 < L$. If h is small enough, then also $L_1 + \delta L h^r < L$, which explains the last inequality. This proves that indeed $f_\theta \in \mathcal{D}(\alpha, L)$ for h small enough.

Proving the second statement:

$$\begin{aligned} I(\theta) &= E_{f_\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(X_1)\right)^2\right] = E_{f_\theta}\left[\left(\frac{1}{f(X_1)}\frac{\partial}{\partial\theta} f(X_1)\right)^2\right] = E_{f_\theta}\left[\frac{1}{f^2(X_1)}\frac{1}{h^2}k^2\left(\frac{X_1-x}{h}\right)\right] \\ &< \frac{1}{h^2}\frac{16}{\varepsilon^2} E\left[k^2\left(\frac{X_1-x}{h}\right)\right] = \frac{1}{h^2}\frac{16}{\varepsilon^2}\int_{-\infty}^{\infty} k^2\left(\frac{u-x}{h}\right) f_\theta(u)du \overset{(**)}{=} \frac{1}{h}\frac{16}{\varepsilon^2}\int_{-\infty}^{\infty} k^2(z) f_\theta(x + z h_n)dz \\ &\leq \frac{1}{h}\frac{16}{\varepsilon^2} f_{\max}\int_{-\infty}^{\infty} k^2(z)dz = \frac{C_1}{h} \end{aligned}$$

Note that $f_0(x) \geq \varepsilon > 0$. Then the first inequality holds, because for sufficiently small h and in a sufficiently small neighborhood of x, $f_\theta(x) = f_0(x) + k(0)\theta/h > \varepsilon/2$ and thus $f_\theta(y) > \varepsilon/4$.

Assignment 6(ii)

a could not be of a smaller order than $h^{\alpha+1}$, because of its definition: $a = \delta h^{\alpha+1}$, where δ is some known parameter.

Assignment 6(iii)

$\lambda_0(u)$ is a density on $(-1,1)$ means it is a function $\lambda_0 : (-1,1) \to \mathbb{R}$ such that $\int_{-1}^{1} \lambda_0(u)du = 1$. Consequently, $\lambda(u)$ is a density on $(-a,a)$, since:

$$\int_{-a}^{a} \lambda(u)du = \frac{1}{a}\int_{-a}^{a} \lambda_0\left(\frac{u}{a}\right) du = \int_{-1}^{1} \lambda_0(v)dv = 1$$

The second last equality is due to a change of variables: $v = u/a \Leftrightarrow dv = 1/a\, du$. Now, let $X \sim \lambda$, then $Y := X/a \sim \lambda_0$ and thus:

$$\begin{aligned}
I(\lambda) &= E\left[\left((\log\lambda(X))'\right)^2\right] = E\left[\left(\frac{\lambda'(X)}{\lambda(X)}\right)^2\right] = \int_{-a}^{a}\left(\frac{\lambda'(u)}{\lambda(u)}\right)^2 \lambda(u)du = \int_{-a}^{a}\frac{(\lambda'(u))^2}{\lambda(u)}du \\
&= \int_{-a}^{a}\frac{(\frac{1}{a}\lambda_0'(\frac{u}{a}))^2}{\frac{1}{a}\lambda_0(\frac{u}{a})}du \overset{(\#)}{=} \int_{-1}^{1}\frac{(\frac{1}{a^2}\lambda_0'(v))^2}{\frac{1}{a}\lambda_0(v)}adv = \frac{1}{a^2}\int_{-1}^{1}\frac{(\lambda_0'(v))^2}{\lambda_0(v)}dv = \frac{1}{a^2}\int_{-1}^{1}\left(\frac{\lambda_0'(v)}{\lambda_0(v)}\right)^2\lambda_0(v)dv \\
&= \frac{1}{a^2}E\left[\left(\frac{\lambda_0'(Y)}{\lambda_0(Y)}\right)^2\right] = \frac{1}{a^2}E\left[\left((\log\lambda_0(Y))'\right)^2\right] = \frac{I_0}{a^2}
\end{aligned}$$

The equality (#) holds, because of the same change of variables and the chain rule:

$$\lambda_0'\left(\frac{u}{a}\right) = \frac{\partial}{\partial u}\lambda_0\left(\frac{u}{a}\right) \overset{def}{=} \frac{\partial}{\partial u}\lambda_0\left(v\right) = \frac{\partial}{\partial u}\lambda_0\left(v(u)\right) = \frac{\partial\lambda_0(v)}{\partial v}\frac{\partial v}{\partial u} = \frac{1}{a}\frac{\partial\lambda_0(v)}{\partial v} = \frac{1}{a}\lambda_0'(v)$$