Topics — Example Sheet 2

Lucas Riedstra

...

Question 1. *Recall that the Epanechnikov kernel is a second-order kernel defined by*

$$K_E(x) = \frac{3}{4\sqrt{5}}\left(1 - \frac{x^2}{5}\right)\mathbb{1}_{|x|\le\sqrt{5}},$$

and that $\mu_2(K_E) = 1$. Let K_0 be another non-negative second-order kernel with $\mu_2(K_0) = 1$. By considering $e(x) \coloneqq K_0(x) - K_E(x)$, or otherwise, show that $R(K_0) \ge R(K_E)$.

Proof. We recall the definitions

$$R(K) = \int_{\mathbb{R}} K^2(u)\,\mathrm{d}u\,, \quad \mu_2(K) = \int_{\mathbb{R}} u^2|K(u)|\,\mathrm{d}u\,,$$

and we recall that K has order 2 if and only if

$$\int_{\mathbb{R}} uK(u)\,\mathrm{d}u = 0.$$

?? $\square$

Question 2. *For $\beta \in (0,1]$ and $L > 0$, let $\mathcal{F}_2(\beta, L)$ denote the class of densities on $\mathbb{R}^2$ that satisfy*

$$|f(x,y) - f(x_0,y_0)| \le L(|x-x_0|^\beta + |y-y_0|^\beta)$$

for all $(x,y),(x_0,y_0) \in \mathbb{R}^2$. Let K be a non-negative kernel on $\mathbb{R}$ with $\mu_\beta(K)$ and $R(K)$ finite. Given i.i.d. pairs $(X_1,Y_1),\ldots,(X_n,Y_n)$, consider the kernel density estimator $\hat{f}_n$ obtained using a product kernel, i.e.,

$$\hat{f}_n(x_0,y_0) \coloneqq \frac{1}{nh^2}\sum_{i=1}^n K\left(\frac{x_0-X_i}{h}\right)K\left(\frac{y_0-Y_i}{h}\right).$$

Find a bound on $\mathrm{MSE}\left\{\hat{f}_n(x_0,y_0)\right\}$ *that holds uniformly for all $f \in \mathcal{F}_2(\beta,L)$ and $(x_0,y_0) \in \mathbb{R}^2$.*

Proof. First we compute a variance bound following the proof of proposition 19: noting that $\hat{f}_n(x,y) = \frac{1}{n}\sum_{i=1}^n K_h(x-X_i)K_h(y-Y_i)$, we have

$$\begin{aligned}
\operatorname{Var}\hat{f}_n(x,y) &= \frac{1}{n}\operatorname{Var}(K_h(x-X_i)K_h(y-Y_i)) \le \frac{1}{n}\mathbb{E}[K_h^2(x-X_i)K_h^2(y-Y_i)] \\
&= \frac{1}{nh^4}\iint_{\mathbb{R}^2} K^2(\frac{x-w}{h})K^2(\frac{y-z}{h})f(w,z)\,\mathrm{d}w\,\mathrm{d}z \\
&\le \frac{\|f\|_\infty}{nh^2}\iint_{\mathbb{R}^2} K^2(s)K^2(t)\,\mathrm{d}s\,\mathrm{d}t = \frac{\|f\|_\infty R(K)^2}{nh^2}.
\end{aligned}$$

Next, we compute a bias bound following the proof of proposition 22. Note that we have

$$\begin{aligned}\operatorname{Bias}\hat{f}_n(x,y) &= \frac{1}{h^2}\iint_{\mathbb{R}^2} K(\frac{x-w}{h})K(\frac{y-z}{h})f(w,z)\,\mathrm{d}w\,\mathrm{d}z - f(x,y)\\ &= \iint_{\mathbb{R}^2} K(s)K(t)\{f(x-sh,y-th)-f(x,y)\}\,\mathrm{d}s\,\mathrm{d}t\,,\end{aligned}$$

and taking absolute values gives

$$\begin{aligned}\left|\operatorname{Bias}\hat{f}_n(x,y)\right| &\le \iint_{\mathbb{R}^2} K(s)K(t)|f(x-sh,y-th)-f(x,y)|\,\mathrm{d}s\,\mathrm{d}t\\ &\le L\iint_{\mathbb{R}^2} K(s)K(t)\Big(|sh|^\beta+|th|^\beta\Big)\,\mathrm{d}s\,\mathrm{d}t\\ &= 2Lh^\beta\int_{\mathbb{R}} K(s)\int_{\mathbb{R}}|t|K(t)\,\mathrm{d}t\,\mathrm{d}s\\ &= 2Lh^\beta\mu_\beta(K)\int_{\mathbb{R}}K(s)\,\mathrm{d}s = 2Lh^\beta\mu_\beta(K).\end{aligned}$$

We therefore have

$$\operatorname{MSE}\hat{f}_n(x,y) \le \frac{1}{nh^2}\|f\|_\infty R(K)^2 + 4L^2\mu_\beta(K)^2h^{2\beta}.$$

Completely analogous to the proof of theorem 23, we can show that $\|f\|_\infty$ is bounded uniformly over $\mathcal{F}_2(\beta,L)$, and the minimiser of MSE is of order $n^{-1/(2\beta+2)}$. Plugging this into the expression gives

$$\sup_{(x,y)}\sup_{f\in\mathcal{F}}\operatorname{MSE}\hat{f}_n(x,y)\le Cn^{-2\beta/(2\beta+2)},$$

for some C depending only on β, L, K. □

Question 3. *Let $\{w_i(x) \mid= 1,\dots,n\}$ denote the effective kernel of the local polynomial estimator of order p based on $(x_1,Y_1),\dots,(x_n,Y_n)$, and let R denote a polynomial of degree at most p. Prove that if $X^\top WX$ is positive definite, then*

$$\frac{1}{n}\sum_{i=1}^n w_p(x,x_i)R(x_i) = R(x)$$

for every $x\in\mathbb{R}$.

Proof. Note that $\frac{1}{n}\sum w(x,x_i)R(x_i)$ is exactly the local polynomial estimator for data $(x_i,R(x_i))_{i=1}^n$ in the point x. Therefore, write $Y=(R(x_1),\dots,R(x_n))^\top\in\mathbb{R}^n$. We know that if $X^\top WX$ is positive definite, then

$$\hat{m}_n(x)=\hat{\beta}_0,\quad \hat{\beta}=(X^\top WX)^{-1}X^\top WY.$$

Now, since R is a polynomial of degree p, there exists a vector $\boldsymbol{v}$ such that

$$Q_h(\cdot-x)^\top\boldsymbol{v}=R(\cdot),$$

and we now have

$$Y=\begin{pmatrix}R(x_1)\\ \vdots\\ R(x_n)\end{pmatrix}=\begin{pmatrix}Q_h(x_1-x)^\top\boldsymbol{v}\\ \vdots\\ Q_h(x_n-x)^\top\boldsymbol{v}\end{pmatrix}=X\boldsymbol{v},$$

and therefore

$$\hat{\beta}=(X^\top WX)^{-1}X^\top WX\boldsymbol{v}=\boldsymbol{v}.$$

It is immediate that

$$R(x)=Q_h(x-x)^\top\boldsymbol{v}=Q_h(0)^\top\boldsymbol{v}=v_1=\hat{m}_n(x),$$

which proves the claim. □

Question 4. *Fix $\beta, L > 0$ and let $m \coloneqq \lceil \beta \rceil - 1$. Recalling the definition of the Hölder class of densities $\mathcal{F}(\beta, L)$, prove that there exists $A = A(\beta, L) > 0$ such that*

$$\sup_{f \in \mathcal{F}(\beta,L)} \max_{j=0,\dots,m} \left\| f^{(j)} \right\|_\infty \le A.$$

Proof. ?? (problem: f being a density does not mean that $f', f'', \dots$ are densities). □

Question 5. *Verify that the local constant and local linear kernel regression estimators have the forms given in the lectures.*

Proof. In the local constant case, we have $Q_h(u) = 1$ and therefore $X = \boldsymbol{e}$ (with $\boldsymbol{e}$ the all-ones vector). It follows that

$$\hat{m}_n(x) = \hat{\beta} = (\boldsymbol{e}^\top W \boldsymbol{e})^{-1} \boldsymbol{e}^\top W Y = \frac{1}{\sum_{i=1}^n W_{ii}} \cdot \sum_{i=1}^n W_{ii} Y_i = \frac{\sum_{i=1}^n K_h(x_i - x) Y_i}{\sum_{i=1}^n K_h(x_i - x)},$$

which corresponds with the expression from the lecture notes.

In the local linear case, we have $Q_h(u) = (1, u/h)^\top$. Define $z_i = (x_i - x)/h$ and $W_i = K_h(x_i - x)$, we have

$$X = \begin{bmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{bmatrix}, \quad X^\top W = \begin{bmatrix} W_1 & \cdots & W_n \\ z_1 W_1 & \cdots & z_n W_n \end{bmatrix}.$$

We can therefore compute, using $\sum_i z_i^r W_i = n h^{-r} s_r(x)$, that

$$X^\top W X = \begin{bmatrix} \sum_i W_i & \sum_i z_i W_i \\ \sum_i z_i W_i & \sum_i z_i^2 W_i \end{bmatrix} = n \begin{bmatrix} s_0(x) & h^{-1} s_1(x) \\ h^{-1} s_1(x) & h^{-2} s_2(x) \end{bmatrix},$$

which gives

$$(X^\top W X)^{-1} = \frac{h^2}{n} \cdot \frac{1}{s_0(x) s_2(x) - s_1(x)^2} \begin{bmatrix} h^{-2} s_2(x) & -h^{-1} s_1(x) \\ -h^{-1} s_1(x) & s_0(x) \end{bmatrix}.$$

Therefore we have

$$X^\top W Y = \begin{bmatrix} \sum_i W_i Y_i \\ \sum_i z_i W_i Y_i \end{bmatrix},$$

and so

$$\begin{aligned} \hat{\beta}_0 &= \frac{h^2}{n} \cdot \frac{1}{s_0(x) s_2(x) - s_1(x)^2} \left(h^{-2} s_2(x) \sum_i W_i Y_i - h^{-1} s_1(x) \sum_i z_i W_i Y_i \right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{s_2(x) - s_1(x)(x_i - x)}{s_0(x) s_2(x) - s_1(x)^2} W_i Y_i, \end{aligned}$$

which corresponds with the expression from the lecture notes. □

Question 6. *In the random design nonparametric regression model for i.i.d. pairs $(X_1, Y_1), \dots, (X_n, Y_n)$, each having joint density $f_{X,Y}$, observe that the regression function m may be expressed as*

$$m(x) = \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x, y)}{f_X(x)} \, \mathrm{d}y\,,$$

where f_X is the marginal density of X_1. Find the estimator $m(x)$ that results from estimating f_X and $f_{X,Y}$ using kernel density estimators with symmetric kernel K (and the corresponding product kernel in the latter case) and a common bandwidth.

Proof. We plug in

$$\begin{aligned}\hat{m}(x) &= \int_{-\infty}^{\infty} y \frac{\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)K_h(y-Y_i)}{\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)}\,\mathrm{d}y \\ &= \frac{\sum_{i=1}^n K_h(X_i-x)Z_i(y)}{\sum_{i=1}^n K_h(X_i-x)},\end{aligned}$$

where $Z_i(y) = \int_{-\infty}^{\infty} yK_h(y-Y_i)\,\mathrm{d}y$. We have

$$\begin{aligned}Z_i(y) &= \int_{-\infty}^{\infty} yK_h(y-Y_i)\,\mathrm{d}y = \int_{-\infty}^{\infty}(y+Y_i)K_h(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} yK_h(y)\,\mathrm{d}y + Y_i\int_{-\infty}^{\infty}K_h(y)\,\mathrm{d}y = Y_i,\end{aligned}$$

(where we use that K_h is symmetric to get rid of $\int yK_h(y)\,\mathrm{d}y$), and so $\hat{m}(x)$ is simply the Nadaraya-Watson estimator (the local polynomial estimator with $p=0$). $\square$

Question 7. *Let $a \le x_1 < \cdots < x_n \le b$, and let $h_i = x_{i+1} - x_i$ for $i = 1,\dots,n-1$. Given $\boldsymbol{g} \in \mathbb{R}^n$ and $\boldsymbol{\gamma} = (\gamma_2,\dots,\gamma_{n-1})^\top \in \mathbb{R}^{n-2}$, show that if there is a natural cubic spline g with $g(x_i) = g_i$ for $i = 1,\dots,n$ and $g''(x_i) = \gamma_i$ for $i = 2,\dots,n-1$ then*

$$g(x) = \frac{(x-x_i)g_{i+1} + (x_{i+1}-x)g_i}{h_i} - \frac{1}{6}(x-x_i)(x_{i+1}-x)\left\{\left(1+\frac{x-x_i}{h_i}\right)\gamma_{i+1} + \left(1+\frac{x_{i+1}-x}{h_i}\right)\gamma_i\right\}$$

for $x \in [x_i, x_{i+1}]$ and $i = 1,\dots,n-1$. Find the corresponding expressions for g on $[a, x_1]$ and $[x_n, b]$.

Proof. On the interval $[x_i, x_{i+1}]$, g must be a cubic polynomial p such that $p(x_i) = g_i$, $p(x_{i+1}) = g_{i+1}$, $p''(x_i) = \gamma_i$, $p''(x_{i+1}) = \gamma_{i+1}$ (where $\gamma_1 = \gamma_n = 0$). Therefore, p must have an expansion in the basis

$$\begin{aligned}&\{q_1(x), q_2(x), q_3(x), q_4(x)\} \\ &:= \left\{x-x_i, x_{i+1}-x, (x-x_i)(x_{i+1}-x)(1+\frac{x-x_i}{h_i}), (x-x_i)(x_{i+1}-x)(1+\frac{x_{i+1}-x}{h_i})\right\}.\end{aligned}$$

Therefore, write $p(x) = aq_1(x) + bq_2(x) + cq_3(x) + dq_4(x)$. Note that

$$q_3''(x_i) = 0, \quad q_3''(x_{i+1}) = -6, \quad q_4''(x_i) = -6, \quad q_4''(x_{i+1}) = 0.$$

We compute a, b, c, d:

1. Since q_2, q_3, q_4 all give 0 when evaluated in x_{i+1}, we must have $g_{i+1} = p(x_{i+1}) = aq_1(x_{i+1}) = ah_i$ or $a = g_{i+1}/h_i$;

2. Analogously, we must have $b = g_i/h_i$;

3. Since q_1'', q_2'', q_4'' all give 0 when evaluated in x_{i+1}, we must have $\gamma_{i+1} = p''(x_{i+1}) = cq_3''(x_{i+1}) = -6c$ or $c = -\gamma_{i+1}/6$;

4. Analogously, we must have $d = -\gamma_i/6$.

Combining the above proves the claim.

On $[a, x_1]$, g simply needs to be a linear function with $g(x_1) = g_1$, which means it has the form $g(x) = c(x-x_1) + g_1$ for some c. We must then choose c such that the first derivative is continuous. An analogous argument holds for g on $[x_n, b]$. $\square$

Question 8. *Define tri-diagonal matrices $Q = (q_{ij})_{i=1,j=2}^{n,n-1} \in \mathbb{R}^{n\times(n-2)}$ and $R = (r_{ij})_{i=2,j=2}^{n-1,n-1} \in \mathbb{R}^{(n-2)\times(n-2)}$ by*

$$q_{ij} \coloneqq \begin{cases} 1/h_i, & j = i+1, \\ -1/h_{i-1} - 1/h_i, & j = i, \\ 1/h_{i-1}, & j = i-1, \end{cases} \qquad r_{ij} = \begin{cases} h_i/6, & j = i+1, \\ (h_{i-1}+h_i)/3, & j = i, \\ h_{i-1}/6, & j = i-1. \end{cases}$$

Prove that $\boldsymbol{g}, \boldsymbol{\gamma}$ represent a natural cubic spline if and only if $Q^\top \boldsymbol{g} = R\boldsymbol{\gamma}$.

Proof. In the previous question, we saw that there was at most one candidate g for the natural cubic spline which satisfies $g(x_i) = g_i$ $(i = 1, \dots, n)$ and $g''(x_i) = \gamma_i$ $(i = 2, \dots, n-1)$. For this g, we know that g and g'' are continuous, however, we do not know if g' is continuous. We find that g represents a natural cubic spline if and only if g' is continuous at the points $x_2, \dots, x_{n-1}$.

Fix $j \in \{2, \dots, n-1\}$. On the interval $[x_{j-1}, x_j]$, we have

$$g'(x_j) = \frac{g_j - g_{j-1}}{h_{j-1}} + \frac{1}{6}h_{j-1}\{2\gamma_j + \gamma_{j-1}\},$$

while on the interval $[x_j, x_{j+1}]$, we have

$$g'(x_j) = \frac{g_{j+1} - g_j}{h_j} - \frac{1}{6}h_j\{\gamma_{j+1} + 2\gamma_j\}.$$

Equating the two and rearranging terms, we obtain

$$\frac{1}{h_j}g_{j+1} - \left(\frac{1}{h_j} + \frac{1}{h_{j-1}}\right)g_j + \frac{1}{h_{j-1}}g_{j-1} = \frac{h_j}{6}\gamma_{j+1} + \frac{h_{i-1}+h_i}{3}\gamma_j + \frac{h_{j-1}}{6}\gamma_{j-1}.$$

It is easily seen that these equations can be rewritten as $Q^\top \boldsymbol{g} = R\boldsymbol{\gamma}$, which proves the claim. □

Question 9 (Continuation)**.** *Prove that R is positive definite. Deduce that, given $\boldsymbol{g} \in \mathbb{R}^n$, there exists a unique natural cubic spline g with knots at $x_1, \dots, x_n$ satisfying $g(x_i) = g_i$ for $i = 1, \dots, n$. Show further that there exists a non-negative definite matrix $K \in \mathbb{R}^{n\times n}$ such that*

$$\int_a^b g''(x)^2 \,\mathrm{d}x = \boldsymbol{g}^\top K \boldsymbol{g}.$$

Proof. It is immediate that the columns of R are linearly independent, so R is invertible and therefore positve definite. By question 8, there is exactly one $\boldsymbol{\gamma}$ such that there exists a natural cubic spline represented by $\boldsymbol{g}, \boldsymbol{\gamma}$, namely $\boldsymbol{\gamma} = R^{-1}Q^\top\boldsymbol{g}$. Note that, on $[x_i, x_{i+1}]$, the polynomial g has leading coefficient $c_i = \frac{1}{6}\left(\frac{\gamma_{i+1}-\gamma_i}{h_i}\right)$. Therefore, by partial integration we find (analogous to the next question)

$$\begin{aligned} \int_a^b g''(x)^2 \,\mathrm{d}x &= \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} g''(x)^2 \,\mathrm{d}x = -\sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} g'''(x)g'(x) \,\mathrm{d}x \\ &= -\sum_{i=1}^{n-1} c_i(g_{i+1} - g_i) = -\sum_{i=1}^{n} \frac{1}{6}\left(\frac{\gamma_{i+1}-\gamma_i}{h_i}\right)(g_{i+1} - g_i). \end{aligned}$$

Since every γ_i can be written as a linear combination of the g_i via $\boldsymbol{\gamma} = R^{-1}Q^\top\boldsymbol{g}$, it follows that there exists a matrix K such that $0 \le \int_a^b g''(x)^2 \,\mathrm{d}x = \sum_{i,j} K_{ij}g_ig_j$ for all $\boldsymbol{g}$, and therefore $\int_a^b g''(x)^2 \,\mathrm{d}x = \boldsymbol{g}^\top K\boldsymbol{g}$ and K is positive semi-definite. □

Question 10. *Let $n \ge 3$, let $a \le x_1 < \cdots < x_n \le b$, and let $\boldsymbol{g} \in \mathbb{R}^n$. Prove that the natural cubic spline interpolant to $\boldsymbol{g}$ at $x_1, \dots, x_n$ is the unique minimiser of $R(\tilde{g}) = \int_a^b \tilde{g}''(x)^2 \,\mathrm{d}x$ over all $\tilde{g} \in \mathcal{S}_2[a, b]$ that interpolate $\boldsymbol{g}$ at $x_1, \dots, x_n$.*

Hint: *Let g be the natural cubic spline interpolant, $h \coloneqq \tilde{g} - g$, and consider $\int_a^b g''(x)h''(x) \,\mathrm{d}x$.*

Solution. As in the hint we let g be the natural cubic spline interpolant, $\tilde{g} \in \mathcal{S}_2[a,b]$, and $h \coloneqq \tilde{g} - g$. We know that g is a cubic polynomial p_i on each interval $[x_i, x_{i+1}]$ (denote its leading coefficient by c_i), that g'' is continuous, and that $g'' = 0$ on $[a, x_1]$ and on $[x_n, b]$. Furthermore, we know that $h(x_i) = 0$ for all i. Using this, we can write

$$\begin{aligned}
\int_a^b g''(x)h''(x)\,\mathrm{d}x &= \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} p_i''(x)h''(x)\,\mathrm{d}x \\
&= \sum_{i=1}^{n-1} \left([p_i''(x)h'(x)]_{x=x_i}^{x_{i+1}} - \int_{x_i}^{x_{i+1}} p_i'''(x)h'(x)\,\mathrm{d}x \right) \\
&= -\sum_{i=1}^{n-1} c_i \int_{x_i}^{x_{i+1}} h'(x)\,\mathrm{d}x \\
&= -\sum_{i=1}^{n-1} c_i (h(x_{i+1}) - h(x_i)) = 0.
\end{aligned}$$

Now we find

$$R(g) = \int_a^b g''(x)^2\,\mathrm{d}x = \int_a^b g''(x)(\tilde{g}''(x) - h''(x))\,\mathrm{d}x = \int_a^b g''(x)\tilde{g}''(x) \overset{\text{CS}}{\leq} \sqrt{R(g)}\sqrt{R(\tilde{g})},$$

with equality if and only if $g'' = \tilde{g}''$ (by Cauchy-Schwarz). Rearranging gives $\sqrt{R(g)} \leq \sqrt{R(\tilde{g})} \implies R(g) \leq R(\tilde{g})$. Since $\tilde{g} \in \mathcal{S}_2[a,b]$ was arbitrary, we deduce that g is a minimiser of R over all function in $\mathcal{S}_2$.

For uniqueness: we already know that any other minimiser $h \in \mathcal{S}_2$ must satisfy $g'' = h''$ a.e., so $g - h$ is a polynomial of degree 1 a.e., and by continuity we know that $g - h$ is a polynomial of degree 1. However, since g and h must agree on $n \geq 3$ points it follows that $g = h$.